19 ESOLEC

The Egyptian Society Of Language Engineering Conference

The Nineteenth Conference of Language Engineering

Proceedings of the Conference

26-29 September, 2020
Alexandria, Egypt

Sponsors

Preface

The proceedings of the nineteenth Conference on Language Engineering contain nineteen papers reported to the conference. Two papers are written in Arabic, two papers are written in French and the rest are written in English language. The papers in the present document are classified in ten sessions corresponding to ten scopes of those mentioned in the call for papers.

The proceedings contain three invited papers. The first one is presented by Prof. Mohsen Rashwan, Professor in the Department of Electronics and Electrical Communications, faculty of Engineering, Cairo University, Egypt. He is Managing Director of RDI Corporation. The paper is about New Trends in Developing the Human Language Technologies and it will be presented in session one. The second invited paper is presented by Prof. Wafaa Kamel, Professor in Department of Arabic Language, Faculty of Arts, Cairo University, Egypt. The invited paper is about Lexicon and automatic processing systems and will be presented in session two. The third Invited paper is presented by Prof. Mervat Fashal, Professor in Department of Phonetics and Linguistics, Faculty of Arts, Alexandria University, Alexandria, Egypt. The third invited paper is about Human perception vs machine interface: an overview. It will be presented in session six.

The third session contains two papers about Computational Linguistics (I). The fourth session contains three paper dealing with Artificial Intelligence and NLP, while the fifth session contains four paper deals with Corpus based NLP. The invited paper will be in session six. Session seven contains two papers about Speech and Speaker Recognition.

Session eight includes two papers in the area of Speech perception. Session nine includes two papers about computational linguistics (II), while the four papers of session ten are concerned with speech analysis.

Conference Chairman

Prof. Dr. Mohamed Adeeb Riad Ghonaimy

Scope of the conference

Natural Language Processing has gained a lot of importance nowadays with many applications requiring real-time performance. In order to achieve the real-time requirements, the components of a Natural Language Processing (NLP) system should be made more efficient.

NLP overlaps to a large degree with computational linguistics (CL), especially when both are applied to standard Arabic or spoken varieties.

The Egyptian Society of Language Engineering (ESOLE) is a leading institution that interested in language engineering and computational linguistics especially for Arabic. Over the past 18 years, the ESOLE, in its conferences, has brought together researchers from across the field of natural language processing and computational linguistics and provided a wide-scope forum for discussing natural language processing researches as well as the best practices in its applications.

The 19th Annual Conference of Language Engineering (ESOLEC'19) will be held in Bibliotheca Alexandrina, Alexandria from 26 September to 29 September 2020. It continues this tradition and thus welcomes papers on all topics related to both natural language processing and computational linguistics, with the expectation that papers may include linguistic insight.

The relevant topics for the conference include, but are not limited to, the following topics:

Syntax, Semantics, Grammar, and the Lexicon.
Lexical Semantics and Ontology.
Phonology/Morphology, Word Segmentation, Tagging.
Text Mining, Paraphrasing and Summarization.
Speech Processing, Recognition and Synthesis.
Computational Linguistics.
Natural Language Processing for Information Retrieval.
Word Sense Disambiguation.
Automatic Character Recognition.
Semantic Role Labeling.
Sentiment Analysis and Opinion Mining.
Corpus-Based Modeling of Language.
Machine Translation and Translation Aids.
Multilingual Processing.
Statistical and Machine Learning Methods.
Social Networks and Contents Development Challenges.
Computational Forensic Phonetics and Linguistics.

Two workshops will be held in conjunction with ESOLEC'19 to encourage the exchange of ideas and to discuss challenging research issues in NLP. They will be held pre the main conference on Saturday September 26, 2020 and post the main conference on Tuesday, September 29, 2020.

Organization

Conference Chairman:

Prof. M. A. R. Ghonaimy
Conference Co-chair:
Prof. Sameh Alansary
Prof. Salwa Elramly
Prof. Magdy Nagy
Prof. Ayman Bahaa

Conference Organizing Committee

- Prof .M. Nagi
- Prof. S. Alansary
- Prof .A. Fahmy
- Prof. H. Korashy
- Prof. H. Barhamtoshy
- Prof. P. Elkafrawy
- Prof. Y. Elhamalawy
- M .Z. Abdelmageed

Local Organizing Committee

Maha Mousa
Marwa Saber
Amany Fashwan
Nahla Mohamed
Mayar Saeed
Israa Elhosiny

Technical Program Committee:

Prof. Taghrid Anber, **Egypt**
Prof. I. Abdel Ghaffar, **Egypt**
Prof. W. Fakhre, **Egypt**
Prof. M. Z. Abdel Mageed, **Egypt**
Prof. Khalid Choukri, ELDA, **France**
Prof. Nadia Hegazy, **Egypt**
Prof. Mona T. Diab, **USA**
Prof. Neamat Abdelkader, **Egypt**
Prof. Afaf AbdelFattah, **Egypt**
Prof. Mostafa Aref**, Egypt**
Prof. Allain Ramsey, **UK**
Prof. M. Elhamalawy**, Egypt**
Prof. S. Abdou**, Egypt**
Prof. A. A. Fahmy**, Egypt**
Prof. I. Farag, **Egypt**
Prof. Wafa Kamel, **Egypt**
Prof. S. Krauwer, **Netherlands**
Prof. Bente Maegaard, CST, **Denmark**
Prof. S. Elramly**, Egypt**

Prof. M. Nagy**, Egypt**
Prof. A. Rafae, **Egypt**
Prof. K. Refaat, **Egypt**
Prof. M. Rashwan**, Egypt**
Prof. K. Shalaan, **UAE**
Prof. H.M. AL-Barhamtoshy**, Saudi Arabia**
Prof. M. F. Tolba, **Egypt**
Prof. S. Elkareh, **Egypt**.

Invited Speakers

Prof. Mohsen A. A. Rashwan
Prof. Wafaa Kamel Fayed
Prof. Mervat Fashal

Sunday, 27/9/2020			
8:30	-	**10:00**	Registration
10:00	-	**11:00**	Opening Session
11:00	-	**12:00**	**Session 1:** Chairman**:** Prof. Aly Aly Fahmi **Invited Paper1: New Trends in Developing the Human Language Technologies** *Prof. Mohsen Rashwan* *Electronics& Communications Engineering Department, Faculty of Engineering, Cairo University.*
12:00	-	**12:30**	**Coffee Break**
12:30	-	**13:15**	**Session 2**: Chairman: Prof. Salwa Al-Ramly **Invited Paper 2: Lexicon and automatic processing systems.** *Prof. Wafaa Kamel Fayed* *Department of Arabic Language, Faculty of Arts, Cairo University, Cairo, Egypt.*
13:15	-	**14:15**	**Session 3**: **Computational Linguistics (I)** Chairman: Prof. Wafaa Kamel **1. Bibalex Arabic Linguistic Resources and Tools for Language Engineering** [*†]SamehAlansary, [**†]Magdy Nagi [†]*Bibliotheca Alexandrina, Alexandria, Egypt* [*]*Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt* [**]*Computer and System Engineering Department, Faculty of Engineering, Alexandria University, Alexandria, Egypt* **2. Shallow Parsing for Automatic Arabic Text Summarization** Sameh Alansary *Bibliotheca Alexandrina, Alexandria, Egypt* *Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt*
14:15	-	**15:15**	**Lunch Break**

15:15	-	**16:45**	**<u>Session 4</u>:Artificial Intelligence and NLP** Chairman: Prof. Mohsen Rashwan **1. Chatbot System Architecture** Moataz Mohammed[*] and Mostafa M. Aref[**] *[*]Computer Science Department, Faculty of Computer and Information Sciences, Ain-shams University, Cairo, Egypt.* **2. Arabic Optical Character Recognition using Sequence to Sequence Models** Mohamed Sobhi[*], Yasser Hifny[**], Saleh Mesbah[*] *[*]Arab Academy for Science, Technology and Maritime Transport, Alexandria, Egypt.* *[**]University of Helwan, Cairo, Egypt.* **3. The Methodology and Uses of Whale Swarm Algorithm** Amr M Sauber[*], Passent M. El-Kafrawy[*,**], Amr. F. Shawish[*] *[*]Faculty of Science, Menoufia University, Egypt* *[**]School of Information Technology and Computer Science, Nile University, Egypt*
16:45	-	**18:45**	**<u>Session 5</u>: Corpus based NLP** Chairman: Prof M. Younis Elhamalawy **1. A Critical Review of Language Resources and Tools for Arabic Sentiment Analysis.** Miramar Etman*, SamehAlansary* **Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt.* **2. A Pilot Study of Biber's Model for Language Variation Detection: A Language Engineering Approach** MaramElsaadany[*], SamehAlansary[**] *[*]Pharous University, Alexandria University, Alexandria, Egypt* *[**]Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt.* **3. Semantic role labeling system for modern standard arabic: a rule based approach** AmenaDeif,,SamehAlansary *Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt.* **4. دور القواعد اللغوية في التمييزالآلي بين معاني الحروف** **خالد مصطفى أبو شبانة[*]، د. أحمد عبد الغني[**]** [*]قسم اللغة العربية، كلية الآداب، جامعة الإسكندرية، مصر [**]قسم الصوتيات، كلية الآداب، جامعة الإسكندرية، مصر
18:45	-	**22:00**	Banquet

Monday, 28/9/2020			
10:00	**-**	**10:45**	**<u>Session 6:</u> Artificial Intelligence** Chairman: Prof. Sameh Alansary **Invited Paper 3: Human perception vs machine interface: an overview** Prof. Mervat Fashal *Professor in Department of Phonetics and Linguistics, Faculty of Arts, Alexandria University, Alexandria, Egypt*
10:45	**-**	**11:45**	**<u>Session 7</u>: Speech and Speaker Recognition** Chairman: Prof Waleed Fakhre **1. Automatic Arabic Speaker Recognition Using Gaussian Mixture Model** Mervat Fashal*, Amna Dheif*, Aya Nabil*, Rehab Arafat* **Phonetics and linguistics Department, Faculty of Arts, Alexandria University Alshatby, Alexandria, Egypt* **2. End-to-End Arabic Speech Recognition: A Review** Abdelaziz A. Abdelhamid[*], Hamzah A. Alsayadi[**], IslamHegazy[*], Zaki T. Fayed[*] [*]*Computer Science Department, Faculty of Computer and information Sciences, Ain Shams University, Cairo, Egypt* [**]*Computer Science Department, College of Computing and Information Technology, Shaqra University, Saudi Arabia*
11:45	**-**	**12:15**	**Coffee break**
12:15	**-**	**13:15**	**<u>Session 8</u>: Speech perception** Chairman: Prof Mervat Fashal **1. The Perception of Arabic Vowel Length by Native and Non-native Listeners: An Experimental Investigation** Eman Kassem* and Lamyaa Tawfik* **Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt* **2. An Investigation of the Correlation between Perceived Pauses and Syntactic Structures** Israa Elhosiny*, Mervat Fashal* **Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt*
13:15	**-**	**14:15**	**<u>Session 9</u>: Computational Linguistics (II)** Chairman: Prof. Seham El-Qareh **1. L'Application du Formalisme des FonctionsLexicales sur la Langue Arabe.** Racha Mohammad Salem *Département de Langue et de LittératureFrançaises, Faculté des Lettres, Universitéd'Alexandrie, Alexandrie, Egypt.* **2. QuellesContraintes pour Traduire la Morphologie et la Syntaxe?** <u>AsmaaGaafar Abdel-Rassoul</u> *<u>Faculte de Lettres, Universite de Menoufia</u>*

<table>
<tr><td>14:15</td><td>-</td><td>15:15</td><td>Lunch Break</td></tr>
<tr><td>15:15</td><td>-</td><td>17:15</td><td><u>Session 10</u>: Speech Analysis
Chairman: Amr Gody

1. قياس انفعالات الممثل الصوتية باستخدام تقنيات هندسة اللغة

د. صديقة لاشين
قسم الدراسات المسرحية-كلية الآداب-جامعة الإسكندرية

2. The Acoustic Characteristics of Read and Spontaneous Colloquial Arabic Speech Corpora: A Pilot Study.
Rudyna Ahmed
Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt

3. Syllables Classification of ASR using Hybrid Visual Features in Fixed HMM
Doaa A. Lehabik *, Amr M. Gody *, Mohamed H. Merzban*, Sameh F. Saad **
*Electrical Engineering Department, Fayoum University, Fayoum, EGYPT
**Modern Sciences and Arts University, 6 October City, Giza, Egypt

4. Creating and Implementing ArSL Corpus for Deaf Drivers
Samah A. Abbas*, Hassanin M. Al-Barhamtoshy**, Fahad M. Al-Otaibi**
*Management Information Systems Department, Faculty of Economics and Administration, King Abdulaziz University
Jeddah, Saudi Arabia
**Information Technology Department, Faculty of Computing & Information Technology, King Abdulaziz University
Jeddah, Saudi Arabia</td></tr>
<tr><td>17:15</td><td>-</td><td>18:00</td><td>Closing session</td></tr>
</table>

Table of contents

Computational Linguistics

Bibalex Arabic Linguistic Resources and Tools for Language Engineering

*†1Sameh Alansary, *‡ 2Magdy Nagi
[1]sameh.alansary@bibalex.org
[2]Magdy.nagi@bibalex.org

Bibliotheca Alexandrina, Alexandria, Egypt

†*Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt*

‡ *Computer and System Engineering Department, Faculty of Engineering, Alexandria University, Alexandria, Egypt*

Abstract— **The need for linguistic resources and tools for natural language processing (NLP) is constantly increasing due to the information revolution and technological development. Researchers in the field of language engineering have suffered from little or no presence of these resources. Bibliotheca Alexandrina (BA) has adopted a center for building linguistic resources and tools needed for natural language processing tasks and applications in order to contribute in building computational applications to keep pace with the huge technological development. This paper review the efforts of the ICT sector in BA over the last 15th year.**

Keywords: **Arabic linguistic resources, Arabic corpus, computational lexicon, Arabic automatic diacritization, Arabic automatic summarization, semantic annotation, tools for NLP.**

1 INTRODUCTION

There is no doubt that the technological development that we are witnessing in our live has begun to penetrate into our public and private worlds and Interject itself into our most personal matters with our desire or against our will. This technology was not only limited to man, but also included all aspects of life, and entered the worlds of languages, literatures, science, information, mathematics and philosophy, etc.[1]. And to keep up with this phenomenal development and deal with it the computer science has emerged, which constituted a specific leap in the field of science and technology, and this science has imposed itself urgently in all fields of knowledge, and this computer science or computer engineering has become the axis around which all aspects of human life revolve. This led to the emergence of language engineering or natural language processing. Natural language processing (NLP), including Information Retrieval, Machine Translation and other Natural Language-related disciplines, is showing more interest in the Arabic language in recent years.

Interest began in the engineering of the Arabic language, more than two decades ago, Suitable resources for Arabic are becoming a vital necessity for the progress of this research, but it was individual efforts and not organized and institutional work as the researcher and his team work, write and conduct his experiments and present important and effective results, and the one who comes after does not complete what he begins with then he starts from scratch, so the efforts Scattered, refined and construction horizontally, not vertically [1]. This resulting in the lack of linguistic resources and tools that can be adopted for building intelligent applications. For example, and not limited to, corpora are an important resource but Arabic lacks sufficient resources in this field, so a research projects need to compile a corpus, which represents the state of the Arabic language at the present time and the needs of end-users. Therefore many trials have been conducted to build Arabic corpora but some of them were unsuccessful trials and others were for commercial purposes. Another problem is the share of poorly resourced languages like Arabic in contributing to the field of computational linguistics and lexical resources is much less than the share of more well-resourced languages like English. Therefore, concerted efforts must be made and institutions and bodies supporting the field strive to establish research centers that work to build resources and tools.

In this regard, in 2005, Bibliotheca Alexandrina adopted the idea of building language tools and resources to serve the Arabic language in general and the field of Arabic natural language processing in particular. This is one of its objectives "A leading institution of the digital age" as it is a center of excellence in the production and dissemination of knowledge. ICT sector in Bibliotheca Alexandrina built a trained and dedicated team of linguists and engineers as a human resources and the infrastructure in order build the linguistic resources and tools needed for Arabic NLP tasks and applications. In this paper we will review the efforts of the ICT sector over the last 15th year. Section 2 discusses the linguistic resources that have been built, section 3 shows the tools that have been developed. Section 4 is a conclusion.

2 LINGUISTIC RESOURCES

A. International Corpus of Arabic (ICA)

Bibliotheca Alexandrina (BA) has initiated a big project to build the "International Corpus of Arabic (ICA)", a real trial to build a representative Arabic corpus as being used all over the Arab world to support research on Arabic. The International Corpus of Arabic is planned to contain 100 million words. It is planned to be analyzed morphologically, syntactically and semantically. The collection of samples is limited to written Modern Standard Arabic selected from a wide range of sources designed to represent a wide cross-section of Arabic; it is stimulating the first systematic investigation of the national variety as being used all over the Arab world [2].

It is designed to include 11 genres, namely; Strategic Sciences, Social Sciences, Sports, Religion, Literature, Humanities, Natural Sciences, Applied Sciences, Art, Biography and Miscellaneous which are in turn further classified into 24 sub-genres, namely; Politics, Law, Economy, Sociology, Islamic, Pros etc. Moreover, there are 4 sub-sub-genres, namely; Novels, Short Stories, Child Stories and plays

Planning of ICA data collection is based on some criteria related to corpus design such as representativeness, diversity, balance and size were taken into the consideration. In collecting a corpus that represents the Arabic Language, the focus was to cover the same genres from different sources and from all around the Arab nations. Almost all publications of the Arab nations have been covered and other publications from outside the Arab nations as al-Hayat magazine which is published in London.

ICA data is composed of Modern Standard Arabic (MSA) written texts. There are different resources for compiling the data. It is composed of four sources, namely; (1) Press source that is divided into three sub-sources, namely; (a) Newspapers, (b) Magazines and (c) Electronic. (2) Net articles, (3) Books and (4) Academics.

Corpus analysis is both qualitative and quantitative. One of the advantages of corpora is that they can readily provide quantitative data that intuitions cannot provide reliably. The use of quantification in corpus linguistics typically goes well beyond simple counting. Table 1 shows some of the numbers of ICA data coverage. It must be noted that total number of "Tokens" refers to all word forms except numbers, foreign words and punctuations to reflect the real size of the used word forms before the analysis stage.

TABLE I

QUALITATIVE LINGUISTIC ANALYSIS FOR ICA STATISTICS

Statistics	Total Number
No. of texts	70,022
No. of words	79,569,384
No. of Tokens	76,199,414
No. of Types	1,272,766
No. of ICA sources	4
No. of sub sources	3
No. of genres	11
No. of sub genres	24
No. of sub sub-genres	4
No. of countries	20
No. of covered years	22

No. of writers	1021

B. MASAR: A Morphologically Annotated Gold Standard Arabic Resource

The first stage of linguistic analysis of the International corpus of Arabic is to analyze the 100 million words of the ICA corpus morphologically that began at Bibliotheca Alexandrina (BA) in 2007. Before beginning the morphological analysis process, each text is preprocessed and marked up with some structural markup such as beginning and end of document, title, paragraph or question. It was preferred to develop our own morphologically annotated gold standard analyzed resource to be used while analyzing the whole ICA data, since it contains more information and details than any other annotated data. It has two releases; the first one consists of about 500 thousand manually annotated words and the second one consists of about 1.5 million automatically annotated words using BASMA [3] that are verified for quality assurance. In this paper, the first release is our concern. The following is the description of the selected data of the first release and the issues that were faced during the analysis process:

The stem-based approach (concatenative approach) has been adopted as the linguistic approach to analyze the ICA. The second version of Buckwalter Morphological Arabic Morphological Analyzer (BAMA 2.0) [4] has been selected since it is a well-known analyzer in the literature and has even been considered as the "most respected lexical resource of its kind". Although Buckwalter has many advantages including its ability to provide a lot of information such as Lemma, Vocalization, Part of Speech (POS), Gloss, Prefix(s), Stem, Word Class, Suffix(s), Number, Gender, Definiteness and Case or Mood, it does not always provide all the information the ICA requires, and in some cases, the analyses provided would need some modification. Its results may give the right solution for the Arabic input word, provide more than one result that needs to be disambiguated to reach the best solution, provide many solutions, but none of them is right, segment the input words wrongly without taking the segmentation rules in consideration or provide no solutions. Consequently, solutions enhancement is needed in these situations. For more details about BAMA problems and how these problems have been handled see [5].

As a results of BAMA's problems, Number, gender and definiteness need to be modified according to their morph-syntactic properties. Some tags had been added to BAMA's lexicons, some lemmas and glossaries had been modified and others had been added. In addition, three new features had been used while developing MASAR; Name Entity, Root and Stem Pattern.

In the first release of MASAR 1,111 text documents are selected from ICA corpus for texts that were published in 2006-2007. It contains 570,137 tokens of which 69,937 are punctuations, numbers, and Latin strings, and 500,200 are Arabic word tokens (81,487 word types). These texts are selected from different sources in ICA; Press, Net Articles and Books. Moreover, these selected texts covered more than one genre as Table 1 shows. In Press Source, the texts are selected from newspapers, magazines and electronic press covering different countries.

The data of MASAR have been morphologically annotated by ten well-trained linguistic annotators using their linguistic information behind the knowledge of traditional MSA grammar. For a quality control comparison of annotators, nine files with total of 9,153 words (and varying number of POS choices per word) were each tagged independently. Out of 9,153 words, 449 words show some disagreement. All three agreed on 89% of the words; the pairwise agreement is at least 94.8%.

Once the annotation process is done, the annotated files are saved in a database in way where each feature is saved separately in order to ease the next stages of syntactic and semantic analysis processes as shown in figure 1.

word	lemmai	voc	gloss	pr1	stem	suf1	gen	num	def	casee	root	Stem_Pattern
في	fiy	fiy	in		fiy/PREP						NONE	NONE
أثناء	vanaY	>avonA'i	during		>avonA'/I		FEM	PL_BR	DEF (EL	i/GEN	vny	>afoEaAl
توجههم	tawaj~uh	tawaj~uhi	attitude		tawaj~uh,	him/PO	MASC	SG	DEF (EL	i/GEN	wjh	tafaE~ul
بسيارته	say~Arap	bisay~Ara	by/with	bi/PRE	say~Ar/N	at/NSU	FEM	SG	DEF (EL	i/GEN	syr	faE~aAl
إلى	<ilaY	<ilaY	to/tow		<ilaY/PRE						NONE	NONE
مدرستهم	madorasa	madorasa	school +		madoras/	at/NSU	FEM	SG	DEF (EL	i/GEN	drs	mafoEal
في	fiy	fiy	in		fiy/PREP						NONE	NONE
شارع	$AriE	$AriEi	street		$AriE/NO		MASC	SG	DEF (EL	i/GEN	$rE	faAEil
المدارس	madorasa	AlmadAri	the + sc	Al/DET	madAris/		FEM	PL_BR	DEF	i/GEN	drs	mafaAEil
بحي	Hay~	biHay~i	by/with	bi/PRE	Hay~/NO		MASC	SG	DEF (EL	i/GEN	Hyy	faEol
الرمال	ramol	Alr~imAli	the + sa	Al/DET	rimAl/NO		FEM	PL_BR	DEF	i/GEN	rml	fiEaAl
المكتظ	mukotaZ^	Almukota	the + ov	Al/DET	mukotaZ^		MASC	SG	DEF	i/GEN	kZZ	mufotaEal/mufotaEil
بالمدارس	madorasa	biAlmadA	with/by	bi/PRE	madAris/		FEM	PL_BR	DEF	i/GEN	drs	mafaAEil
الابتدائية	{ibotidA}i	Al{ibotid/	the + el	Al/DET	{ibotidA}i	ap/NSU	FEM	SG	DEF	i/GEN	bd'	{ifotiEaAliy~
غرب	garob	garoba	west/W		garob/NO		MASC	SG	DEF (EL	a/ACC	grb	faEol
غزة	gaz~ap	gaz~ap	Gaza		gaz~ap/N		FEM	SG	DEF		NONE	NONE
.	Punc	Punc	Punc	Punc	Punc	Punc	Punc	Punc	Punc	Punc	Punc	Punc
P/	EOF_Prg	EOF_Prg	EOF_Pr	EOF_P	EOF_Prg	EOF_Pr	EOF_Prg	EOF_P	EOF_Pr	EOF_Prg	EOF_Prg	EOF_Prg
/P	BOF_Prg	BOF_Prg	BOF_Pr	BOF_P	BOF_Prg	BOF_Pr	BOF_Prg	BOF_P	BOF_Pr	BOF_Prg	BOF_Prg	BOF_Prg
وذكرت	*akar-u	wa*akara	and + m	wa/CC	*akar/PV	at/PVSU					*kr	faEal
مصادر	maSodar	maSAdiru	sources		maSAdir/		FEM	PL_BR	INDEF	u/NOM	Sdr	mafaAEil
أمنية	>amoniy~	>amoniy~	security		>amoniy~	ap/NSU	FEM	SG	INDEF	N/NOM	'mn	faEoliy~
فلسطينية	filasoTiyn	filasoTiyn	Palestin		filasoTiyn	ap/NSU	FEM	SG	INDEF	N/NOM	NONE	NONE
أن	>an~a	>an~a	that		>an~a/SU						NONE	NONE
مسلحين	musal~aH	musal~aH	armed/		musal~aH	iyna/NS	MASC	PL	INDEF	ACC	slH	mufaE~al
ملثمين	mulav~an	mulav~an	masked		mulav~an	iyna/NS	MASC	PL	INDEF	ACC	lvm	mufaE~al
يستقلون	{isotaqal~	yasotaqil'	they (p	ya/IV3	sotaqil~/I	uwna/I'				MOOD:I	qll	sotafoEil

Figure 1: Sample of ICA Gold Standard Resource

C. LESAN: Lexical Semantic Annotated Resource

The second stage of linguistic analysis of the International corpus of Arabic is to analyze the 100 million words of the ICA corpus on the lexical semantic level which is repented in LESAN; LExical Semantic ANnotated Resource. It is built during the process of developing the International Corpus of Arabic (ICA) and benefited from its morphological analysis stage using a built semantic lexicon. The used semantic lexicon is a lemma-based lexicon and each word is assigned the suitable lexical semantic meaning according to its context and its selected lemma/tag in the morphological analysis stage.

To detect the different meanings of the same word form, we need first to detect its different morphological analyses according to the context in which it occur as table II shows:

TABLE III
Morphological analyses of the word form ‘عين’ with some of its meaning

Morphological Analysis	Tag	Sense	Context
عَيَّنَ	Past Verb	حَدَّدَ	عَيَّنَ الْمُدِيرُ أَسَاسَ الْمُشْكِلَةِ
		وَظَّفَ – اِخْتَار - قَلَّدَ	عَيَّنَ رَئِيسُ الْعَمِلِ الشَّخْصَ الْمُنَاسِبَ لِلْوَظِيفَةِ
		خَصَّصَ	عَيَّنَ الْمَلِكُ حِصَّةً مِنْ الْمَالِ لِلْفُقَرَاءِ
عَيْن	NOUN	عُضْوُ الْإِبْصَارِ لِلْإِنْسَانِ وَغَيْرِهِ مِنْ الْكَائِنَاتِ	عَيْنُ الذَّبَابَةِ مُرَكَّبَةٌ
		جَاسُوس	وَجَعَلَهُ عَيْنَ مِصْرَ عَلَى إِسْرَائِيلَ
		نَبْع	عَيْنٌ جَارِيَةٌ

There are two competing models of lexical processing in the literature. The first proposes that we rely on mental lexicons. The second claims there are no mental lexicons; we identify certain items as words based on semantic knowledge. Thus, the former approach – the multiple-systems view – posits that lexical and semantic processing are sub-served by separate systems, whereas the latter approach – the single-system view – holds that the two are interdependent [4].

We have developed a lexicon in which the possible senses for a word are defined. Various lexicographic sources have been used as references to build the inventory of senses for each word, mainly the electronic lexicon "قاموس المعاني"[1] . The developed lexicon is lemma based; it contains 44358 lemmas. The lexicon entries cover all content words in ICA[2].

Regarding Modern Standard Arabic, some senses need to be added in the semantic lexicon and have been dealt with in the manual annotation process (section 4) due to two main reasons:

1. The word is newly used in Arabic and is not included in classical lexicons.
2. The senses of certain words are found in the classical Arabic lexicons, but the modern usage of these words require new senses to be added. For example, the word "فاجأ" "fAja>" has a new sense "أَدْهَشَ" ">adoha$" "surprise" as in example [1].

   تَمَسُّكُ التُّونِسِيِّينَ بِالْإِسْلَامِ فَاجَأَ الْغَرْبَ [1]

   tamas~uku Alt~unisiy~iyna biAl<isolAmi fAja>a Algaroba

   The Tunisians' adherence to Islam surprised the West

Once the annotation process is done, the annotated files are saved in a database in a way where the suitable sense of each word depending on the context in which it occurs is saved as shown in figure 2:

word	lemmaid	tags	Lexical_Semantic_Annotation
/P		BOF_Prg	
(		Punc	
ساحة	sAHap	NOUN	ساحة: مكان واسع
كبيرة	kabiyr	ADJ	كبير : هائل عظيم / ذو مرتبة عالية
في	fiy	PREP	
قرية	qaroyap	NOUN	قَرْيَة: عَدَدٌ قَلِيلٌ مِنَ الدُّورِ فِي بُقْعَةٍ مِنَ الأَرْضِ فِي السَّهْلِ أَوِ الْجَبَلِ .
.		Punc	
نوافذ	nAfi*ap	NOUN	نَافِذَةٌ : شُبَّاك
وأبواب	bAb	NOUN	باب: مدخل
تطل	>aTal~	IV	أطل: أشرف
على	EalaY	PREP	
الساحة	sAHap	NOUN	ساحة: مكان واسع
-		Punc	
بعض	baEoD	NOUN	بَعْض: جزء ، نوع من ، طائفة
الأشجار	Sajarap	NOUN	شَجَرةُ : نباتٌ يقوم على ساقٍ صُلبة وقد يُطلَقُ على كلّ نبات غير قائم
الذابلة	*Abil	ADJ	ذَابِل : يابس ، أصفر ، ميت
،		Punc	
وبعض	baEoD	NOUN	بَعْض: جزء ، نوع من ، طائفة
الأشجار	$ajarap	NOUN	شَجَرةُ : نباتٌ يقوم على ساقٍ صُلبة وقد يُطلَقُ على كلّ نبات غير قائم
المقطوعة	maqoTuwE	ADJ	مقطوع : مفصول بعضه عن بعض
أغصانها	guSon	NOUN	غصن: فرع ، ساق
،		Punc	
أو	>aw	CONJ	
جذوعها	ji*oE	NOUN	جذع : ساق النخلة ونحوها
،		Punc	
تفاجئ	fAja>	IV	فاجأ : جاء في وقت غير متوقع، باغت
عين	Eayon	NOUN	عَيْنٌ : عضو الإبصار للإنسان وغيره من الحيوان
الناظر	nAZir	NOUN	ناظر : باصر بعينيه
فتصدمها	Sadam-i	IV	صدم: أفزع
.		Punc	

Figure 2: Semantically Annotated Sample

Twenty well-trained linguistic annotators have semantically annotated the data of LESAN. In order to make sure that the annotators follow the same guidelines and of almost the same level of professionalism, nineteen files with total of about 19,225 words (and varying numbers of senses choices per word) were annotated independently by each annotator and they were compared together. Out of 19, 225 words, only

[1] https://www.almaany.com/ [Accessed 6-2-2020]

[2] https://www.bibalex.org/ica/ar/default.aspx [Accessed 6-2-2020]

2884 words show some disagreement. All twenty annotators agreed on 85% of the words; the pairwise agreement is at least 92.3%.

D. MUHIT

There is a clear need for dictionaries translating between a large number of languages. The creation of a dictionary of good quality takes a lot of time, and given the fact that 5000-6000 languages yield 25-30 million pairs of languages, it is important to have a database that provides the possibility to translate directly between pairs of languages. A well-known problem is that words are often hard to match across languages i.e. different words from different languages do not have the same range of meanings, not all words from one language have an equivalent in the other, etc. Moreover, a multilingual lexical database should meet a number of requirements [6].

Over the last few decades, a large amount of new lexical resources have arisen: machine readable dictionaries, lexical databases, full-form lexicons, morphological databases, semantic networks, dictionary databases, etc.

The purposes of usage of a lexical database are different from those of a dictionary. MUHIT database differs from the design of dictionary databases in a number of things. Since it does not list only lemmas, but complete inflected forms as well as the amount and type of information stored for each lemma is different [7].

In this section, we will be presenting MUHIT (Multilingual Harmonized database). MUHIT database differs from the design of dictionary databases in a number of things. Since it does not list only lemmas, but complete inflected forms as well as the amount and type of information stored for each lemma is different [7].

1) What is MUHIT?:

"MUHIT" is an abbreviation for (MUltilingual Harmonized dIcTionary) but it is not just an abbreviation, it constitutes a meaningful word. The name "MUHIT" has been inspired by the Arabic word "المحيط" (al-Muhit), which means "Ocean" and "comprehensive". Moreover, it is part of one of the most celebrated Arabic dictionaries (al-Qamus al-Muhit), compiled by al-Firuzabadi (1329–1414) that has been widely used for centuries [8].

MUHIT is a multilingual electronic lexical database which has been developed within the universal networking language (UNL) framework [9], [10], [11] and it is one of the UNDL Foundation [12] products in cooperation with Bibliotheca Alexandrina. MUHIT is available on the UNLLab (http://www.unlweb.net/lab/) where entries have been interlinked by sense, and natural language word forms have been associated to a uniform concept identifier (UWs), the words of UNL [10], [13]. In 2013 MUHIT contained about 10,000,000 word forms collected from more than 40 languages, by 2019 updated and the number of words reached 20,000,000 words collected from more than 140 languages.

MUHIT was developed mainly for cross-language word search. This means that MUHIT can help users in finding and using information in their native or non–native languages. This is clear from the design of MUHIT. The methodology adopted in developing the computational lexicon "MUHIT" depends on the combination of WordNet approach and corpus based approach through many projects (BRUNO, MIR, LPP, LIS and LEWIS & SHORT) which have been developed through the UNLarium environment.

2) MUHIT linguistic infrastructure

It is important to explain the linguistic infrastructure of MUHIT. Linguistic knowledge that appears in MUHIT has been assigned to all words through $\text{UNL}^{\text{arium}}$ encompassing different linguistic levels: morphological information, morpho-syntactic information, syntactic information and semantic information. UNL uses a standard and universal list of features (Tagset) to describe all types of the linguistic information concerning every natural language word in MUHIT. The linguistic infrastructure of MUHIT discussed in details in [14].

3) How toUse the System?

MUHIT is available on the UNL^{lab} and it is accessible through the websitehttp://www.unlweb.net/lab/. Being an online application provides the user with a number of advantages; data is stored remotely hence requiring no disk space from the part of the user, no installation or updating is required, and most importantly providing an easy access through the internet. Moreover, the user is not required to have an account in the UNL to access this application. MUHIT is a free and open source application. MUHIT provides a variety of search options for more comprehensive results. A distinct advantage that the system provides is that there is no need to choose a specific language ahead, either as a source language or as a target language, rather the system searches for the string in all existing dictionaries belonging to the different participating languages. The search does not only include base forms, which are the typical headwords of most dictionaries, but all existing inflected forms as well as. MUHIT also has adopted the regular expression system to provide a concise and flexible means for matching strings of text, such as particular characters, words, or patterns of characters. Figure 3 shows the results of searching the word "cate in MUHIT in different languages.

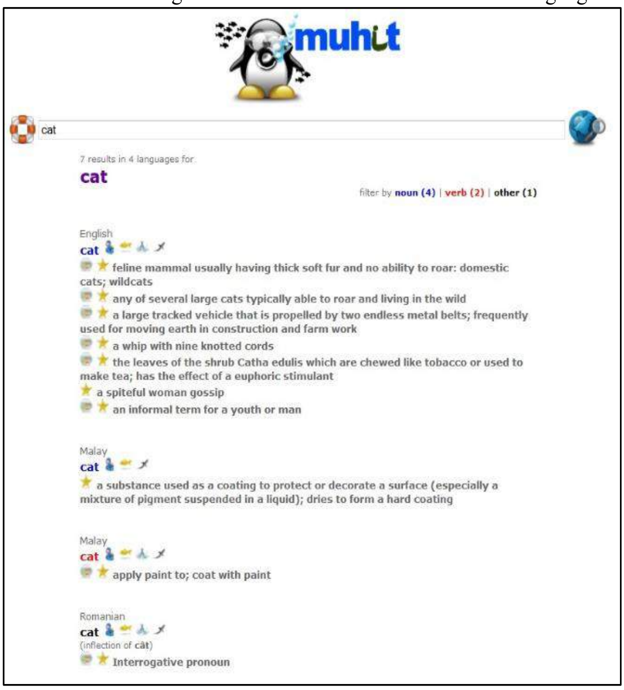


Figure 3: Searching the word "cate in MUHIT

The result appearing for the search word is accompanied with either three or four icons on its left. Each icon is responsible for providing the user with certain information. The information displayed are the author of the entry, The features of the entry, such as part of speech, number, gender etc., The inflections of the entry, if any. For each sense of the entry, it is also provided: The set of synonyms in the same language. The set of synonyms in different languages.

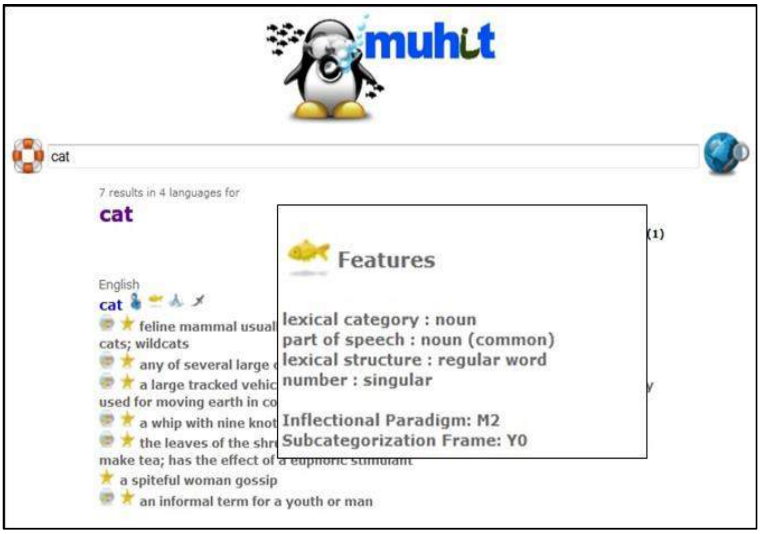


Figure 4: The lexical description of the entry

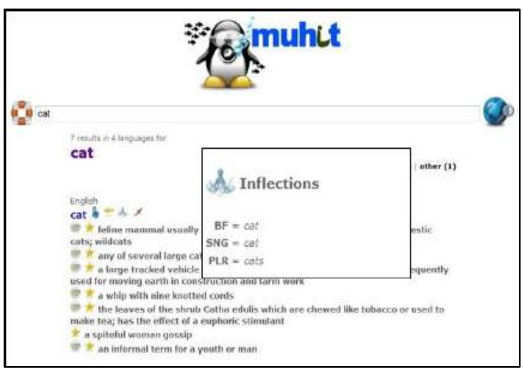


Figure 5: The inflected forms of the search word

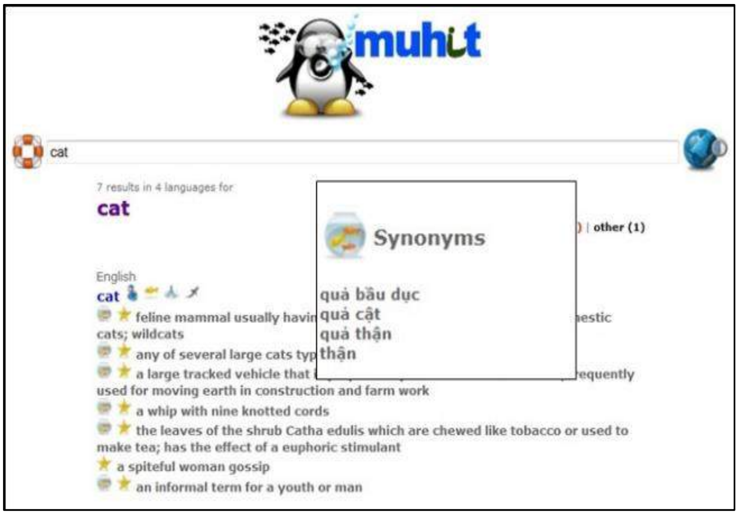


Figure 5: The synonyms of the search word

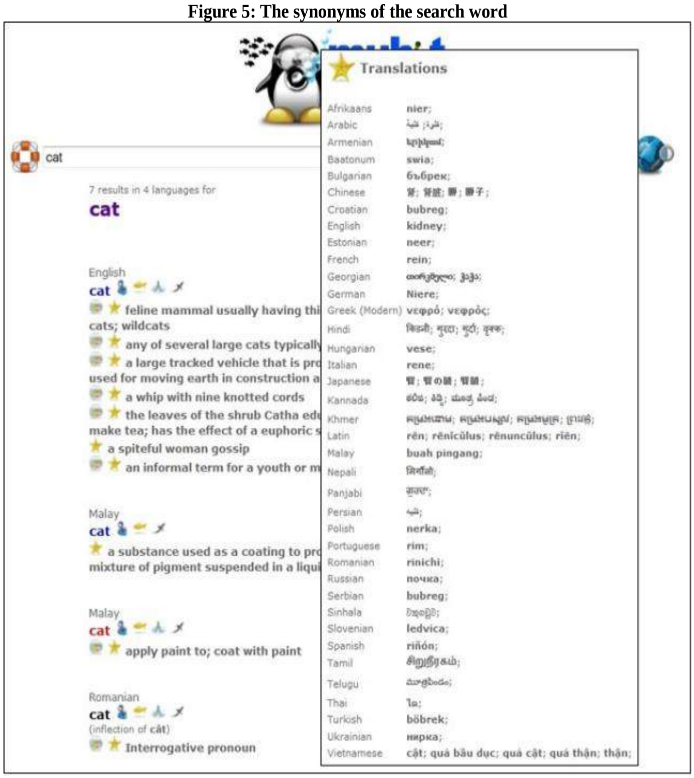


Figure 6: The available translations of the search word

The options mentioned above have combined both the advantages of a monolingual dictionary and a multilingual one. On the one hand, it provides a detailed linguistic description, that is usually found in learners dictionary only, for each result. On the other hand, it provides a translation for each single sense.

MUHIT is an open initiative. Hence, it is designed in a way that makes the participation easily accessible. The participation is done on different levels. All details about MUHIT are discussed in [14].

E. Arabic Computational Lexicon

The quality of a NLP system depends to a great extent on the quality of the linguistic resources it uses, with dictionaries being the single most important variety of lexical resources as they contain an enormous amount of linguistic knowledge; lexical, semantic, morphological and syntactic [15].

Bibliotheca Alexandrina (BA) built an Arabic computational lexicon using the UNL framework. In this attempt, we sought to overcome the problems of accessibility and usability. Since most of the available lexical resources are not really available, as they are usually exclusively used within the software they are embedded in, we tried to offer a solution by making our dictionary open source thus making it a universally available lexical resource. Lexical entries in the Arabic computational lexicon are composed of an Arabic headword that is accompanied by a list of linguistic attributes. After a process of experimentation, it has been decided that the most suitable base form for our Arabic computational lexicon is the lexeme. This result has been reached after proving that a lexeme-based Arabic headword can be more efficiently and economically transformed into the required word form; plural, dual, feminine, past tense...etc. by using the linguistic information assigned to the headword within the entry. The use of a base form has eliminated the need to include all the possible word forms of a single lexical item in the dictionary; and has thus cut down the redundancy to a great degree. Moreover, this has rendered the dictionary more efficient and robust.

1) Linguistic description

This subsection presents how Arabic words are described in the Arabic computational lexicon using a list of features extracted from the UNDL Foundation tagset[3]. The Arabic computational lexicon utilizes a semantic ontology. This ontology classifies the entities existing in the natural world in a semantic hierarchy. This hierarchy points out the particular type of each concept and the kind of relation it holds with other concepts in the ontology. Each entry in this hierarchy carries a set of features and attributes and all subclasses of this concept inherit the properties of that class [16].

The linguistic information contained in the lexicon is divided into three types. 1) a list of simple features describing the lexical structure of words such as part of speech, gender, number, voice, transitivity and etc. 2) inflection paradigms to describe the morphological behavior of the Arabic words, these are a kind of morphological rules that are responsible for generating the different word forms out of the stored Arabic word as shown in figure 8.

(1) MCL&SNG:=0>"";	→	مدرس
(2) FEM&SNG:=0>"ة";	→	مدرسة
(3) MCL&DUA&NOM:=0>"ان";	→	مدرسان
(4) FEM&DUA&NOM:=0>"تان";	→	مدرستان
(5) MCL&DUA&ACC:=0>"ين";	→	مدرسين
(6) FEM&DUA&ACC:=0>"تين";	→	مدرستين
(7) MCL&DUA&GNT:=0>"ين";	→	مدرسين
(8) FEM&DUA&GNT:=0>"تين";	→	مدرستين
(9) MCL&PLR&ACC:=0>"ين";	→	مدرسين
(10) MCL&PLR&GNT:=0>"ون";	→	مدرسون
(11) FEM&PLR:=0>"ات";	→	مدرسات

Figure 7: The rules used to generate the different word forms out of the lexeme "مدرس" paradigm M532

[3] http://www.unlweb.net/wiki/Tagset

(1) MCL,SNG,3PS,PAS,ACV	→	قال
(2) MCL,SNG,3PS,PRS,ACV	→	يقول
(3) FEM,SNG,3PS,PAS,ACV	→	قالت
(4) FEM,SNG,3PS,PRS,ACV	→	تقول
(5) MCL,PLR,3PP,PAS,ACV	→	قالوا
(6) FEM,PLR,3PP,PAS,ACV	→	قلن
(7) MCL,SNG,2PS,PAS,ACV	→	قلت
(8) MCL,SNG,2PS,PRS,ACV	→	تقل
(9) MCL,DUA,2PP,PRS,ACV	→	قلتما
(10) MCL ,PLR,2PP,PAS,ACV	→	قلتم
(11) SNG,1PS,PRS,ACV	→	أقول
(12) PLR,1PS,PAS,ACV	→	قلنا
(13) MCL,SNG,3PS,PAS,PSV	→	قيل
(14) MCL,SNG,3PS,PRS,PSV	→	يقال
(15) FEM,SNG,3PS,PAS,PSV	→	قيلت
(16) FEM,SNG,3PS,PRS,PSV	→	تقال

Figure 8: The rules used to generate the different word forms out of the lexeme "قال" paradigm M103

2) Subcategorization rules to describe the syntactic behaviour of the words. These are the rules that determine the number and types of the necessary syntactic arguments (specifiers, complements and adjuncts) of the verb. For example in the sentence "نقل الإسكندر علوم فاس وبابل إلى أثينا"

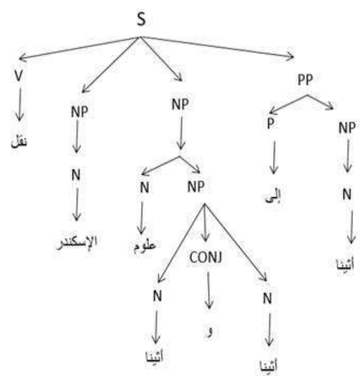


Figure 9: The necessary syntactic arguments of the verb "نقل" Subcategorization Frame Y521

All of the above mentioned features enable us to build a comprehensive Arabic lexical resource especially designed to suit the different applications of NLP. Currently, the BA Arabic computational lexicon contains about 1,500,000 Arabic headwords. Figure 11 shows a sample of the dictionary entries in their final form as they are ready to be exported.

```
☐ [كاتب]{113957}"102870092"
(LEMMA=كاتب,BF=كاتب,LEX=N,POS=NOU,LST=WRD,GEN=MCL,NUM=SNG,PAR=M628,FRA=Y0,ANI=NANM,ABN=CCT,ALY=ALI,ANI=NANM,CAR=CTB,SEM=ARF,SFR=K0)
<ar,0,1>;
☐ [قال]{204288}"201009240"(LEMMA=قال,BF=قال,LEX=V,POS=VER,LST=WRD,TRA=TSTD,PAR=M103,FRA=Y0,SEM=CMV)<ar,255,3>;
☐ [مدرسة]{107624}"108283180"(LEMMA=مدرسة,BF=مدرسة,LEX=N,POS=NOU,LST=WRD,GEN=FEM,NUM=SNG,PAR=M583,FRA=Y0,SEM=GRO)<ar,1,1>;
☐ [illegible]{209524}"201018352"(LEMMA=[illegible],BF=[illegible],LEX=V,POS=VER,LST=WRD,TRA=TST2,PAR=M242,FRA=Y518,SEM=CMV)<ar,9,2>;
☐ [طالب]{108524}"110274815"(LEMMA=طالب,BF=طالب,LEX=N,POS=NOU,LST=WRD,GEN=MCL,NUM=SNG,PAR=M600,FRA=Y0,ABN=CCT,ANI=ANM,SEM=HUM)<ar,1,1>;
☐ [illegible]{92117}"300885099"(LEMMA=[illegible],BF=[illegible],LEX=J,POS=PTL,LST=WRD,DEG=PST,PAR=M467,FRA=Y0)<ar,1,1>;
☐ [illegible]{109004}"300957176"(LEMMA=[illegible],BF=[illegible],LEX=J,POS=PTL,LST=WRD,DEG=PST,PAR=M467,FRA=Y0,SEM=HPP)<ar,2,0>;
☐ [illegible]{94432}"301876670"(LEMMA=[illegible],BF=[illegible],LEX=J,POS=PTL,LST=WRD,DEG=PST,PAR=M467,FRA=Y0)<ar,1,2>;
☐ [illegible]{202714}"115669360"(LEMMA=[illegible],BF=[illegible],LEX=N,POS=NOU,LST=WRD,GEN=MCL,NUM=SNG,PAR=M551,FRA=Y0)<ar,0,0>;
☐ [عنوان]{124324}"106787150"(LEMMA=عنوان,BF=عنوان,LEX=N,POS=NOU,LST=WRD,GEN=MCL,NUM=SNG,PAR=M585,FRA=Y0,ANI=NANM,SEM=CMN)<ar,1,4>;
☐ [illegible]{209434}"105839024"
(LEMMA=[illegible],BF=[illegible],LEX=N,POS=NOU,LST=WRD,GEN=MCL,NUM=SNG,PAR=M599,FRA=Y0,ABN=ABT,ALY=ALI,ANI=NANM,CAR=CTB,SEM=CGN,SFR=K0)<ar,25,0>;
☐ [illegible]{82527}"400114029"(LEMMA=[illegible],BF=[illegible],LEX=A,POS=AAV,LST=WRD,PAR=M0,FRA=Y0,SEM=MAN)<ar,1,1>;
```

Figure 10: The Arabic UNL dictionary

The entry in the Arabic computational lexicon contains the Arabic Headword, a number string representing the Headword meaning or rather its place in the hierarchy of concepts, a list of linguistic features including the list of simple features and the number of the inflectional paradigm that describes the morphological behavior of the headword as well as the number of the Subcategorization rule that describes it syntactic behavior, finally, the entry contains frequency and priority numbers that are used to indicate the frequency of usage of the headword; this piece of information is necessary to determine the most used senses of a particular Arabic word and can, thus, aid the process of word-sense disambiguation. All details about the description of the Arabic computational lexicon are discussed in [17].

3 TOOLS

A. *Arabic Diacritization System (Alserag)*

In Modern Standard Arabic, texts are typically written without diacritical markings. The diacritics are important to clarify the sense and meaning of words. The process of automatically restoring diacritical marks is called diacritization. Diacritization helps the reader in disambiguating the text or simply in articulating it correctly. As Arabic is a language where the intended pronunciation of a written word cannot be completely determined by its standard orthographic representation; it rather depends on a set of special diacritics. The absence of these diacritics in Arabic text increases lexical and morphological ambiguity, because one written form can have several vocalizations, each vocalization may have different meaning(s) [18, 19]. However, these diacritics are generally left out in most genres of written Arabic, which results in widespread ambiguities in vocalizations and meaning Diacritizing [20]. Arabic written text is crucial for many NLP tasks, translation can be enumerated among a longer list of applications that vitally benefit from automatic diacritization [21, 22, and 23].

Much work has been done on Arabic diacritization. The actually implemented systems can be divided into two categories [24]: Systems implemented by individuals as part of their academic activities and systems implemented by commercial organizations for realizing market applications. There are also other available systems as Mishkal Arabic diacritizer, and Harakat Arabic diacritizer; they are free Arabic diacritizers, which are available online. Finally, on March Google has launched an innovative new Google Labs Arabic tool called Tashkeel, a tool that adds the missing diacritics to Arabic text. Unfortunately, the tool is not available now [20].

The developed system (Alserag) is a rule based system. Alserag has been developed in 2016, it depends on two resources: the Arabic diacritized dictionary and a set of linguistics rules. The Arabic diacritized dictionary is a dictionary where Arabic natural language words exist with their diacritics, along with the corresponding linguistic features, which describe the Arabic word morphologically, syntactically and semantically. While the linguistic rules work through three modules in order to provide fully diacritized Arabic words namely, morphological analysis module, syntactic analysis module and morph-phonological processing module [20, 25]. These modules are achieved through 7 main phases: (i) Preprocessing which is responsible for auto-correcting the raw text and segmenting the Arabic text into sentences. (ii) Tokenization which is the process of splitting the natural language input into lexical items. (iii) Disambiguation which is a process of choosing the right internal diacritization for the word from the dictionary. (iv) Name entity recognition. (v) Syntactic shallow parsing which is an analysis of a sentence by identifying its constituents (NPs, JPs---etc.). (vi) Case ending module which is responsible for predicting the arguments of the predicate and assigning the diacritical marks that are attached to the ends of words to indicate their grammatical function. (vii) Morph-phonological module which is a series of rules that focus on the sound changes that take place in morphemes (minimal meaningful units) when they are combined to form words [20, 25].

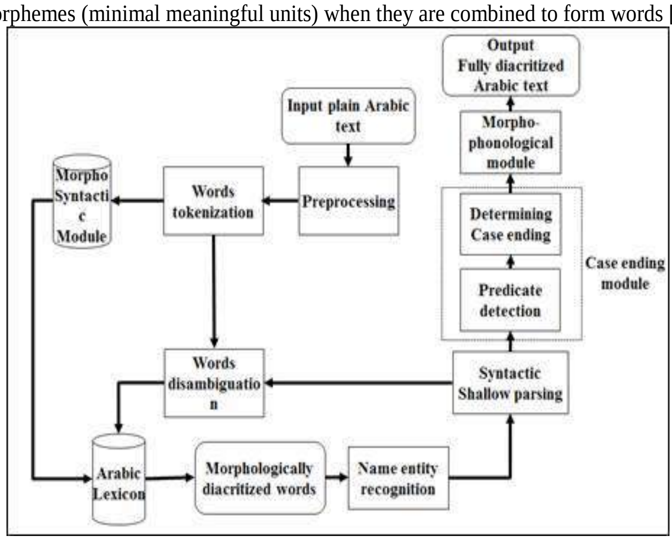


Figure 11: Architecture of ALSERAG

There are two engines that are used in Alserag, the first is Interactive ANalyzer (IAN), which is used in the analysis process, and includes a grammar for natural language analysis. The syntactic processing is done automatically through the natural language analysis grammar, the second is dEep-to-sUrface natural language GENErator engine (EUGENE) which is used in the generation process, and receives the analyzed input and provides a diacritized output without any human intervention [26].

Although the system was able to overcome many challenges, it still has some limitations. The system has been improved in this phase and still there is more potential for further improvements. One of the most difficult problems that faces a parser is structural ambiguity, since it leads to problems in determining the boundaries between constituents. Despite these limitations, the system is proved to be promising when tested and demonstrated [25].

The corpus has been selected from the International Corpus of Arabic (ICA). The selected corpus size is 400,000 Modern Standard Arabic words; they are divided into 300,000 words as training data and 100,000 words as testing data. The selected texts are from different sources; Newspapers, Net Articles and Books representing the following genres; politics: 148,211, miscellaneous: 100,253, child stories: 57,174, economy: 34,930, society: 32,955 and sports: 26,477 [20]. The results of the system were evaluated for accuracy against the reference using two metrics; diacritization error rate (DER) and word error rate (WER). In addition to calculating DER and WER, the evaluation system calculates internal diacritics and case ending separately [25]. Table IV shows the evaluation history of the automatic diacritization system (Alserag) over the last three years from 2017 to 2019 and figure 13 shows the latest evaluation results of the 100.000 words.

TABLE V

EVALUATION HISTORY OVER LAST THREE YEARS (2017-2019)

Year	Morphology	Case Ending	DER	WER
2017	2.56%	13.51%	4.86%	15.79%
2018	1.77%	9.39%	3.37%	11.7%
2019	1.53%	8.44%	2.98%	10.50%

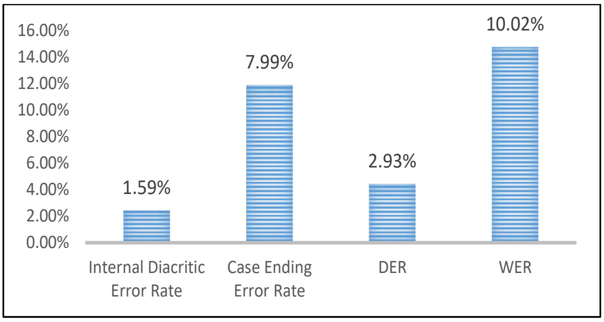


Figure 12: Evaluation of the whole data of Alserag

Alserag system, which is a rule-based system, is benchmarked against three known diacritization systems: Harakat, Mishkal, and Aldoaly, which are statistical based systems as shown in table VI [25].

TABLE VII

BENCHMARKING OF THE WHOLE DATA OF ALSERAG AMONG THE OTHER THREE SYSTEMS

	Alserag	Harakat	Mishkal	Aldoaly
Internal Diacritic Error Rate	1.59%	43.30%	32.53%	80.92%
Case ending Diacritic Error Rate	7.99%	16.23%	31.15%	89.72%
Diacritic Error Rate (DER)	2.93%	37.63%	32.24%	82.76%
Word Error Rate (WER)	10.02%	43.49%	65.00%	97.87%

According to the recent results obtained by the benchmarking process, our system scored the least error rate followed by Harakat and Mishkal and finally Aldoaly, which scored over 80% error rate.

B. *Arabic Summarization System*

Due to the daily increasee of the electronic documents on the Internet, everyone should benefit from this revolution of information. The important way to access these documents and get the core content of them and utilize from the information existing in these documents became an urgent need. Hence the need for tools to facilitate this appeared, which is called Automatic Text Summarization. Radev et al. (2002) defines a summary as “a text that is produced from one or more texts, that conveys important information in the original text(s), and that is no longer than half of the original text(s) and usually significantly less than that”. In order to generate a summary, we have to identify the most significant pieces of information exist in the text, omitting the redundant information and reducing details.

Text summarization strongly appeared in many applications that will help in facilitating life, for example we do not need to consume time in reading news with all its details, so we can summarize news to SMSs or to news mobile applications that will save a lot time. In addition, businessmen need fast summaries for their reports and documents, etc. Therefore, when we extract this summarized version from a certain document by means of a computer, automatically, we call this Automatic Text Summarization.

There are two main approaches of summarization, extraction approach, where all sentences are first rated according to their importance, and then a summary is generated after choosing a number of top scoring sentences. On the other hand, the abstraction approach in which documents have to be recognized and interpreted, and then the summary is generated.

Arabic language is more sophisticated and needs much time compared with English and other European languages. There are different existing systems of automatic text summarization, which classified into two categories: Systems implemented on Arabic languages, such as AQBTSS, ACBTSS, Sakhr Arabic Summarizer, Lakhas, Aramedia, Ikhtasir. Moreover, other systems implemented on other different languages, such as Copernic Summarizer, Pertinence summarizer, Kify Text Summarizer, SweSum, MEAD.

In a series of developments, an Automatic Arabic Text Summarization System is proposed, this system uses different stages in order to summarize Arabic documents. The first stage is responsible for extracting the most informative sentences of the document. There are many factors should be considered while selecting the important sentences such as: the type of the document, the sentence length, word frequency, and the number of topic words the sentence contain, they are all helping factors in determining the most important sentences. The second stage is responsible for excluding the non-informative constituents after identifying which constituents are incident and could be omitted, and which are principal and vital. The system represents the constituents syntactic structure of each sentence using X-bar theory in order to categorize them into incident and principal. Our proposed automatic summarizer adopts the X-bar theory, because of its compatibility with Arabic language, and because it supplies a systematic description of Standard Arabic sentence formation.

There are two engines that are used in our proposed Arabic automatic summarization system during the linguistic processing stage, the first is Interactive ANalyzer (IAN) which represents the analysis process, the second is dEep-to-sUrface natural language GENErator engine (EUGENE) which represents the generation process.

The system depends on three linguistics resources, namely; the word-net, UNL encycolopedia, and EDGES. EDGES: is the Entity Discovery and Graph Exploration System, a user-friendly visualization tool used for exploring semantic networks by enabling concept (words of the document) expansion, collapsing and navigation.

UNL Encyclopedia: It is also known as UNL Example Base; it contains semantic relations between UWs along with a degree of probability. It comprises information that is related to the probability of occurrence rather than the possibility of occurrence.

There is an inventory study has been done on a number of documents in order to measure the accuracy of our proposed system and the established summarization grammar. A sample of this test was taken from a random document whose total number is 115 words as shown in figure 14, while when automatically summarized it becomes 43 words as in figure 15. The summarization grammar could detect that some constituents, which have specific grammatical functions, could be omitted from a document during the summarization process, such as the highlighted constituents in the document in figure 14 which might be Appositions, phrases starts with specific keywords, Relative clauses. All of those do not add additional information to the sentence`s meaning.


أكد المعلق الرياضي أحمد الطيب، أن محمد أبو تريكة لاعب الأهلي المعتزل، رفض تقاضي أي أموال نظير مشاركته مع نجوم العالم في احتفال دولة الكويت بافتتاح ملعب جابر الدولي.

وقال الطيب لمصراوي: "أبو تريكة اعتذر عن تقاضي مبلغ 50 ألف يورو نظير مشاركته في الاحتفال، وقال إنه لشرف أن يشارك في يوم عظيم وحدث كبير بدولة الكويت".

وضمت تشكيلة نجوم العالم كل من: "الحارس جيمس ليمان، والحارس ديفيد جميز، وديفيد بيكهام، ميشيل سلجادو، اليساندرو نيستا، وكارلوس بويول، وزامبروتا، وربرتو كارلوس، ولويس فيجو، وروبيرتو بيرس، رونالدينهو، واليساندور ديل بيرو، وداوت يورك، أندريه شفشينكو، ومحمد أبو تريكة".

وطلب الشيخ صباح الأحمد الجابر الصباح مصافحة أبو تريكة بين شوطي المباراة، التي انتهت بفوز نجوم الكويت برباعية مقابل هدفين.

Figure 13: Original Document

أكد أحمد الطيب، أن محمد أبو تريكة لاعب الأهلي المعتزل، رفض تقاضي أي أموال نظير مشاركته مع نجوم العالم في احتفال دولة الكويت بافتتاح ملعب جابر الدولي.

وضمت تشكيلة نجوم العالم

وطلب الشيخ صباح الأحمد الجابر الصباح مصافحة أبو تريكة بين شوطي المباراة

Figure 14: The automatic summarized document

There are many tools and measures that could be used in the evaluation of the automatic summary; however, many researches admitted that there are two main points have to be considered in the evaluation, the Compression Ratio (CR) and the Omission Ratio (RP). Therefore, our proposed automatic summarizer follows this pace and the results are significantly accurate.

C. BASMA : BibAlex Standard Arabic Morphological Analyzer

The process of developing a morphological analyzer tool for ICA began in 2007 which is known as BibAex Arabic Morphological Analyzer Enhancer (BAMAE) but later known as BibAlex Standard Arabic Morphological Analyzer (BASMA). It is a system that has been built to morphologically analyze and disambiguate the Arabic texts depending on BAMA's enhanced output of ICA. It was preferred to use BAMA's enhanced output of ICA since it contains more information than any other systems of BAMA's enhanced output. This is the reason why the members of ICA team aimed to build their own morphological disambiguator (BAMAE), figure 16 shows the architecture of BASMA.

In order to reach the best solution for the input word, BAMAE preforms automatic disambiguation process carried on three levels, depends primarily on the basic POS information (Prefix(s), Stem, Tag and Suffixes) obtained from enhanced BAMA's output:

- Word level which avoids or eliminates the impossible solutions that Buckwalter provides due to the wrong concatenations of prefix(s), stem and suffix(s).
- Context level where some linguistic rules have been extracted from the training data to help in disambiguating words depending on their context.
- Memory based level which is not applicable in all cases; it is only applicable when all the previous levels failed to decide the best solution for the Arabic input word.

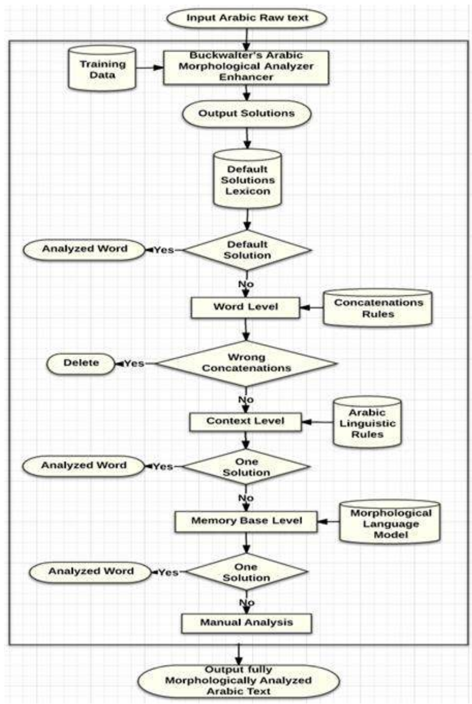


Figure 15: BAMAE Architecture

After selecting the best POS solution for each word, BAMAE detects the rest of information accordingly. It detects the lemmas, roots (depending primarily on the lemmas), stem patterns (depending on stems, roots and lemmas), number (depending on basic POS and stem patterns), gender (depending also on basic POS, stem patterns and sometimes depending on number), definiteness (depending on POS or their sequences), case (depending on definiteness and sequences of POS) and finally it detects the vocalization of each word. For more details see [3].

D. UNL Editor (An Annotation tool for Semantic Analysis)

Semantic Annotation has become an increasingly important research topic being a fundamental element of many Natural Language Processing applications like information retrieval, query answering and information extraction. Semantic annotation is additional information in a document that identifies or defines the semantics of a part of that document [27].

In the context of the UNL (The Universal Networking Language), a semantically based interlingua to break language barriers between human languages, the UNDL Foundation in cooperation with Bibliotheca Alexandrina has started an initiative for building a tool for semantic annotation called the UNL Editor; a visual editor designed with the intention of providing full semantic annotation, thus analyzing natural language texts and, generating UNL documents. This tool is based upon a comprehensive visualization of the entire process of the annotation. It is uniquely designed on linguistic background; adopting certain linguistic theories closely related to computational linguistics in terms of using unified super sets of semantic relations [28] thus overcoming the problem of conflicting and confusing names [29], and making use of renowned lexical recourses; WordNet [30]. Moreover, it provides a powerful visual interface for working with UNL data both in a textual and graphical mode with friendly interface creating an appropriate environment for navigating through the needed steps of providing the analysis; it offers a visualization of the analysis through graphs which aids the representation of the semantic network created with every sentence analyzed.

The UNL Editor provides a means enabling the analysis of the underlying semantic relations composing the Natural Language sentences. It is designed on linguistic bases. On a semantic assumption or rather on semantic theory stating that a deep semantic analysis for a natural language text requires two levels of semantics; lexical semantics and grammatical semantics and it is discussed in details in S. Alansary, M. Nagi and N. Adly, 2011 [31]. Lexical semantics in UNL Editor is expressed through creating the nodes, a process in which every word or rather every concept in the sentence to be analyzed is matched with its corresponding ID, meaning that a single node may contain more than one lexical item; a compound word, as long as it is representing a single concept. Grammatical Semantics in the UNL Editor is expressed in terms of a range of semantic relations, and a list of attributes. UNL Editor has proposed a unified super set of the semantic relations. These relations are highly standardized as each relation is clearly defined in the UNL framework. The tool includes 45 semantic relations and they are a closed set of relations. Relations are used to describe the objectivity information of sentences. In the UNL, relations are normally regarded as representations of semantic cases or thematic roles (such as agent, object, instrument, etc.) between concepts. They are used in form of arcs connecting a node to another node in a UNL graph. They correspond to two-place semantic predicates holding between two concepts. Relations are always used to describe semantic dependencies between syntactic constituents. For more information about the semantic relation within the UNL frame work see [32].

Other additional information are being presented through attributes, representing information conveyed by natural language grammatical categories (such as tense, mood, aspect, number, etc) [33]. In opposition to relations, attributes correspond to one-place predicates; attributes are intended to be used as annotations made to nodes or hypernodes of a U

NL hypergraph. Moreover, they are also a closed set. Attributes modify concepts or semantic networks to indicate subjectivity information such as about how the speaker views these states-of-affairs and his attitudes toward them and to indicate the property of the concepts, for more information about the attributes within the UNL frame work see [32].

1) How to Use the UNL Editor?

In order to use the tool, the user will have to sign in the UNL web then access the UNL Editor via UNL dev application (The UNL Integrated Development Environment). Figure 17 describes the steps for reaching the semantic graphic representation. Within the UNL Editor Frame work, the process of decision making is completely human: the user uploads the text to be analyzed; selects the corresponding IDs; relate nodes through creating semantic relations; and assigns attributes to nodes. The first step will be the text input and text segmentation followed by concepts selection to create the nodes and adding the appropriate attributes to each node then the final step in order to reach the semantic graph will be linking the created nodes by semantic relations [34].

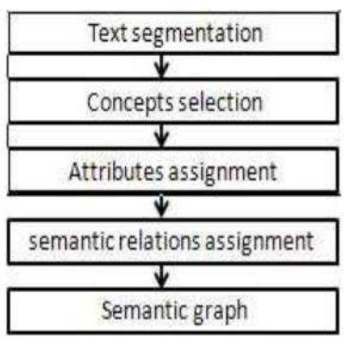
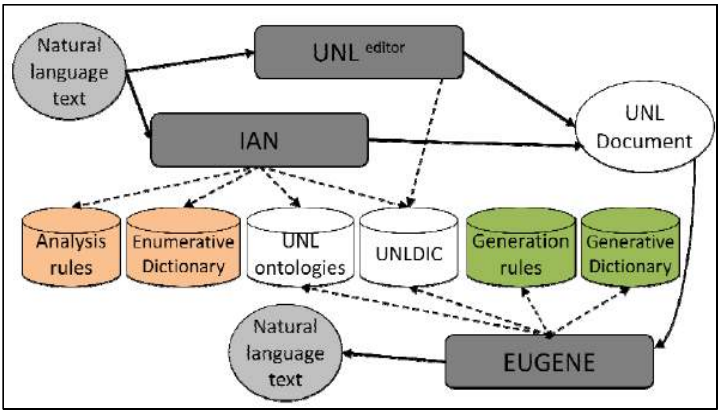


Figure 16: Steps for reaching the semantic graph

The detailed description of how the UNL Editor works is illustrated with examples and figures in [35].



4 CONCLUSIONS

The paper shed light on the achievements and contributions of Bibliotheca Alexandrina in the field of language engineering over the last 15th years. Where Bibliotheca Alexandrina made efforts to build infrastructure and human resources to work on building linguistic resources and tools to be available for researchers and specialists in the field of language engineering. The paper briefly discussed these resources and tools.

REFERENCES

[1] N. Ahmed "*هندسة اللغة العربية: مطلب قومي وهدف استراتيجي*" Arabiyât: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban, 4, (1), 2017, 88-101.

[2] S. Alansary, M. Nagi and N. Adly, *Building an International Corpus of Arabic (ICA): progress of compilation stage.* In proceedings of the 7th International Conference on Language Engineering. Cairo, Egypt, 5–6 December 2007.

[3] S. Alansary, *BASMA: BibAlex Standard Arabic Morphological Analyzer*. In the proceedings of (ESOLE). Egypt, Cairo, 9-10 December.

[4] T. Buckwalter, *Buckwalter Arabic Morphological Analyzer Version 2.0. Linguistic Data Consortium*, University of Pennsylvania, 2004. LDC Catalog No.: LDC2004L02.

[5] S. Alansary, *BAMAE: Buckwalter Arabic Morphological Analyzer Enhancer*. In proceedings of Arabic Language Processing Conference. Rebate, Morocco, 2-3 May: Mohamed Vth University.

[6] M. Janssen, "*SIMuLLDA*: *a Multilingual Lexical Database Application using a Structured Interlingua*", Doctoral dissertation, Utrecht University, June, 2002.

[7] M. Janssen. *Lexical vs. Dictionary Databases: design choices of the MorDebe system. Papers in Computational Lexicography* - COMPLEX, Budapest, Hungary, 2005.

[8] The UNLweb website: http://www.unlweb.net/muhit/index.php?muhit=help, (accessed in October 2013).

[9] H. Uchida, M. Zhu, T. G. Della Senta, "A Gift for a Millennium", November 1999.

[10] S. Alansary, M.Nagi, N.Adly, UNL+3: The Gateway to a Fully Operational UNL System. In Proceedings of 10th International Conference on Language Engineering, Cairo, Egypt, 2010.

[11] J. Cardeñosa, A. Gelbukh, E. Tovar (eds.): Universal Networking Language: advances in theory and applications.(Research on Computer Science, 12). Mexico City: National Polytechnic Institute. 443pp, 2005.

[12] The UNDL Foundation website: www.undl.org

[13] R. Martins, V. Avetisyan, "Generative and Enumerative Lexicons in the UNL Framework," in Proc. Of Seventh International Conference on Computer Science and Information Technologies (CSIT 2009), 28 September - 2 October, 2009, Yerevan, Armenia Proceedings of CSIT 2009.

[14] S. Alansary, *MUHIT: A Multilingual Lexical Database*, 13th International Conference on Language Engineering, Ain Shams University, Cairo, Egypt, December 11 - 12 2013.

[15] P. Bjakman and V. Raskin. *What Linguists Might Contribute to Dictionary Making If They Could Get Their Act Together*. In The Real World Linguistics. Ed. Ablex, Norwood, NJ, 1986.

[16] M. Obitko, *Ontologies description and applications*, Research Report No. 126/01Czech Technical University, Praue, 2001.

[17] S. Alansary, *A UNL Based Approach for Building an Arabic Computational Lexicon* , INFOS2012, Cairo University - Egypt, May 16 - 17 2012.

[18] Bouamor. H., Zaghouani.W., Diab.M., Obeid. O., Oflazer. K., Ghoneim.M., and Hawwari. A.: A Pilot Study on Arabic Multi-Genre Corpus Diacritization Annotation. Proceedings of the Second Workshop on Arabic Natural Language Processing, pages 80–88, Beijing, China, c2014 Association for Computational Linguistics (2015).

[19] EL-Desoky. A., Fayz. M. and Samir, D.*: A smart Dictionary for the Arabic Full-Form Words*. (IJSCE). ISSN: 2231-2307, Volume-2, Issue-5 (2012).

[20] S. Alansary, "*Alserag: An Automatic Diacritization System for Arabic*". The 2nd International Conference on Advanced Intelligent Systems and Informatics (AISI'16), Cairo, Egypt, 2016.

[21] Smr, O.: Yet Another Intro to Arabic NLP,http://ufal.mff.cuni.cz/~smrz/ANLP/anlp-lecture-notes.pdf (2005).
[22] Rashwan., M., Abdou.S., Rafea., A.: *Stochastic Arabic Hybrid Diacritizer*, IEEE trans. Natural Language Processing and Knowledge Engineering, pp.1-8, 24-27 (2009).
[23] Attia, M., Mohsen A. A. Rashwan, Mohamed A.S. A. A. Al-Badrashiny.: Fassieh®, *a Semi-Automatic Visual Interactive Tool for Morphological*, PoS-Tags, Phonetic, and Semantic Annotation of Arabic Text Corpora, IEEE trans. AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 17, NO. 5, pp.916-925 (2009).
[24] Al Badrashiny, M.: *Automatic Diacritizer for Arabic Text*. A Thesis Submitted to the Faculty of Engineering, Cairo University in Partial Fulfillment of the Requirements for the Degree of master of science in electronics & electrical communication (2009).
[25] S. Alansary, (2016, December). Improving Alserag Arabic Diacritization Grammar through Syntactic Analysis. In 16th international conference on language engineering, Cairo, Egypt.2016.
[26] Alansary. S.: A Suite of Tools for Arabic Natural Language Processing: A UNL Approach, the special session on Arabic Natural Language Processing: Algorithms, Resources, Tools, Techniques and Applications, (ICCSPA'13), Sharjah, UAE (2013).
[27] H. Bunt and Ch. Overbeeke," *A note on the definition of semantic annotation Languages*" , Proceedings of the 8th International Conference on Computational Semantics, pages 268–271, Tilburg, January 2009. c 2009 International Conference on Computational Semantics 2009.
[28] Ch. Johnson and Ch. J. Fillmore: "T*he FrameNet tagset for frame-semantic and syntactic coding of predicate-argument structure*". In the Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics (ANLP-NAACL 2000), Seattle WA, pp. 56-62, April 29- May 4, 2000.
[29] D.R Dowty, *Thematic Proto-Roles and Argument Selection*, Linguistic Society of America, 1991.
[30] C. Fellbaum, *WORDNET: An Electronic Lexical Database*, the MIT Press, 1998.
[31] N. Chomsky, *Studies on Semantics in Generative Grammar*, Mouton publisher, 1972.
[32]Sameh Alansary, Magdy Nagi, Noha Adly , *UNL+3: The Gateway to a Fully Operational UNL System* , 10th International Conference on Language Engineering, Ain Shams University, Cairo, Egypt, December 22 - 23 2010.
[33] D. Pisoni and R. Remez, *The Handbook of Speech Perception*, Blackwell Publishing Inc., 2004.

[34] M.Zhu and H.Uchida, "UNL annotation", UNL Center, UNDL Foundation specifications and manuals, 2003.

[35] S. Alansary, M. Nagi, N. Adly *, UNL Editor: An Annotation tool for Semantic Analysis* , 11th International Conference on Language Engineering, Ain Shams University, Cairo, Egypt, December 14 - 15 2011.

BIOGRAPHY

Dr. Sameh Alansary: *Director of Arabic Computational Linguistic Center at Bibliotheca Alexandrina, Alexandria, Egypt.*

He is professor of computational linguistics in the Department of Phonetics and Linguistics and the head of Phonetics and Linguistics Department, Faculty of Arts, Alexandria University. He obtained his MA in Building Arabic Lexical Databases in 1996, and his PhD from Nijmegen University, the Netherlands in building a formal grammar for parsing Arabic structures in 2002. His main areas of interest are concerned with corpus work, morphological analysis and generation, and building formal grammars.

He is also the head of Arabic Computational Linguistics Center in Bibliotheca Alexandrina. He is supervising and managing the Universal Networking Language project in Library of Alexandria since 1-6-2005 till now.

Dr. Alansary is the co-founder of the Arabic Language Technology Center (ALTEC), an NGO aims at providing Arabic Language resources and building a road map for Arabic Language Technology in Egypt and in the Middle East. He has many scientific works in Arabic Natural Language Processing published in international conferences and periodicals, and a member in many scientific organizations: (1) Egyptian Society of Language Engineering, Cairo, (2) Arabic Linguistic Society - USA, (3) Association of Computational Linguistics - USA – Europe, (4) Universal Networking Language foundation, United Nations, Geneva, Switzerland.

Dr. Magdy H. Nagi: *Senior Consultant, ICT Sector at Bibliotheca Alexandrina*

Dr. Nagi is a Professor in the Computer and Systems Engineering department, Faculty of Engineering, Alexandria University. He obtained his Ph.D. from the University of Karlsruhe, in 1974, where he served as Lecturer for two years and as a Consultant to its Computer Center from 1974-1990. During this period, he also served as Consultant to many companies in Germany such as Dr. Otker, Bayer, SYDAT AG, and BEC.

On the national level, he was a Consultant to many projects under the umbrella of either the University of Alexandria or the Faculty of Engineering for designing and/or implementing automation projects for governmental authorities or public sector companies, such as the Ministry of Interior, the Health Insurance Organization (HIO), the Social Insurance Organization (SIO), and the Customs Authorities.

Dr. Nagi has served, since 1995, as Consultant to the Bibliotheca Alexandrina. Among his activities were the design and installation of Bibliotheca Alexandrina's network and information system, namely a trilingual information system that offers full library automation.

In 2001, he got appointed as the Head of the Information and Communication Technology (ICT) Sector of the Bibliotheca Alexandrina and occupied that post till 2012. He currently serves as a senior Consultant to the ICT Sector and continues to oversee the various projects and partnerships established between the ICT Sector and many international institutions.

Dr. Nagi is a member of the ACM and the IEEE Computer Society as well as several other scientific organizations. His main research interests are in operating systems and database systems. He is author/co-author of more than 80 papers.

TRANSLATED ABSTRACT

المصادر والأدوات اللغوية العربية لمكتبة الإسكندرية في مجال هندسة اللغة

تتزايد الحاجة إلى الموارد والأدوات اللغوية لمعالجة اللغة الطبيعية (NLP) باســـتمرار بســـبب ثورة المعلومات والتطوير التكنولوجي .لقد عانى الباحثون في مجال هندســـة اللغة من قلة أو عدم وجود هذه الموارد. وقد تبنت مكتبة الإســـكندرية (BA) مركزًا لبناء الموارد والأدوات اللغوية اللازمة لمهام وتطبيقات معالجة اللغة الطبيعية من أجل المســـاهمة في بناء التطبيقات الحاسوبية لمواكبة التطور التكنولوجي الضخم. تستعرض هذه الورقة جهود قطاع تكنولوجيا المعلومات والاتصالات في مكتبة الإسكندرية على مدار الخامسة عشر عاما الماضية منذ عام 2005 وحتى عام 2020.

Shallow Parsing for Automatic Arabic Text Summarization

Sameh Alansary

Bibliotheca Alexandrina, Alexandria, Egypt
Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt
sameh.alansary@bibalex.org

***Abstract*—in this paper, we propose a summarization method from the viewpoint of the shallow syntactic analysis involves different stages: sentences extraction stage, syntactic analysis stage, and generation stage. Finally, the evaluation summarization process will discussed.**

***Keywords:* Text summarization (TS), text analysis, parts of speech, shallow parsing and abstractive summarization.**

1 INTRODUCTION

The automatic text processing is a research field that is currently extremely active. Automatic text summarization, aims to reduce the size of a text while preserving its information content. A summarizer is a system that produces a condensed representation of its input's for user consumption [1]. In order to generate a summary of a document or a sentence, we have to identify the most significant pieces of information exist in a document or a sentence, omitting the redundant information and reducing details. A summary can be employed in an indicative way – as a pointer to some parts of the original document, or in an informative way – to cover all relevant information of the text . In both cases, the most important advantage of using a summary is its reduced reading time [1]. Summary generation by an automatic procedure has also other advantages: (i) the size of the summary can be controlled; (ii) its content is determinist; and (iii) the link between a text element in the summary and its position in the original text can be easily established [1].
There different systems of automatic text summarization for different languages are listed and discussed in the previous work [2]. The actually implemented systems can be divided into two categories: Systems implemented on Arabic languages and other systems implemented on other languages rather than the Arabic:

There are number of Arabic text summarization systems as referred in the previous study [2]: the Arabic Query-Based Text Summarization System, the Arabic Concept-Based Text Summarization System, Sakhr summarizer, Lakhas Arabic summarizer, and the summarizer of Aramedia.

The Non-Arabic Text Summarization Systems: Copernic Summarizer, Kify Text Summarizer, SweSum, MEAD [2]. The systems are listed and each on is described in details in [2].

In the recent years, many approaches have appeared due to the huge amount of information overloaded on the Web, but there are two dominant approaches of summarization: the extraction approach and the abstraction approach.

Our proposed system depends on different important stages. The previous experiences and approaches were inspiring to build our proposed system. The first stage is the **Normalization**, which plays an important role in summarization process, in addition to, another two essential roles, Firstly, splitting long stream contexts into sentences or phrases to facilitate the processing. Secondly, modifying the spelling errors in the input texts. The second stage is to **disambiguate** the correct parts of speech of the word forms. After that comes the **shallow parsing stage**, how to parse, connect different constituents' boundaries and build the tree of each sentence. Then it comes to the **Summarization** stage (Also called **Deletion** stage) that decides which phrases are going to be remained or to be excluded. Finally, it is the role of **Generation** to convert the summarized trees to a well-formed generated text. Each stage will be described in details in section 3:

In this research, we mainly focus on introducing the effort that have been done in the area of the syntactic analysis of the documents. Our objective is to generate document summaries using the shallow parsing of the sentences document that is an analysis of a sentence, which first identifies constituent parts of sentences (nouns, verbs, adjectives, etc.) and then links them to higher order units that have discrete grammatical meanings. This paper is organized as follows: the following section describes some related works with respect to the use of the syntactic analysis of documents for summarization. Section 3 introduces our idea behind the selection of the shallow parsing of the sentences document that is to be summarized. Section 4 presents the summarization methodology and describes the experiment illustrated by examples, which illustrate the different stages of the analysis. Finally, Section 6 concludes this paper.

2 AUTOMATIC TEXT SUMMARIZATION DIFFERENT APPROACHES

In the recent years, many approaches have appeared due to the huge amount of information overloaded on the Web, but there are two dominant approaches of summarization as discussed before in our previous study [2]:

A. Extraction Approach

This approach makes use of different properties of the text to weight the sentences by using a combination of statistical heuristics or linguistic features. Each sentence in the text is assigned a score using a combination of statistical heuristics. The sentences are sorted in descending order according to their score values and an appropriate number of the highest scored sentences are selected from the text to form the summary according to the summarization ratio. The sentences that have the highest scores are considered very important and included in the summary [2].
Extraction is mainly concerned with judging the importance or the indicative power of each sentence in a given document. All sentences are first rated in terms of their importance, and then a summary is obtained by choosing a number of top scoring sentences [3].

The methods used for determining the weights of the sentences are: Cue method, location method and title method. One of the most shortage of using the extractive approach is that the important and relevant information is usually spread out throughout the document and the extractive techniques are unable to combine all of these unless increasing the size of the summery [3]. In addition, when the sentences are picked up as they are the pronouns often tend to lose their references thus creating a confusion to trace the meaning. If there is a confliction in the information, it may not be presented accurately. In order to overcome these problems abstractive summarization techniques can be used as in [3].

B. Abstraction Approach

The abstraction approach involves simplifying and condensing the text. When text summaries are created manually using the abstraction approach, humans read the text, reinterpret it, and rewrite it [3]. Producing abstracts is not a simple task because central topics have to be identified, topics have to be interpreted, and then the summary is generated. The abstraction approach is much more difficult to be programmed than extraction, therefore extraction is the more commonly used approach in automatic text summarization.

Abstractive summarization includes understanding the main concepts and relevant information of the main text, then expressing that information in short, and clear format. Abstractive summarization techniques be classified into two categories: 1) structured based and 2) semantic based methods. Each method will be classified in details in the following sub section.

1) Structured Based Abstractive Summarization Methods

Structured based approaches determines the most important information through documents by using templates, extraction rules and other structures such as tree, ontology etc. [3].The structured based abstractive category has three different methods are: the first is Rules based method, the second is ontology method and finally, Tree bank method. Each method will be explained in the following sub-subsection.

- Rule Based Method

This method includes identifying the categories of the documents to be summarized and form questions based on these categories such as What happened?, when did it happen?, who got affected ?, what were the consequences? Etc. Then, developing a set of rules base on these questions. verbs and nouns having similar meanings are determined and their positions are correctly identified. The context selection module selects the best candidate amongst these. Generation patterns are then used for the generation of summary sentences [3].

- Ontology Method

In this method, domain ontology is defined by the domain experts. Next phase is document processing phase. Meaningful terms from corpus are produces in this phase. The meaningful terms are classified by the classifier on basis of events of news. Membership degree associated with various events of domain ontology. Membership degree is generated by fuzzy inference [3].

Limitations of this approach are it is time consuming because domain ontology has to be defined by domain experts. Advantage of this approach is it handles uncertain data.

- Tree Based Method

In this approach, the preprocessing is done of similar sentences using shallow parser. After that those sentences mapped to the predicate-argument structure. Different algorithms can be used for selecting the common phrase from the sentences such as Theme algorithm. The phrase conveying the same meaning is selected and also some information added to it and arranged in a particular order. A language generator can be used for making the new summary sentences by combining and arranging the selected common phrase. Use of language generator increases the fluency of the language and reduces the grammatical mistakes. This feature is the main strength of this method [3]. The main problem with this method is that the context of the sentences does not get included while selection of common phrase and it is important part of the sentences even if it is not part of the common phrase.

2) Semantic Based Abstractive Summarization

- Multimodal Semantic Model

Multimodal semantic model captures the concepts and form the relation among these concepts. These selected concepts are expressed in the form of sentences. Multimodal consist of three phases (Semantic Modal, Rated Concepts, sentence Generation).

- Information item based method

In this method, instead of generating abstract from sentences of the input file, it is generated from abstract representation of the input file. The abstract representation is nothing but an information item which is the smallest element of information in a text. The modules of this framework are: Information item retrieval, sentence generation, sentence selection and summary generation [3].

From this method, a short, coherent, information rich and less redundant summary can be formed. In spite of so many advantages, this method has also many limitations. While making grammatical and meaningful sentences, many important information items get rejected. Due to which, linguistic quality of resultant summary gets reduced.

- Semantic Graph Based Methods

The semantic graph approach consists of three phases: The first phase represents input document using rich semantic graph (RSG). In RSG, the verbs and nouns of the input document are represented as graph nodes and the edges correspond to semantic and topological relations between them. The second phase reduces the original graph to a more reduced graph using heuristic rules. The third phase generates an abstractive summery.

The advantage of this method is that it produces less redundant and grammatically correct sentences. The disadvantage of this method is that it is limited to a single document and not multiple documents [3].

3 THE METHODOLOGY OF THE PROPOSED SYSTEM

Our proposed system depends on different important stages. The previous experiences and approaches were inspiring to build our proposed system. The first stage is the **Normalization**, which plays an important role in summarization process, in addition to, another two essential roles, Firstly, splitting long stream contexts into sentences or phrases to facilitate processing. Secondly, modifying the spelling errors in the input texts. The second stage is to **disambiguate** the correct parts of speech of the word forms. After that comes the **shallow parsing stage**, how to parse, connect different constituents' boundaries and build the tree of each sentence. Then it comes to the **Summarization** stage (Also called **Deletion** stage) that decides which phrases are going to be remained or to be excluded. Finally, it is the role of **Generation** to convert the summarized trees to a well-formed generated text. Each stage will be described in details in the following sub-sections:

A. Normalization

Text Normalization has become a common practice in the development of various applications. It helps in saving time, which might be consumed in later stages if not applied. This happens through Preprocessing Phase which is responsible for auto-correcting the raw text and segmenting the Arabic text into sentences [4, 5]. The normalization stage is useful for preparing the input text for later processing by transforming the text to a standard format without dispersed data; later operations are able to work with. Moreover, it plays a significant role in the summarization process in our proposed system. The more the input text is being normalized, the more the parser latterly is ready to parse accurately. Text Normalization stage has three essential steps that will be discussed in the following sub-sub-sections:

1) Extraction: Sometimes the Arabic Texts contains intuitive unnecessary constituents that do not add any addition to the whole meaning, so they are better to be deleted. There is an inventory study was done on a certain data to detect the common constituents that are categorized as unimportant constituents in whatever the surrounding context. This study yielded somehow realistic results when tested on various Arabic texts. The Normalizer extracts those constituents according to the applied rules, which remarked these constituents by punctuations such as comma, semi-colon, parenthesis and full stop. This step helps in the summarization process by excluding such sentences or phrases from the first beginning to get rid of the distraction. These extracted constituents might be words, phrases, or expressions. For example: Parenthetical expressions, such as phrases between brackets or hyphens and Phrases starts with specific Arabic keywords, such as “بالرغم من”, “على حد قوله”, “بسبب”, “مثل”, “التي”, “ذلك” “وبعد”.

2) Modification: Sometimes, the input Arabic texts contains some common spelling mistakes. In this step, the text normalization detects the spelling errors in the text and modify it according to a given rules. The errors can be either a missing character or a wrong character such as:

Missing character can a space or a letter as:

I. عبدالله → عبد الله
II. موسقى → موسيقى

Wrong character can be a wrong written letter such as:

III. الى → إلى

B. Disambiguation

Disambiguation is the process of choosing the right internal diacritization (token) for the word form from the dictionary [4, 5]. It prevents the wrong automatic lexical choices from the dictionary and obtains the right internally diacritized ones. Disambiguation rules decides which nouns appear together, which verbs comes with which nouns, which adjectives describe which nouns, which adverbs characterize which verbs and also which prepositions are correctly used; Therefore disambiguation is the stage of completing the word level which is called Morphology. The morphological analysis of the input text is responsible for analyzing Arabic words and assigning the correct POS and the internal diacritization of words, which is achieved through two processes; tokenization process and disambiguation process. Firstly, tokenization is the process of splitting the natural language input into lexical items; the tokenization process is mainly based on the dictionary, therefore the possibility of ambiguity increases with the increase in the number of entries in the dictionary. Lexical items are tokenized according to longest matched unless the possible longest match is blocked by the developed rules, this stage is explained in details in our previous study in [2]. Secondly, disambiguation is applied over the outcomes of the tokenization process; the disambiguation rules are used to reject the wrong lexical choices and re-obtain the right ones as referred in [4, 5].

As mentioned before disambiguation is concerned with preventing the wrong automatic lexical choices and obtaining the right internally diacritized words. Some linguistic indicators can help in solving the lexical ambiguity, which are morphological, adjacency and linguistic indicators as mentioned in [4, 5]. Morphological indicators: affixation has an important role as the first level of part of speech disambiguation, such as prefixes and suffixes. Adjacency indicators: many qualifiers such as number and gender qualifiers and functional word qualifier could control disambiguating the part of speech. Linguistic indicators: the co-occurrence of specific words with words with specific semantic features is used as an indicator [4, 5].
The disambiguation stage works through several modules; each one has its own important role. To start with, Segmentation or tokenization, which is the first step in disambiguation that segments longest match words, decides the outlines of all parts of speech and rejects some tokens, such as the word "بالمدرسة", this word could be segmented into "بال" verb + "مدرسة" noun or into "ب" preposition + "ال" article + "مدرسة" noun, all these entries exist in the dictionary. The first segmentation "بال" verb + "مدرسة" noun is rejected by the developed rules; because verbs are not adjacent to nouns without a middle space, another segmentation will be detected which is "ب" preposition + "ال" article + "مدرسة" noun, which is the correct one.

Then it comes to Collocation, which is essentially a lexical relation, and not subject to rules but to tendencies [6]. Mitchell (1965) defines collocation as an association of roots or potential lexical meanings rather than actual words; further "a linguistic item or class of items is meaningful not because of inherit properties of its own but because of the contrastive or differential relationships it develops with other items or classes [7]. For example, the sentence "السنة والشيعة", this sentence commonly shows that both words "سنة" sunna + "شيعة" 'shiaa' appear together with a specific meaning. The word "سنة" can be disambiguated as the noun "سَنَة" year or as "سُنَة" sunna, both tokens have the same part of speech, but the only difference between them is the internal diacritic marks that caused variation in meaning. The collocation rules are designed to express either the high possibility of occurrence or the low possibility of co-occurrence of specific words [4, 5]. Therefore, the word "سَنَة" "year" will be rejected in this context, while the word "سُنَة" "sunna" will be selected to match the meaning.
Moreover, there are rules, which are responsible for disambiguating prepositions, adjectives, nouns, etc., if they are wrongly tokenized, according to the surrounding context. Finally, the most important role in the disambiguation stage, after accurately completing all the previous modules, is detecting the agreement between nouns and adjectives and the agreement between verbs and nouns [4, 5]. By this final step, the sentence or the phrase has correctly been disambiguated and the input text is ready for being syntactically parsed.

C. Shallow Parsing

Partial or shallow parsing - the task of recovering only a limited amount of syntactic information from natural language sentences - has proved to be a useful technology for written and spoken language domains [8]. For example, within the Verbmobil project, shallow parsers were used to add robustness to a large speech-to-speech translation system [9]. Shallow parsers are also typically used to reduce the search space for full-blown, `deep' parsers [10]. Yet another application of shallow parsing is question-answering on the World Wide Web, where there is a need to efficiently process large quantities of (potentially) ill-formed documents [11, 12]. Moreover, more generally all text mining applications, e.g. in biology [13].

Abney (1991) is credited with being the first to argue for the relevance of shallow parsing, both from the point of view of psycholinguistic evidence and from the point of view of practical applications. His own approach used hand-crafted cascaded Finite State Transducers to get at a shallow parse [14]. Typical modules within a shallow parser architecture include the following [8]:

1) Part-of-Speech Tagging: given a word and its context, decide what the correct morphosyntactic class of that word is (noun, verb, etc.). POS tagging is a well-understood problem in NLP, to which machine learning approaches are routinely applied.

2) Chunking: given the words and their morphosyntactic class, decide which words can be grouped as chunks (noun phrases, verb phrases, complete clauses, etc.)

3) Relation detecting: given the chunks in a sentence, decide which relations they have with the main verb (subject, object, location, etc.).

Shallow parsing is a challenging domain for machine learning research. Note that shallow parsing does not refer to a single technique. Instead, it is better to consider it to refer to a family of related methods, all of which attempt to recover some syntactic information, at the possible expense of ignoring all other such information [8]. Building shallow parsers is therefore a labor-intensive task. Unsurprisingly, shallow parsers are usually automatically built, using techniques originating within the machine learning (or statistical) community [8].
Our Automated Shallow Parser adopts the "**X-bar theory**", the theory of syntactic category formation. The X-bar theory was chosen because of its compatibility with Arabic language. The goal of using X-bar theory will be to supply a systematic description of Standard Arabic sentence formation. There are many advantages of adopting X-bar format in lieu of the phrase structure rules. The latter suffer a severe failing. Their role seems largely redundant as they simply duplicate information included in the lexical entries of the lexical categories [15]. On the other hand, the consistency of X-bar theory towards all the phrasal categories is revealed. The X-bar format permits to bring out what is common to the different types of phrases. Another significant property of X-bar theory is that it throws light on the hierarchical organization of the phrase instead of the linear order of the constituents, which is intuitively felt to be wrong. Furthermore, the X-bar schema can be extended to embrace the constituents of the clause as a whole [15].

The proposed Parser starts to identify different types of constituents to fulfill the **sentence level**. In the beginning, **Adjectives** such as "جميلة" is projected to minimal projection "JB", and then if there are any complements exists, it will be combined within this "JB" to create a bigger "JB". Finally, it will be projected to maximal projection "JP" anyways, but after confirming the presence of a specifier such as the Determiner "ال" "DP", as shown in figure 1 for the phrase "الجميلة":

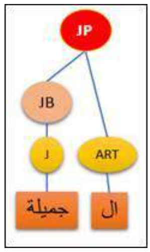


Fig.1. The tree diagram of Adjective phrase in X-bar representation.

For the **nouns**, all nouns is projected to minimal projection "NB", but if it is a **Proper Noun** such as "محمد", it will be projected directly to its maximal projection "NP" if it's not followed by any adjunct related to it as in figure 2. However, if it is a **Common Noun** such as "كتاب", it will be combined with any following *complements* if exists which is necessary to complete the meaning, in order to form "NB", such as "edafa", as in"كتاب محمد" as shown in figure 3, or "subcategorization frame" as in "الاعتماد على". After that nouns combine with *Adjuncts* -which is not syntactically required-, such as the adjective "جميل" in "كتاب جميل" as in figure 4, or "في الفصل" in "الكتاب في الفصل", therefore, "NB" will be formed. If there are no more surrounding complements or adjuncts, "NB" will be projected to maximal projection "NP" if specifiers are found such as the Determiner "ال" as in figure 5.

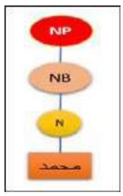


Fig.2. The tree diagram of Proper Noun in X-bar representation.

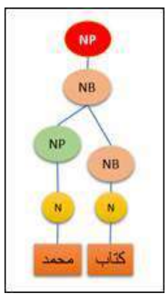


Fig.3. The tree diagram of Noun Phrase "edafa".

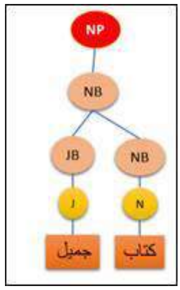


Fig.4. The tree diagram of Noun- Adjective Phrase.

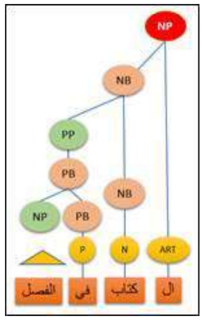


Fig.5. The tree diagram of a final projected Noun Phrase "الكتاب في الفصل".

Then the parser identifies the **Prepositions** such as "إلى", "على", "في", etc…, the preposition is projected to be a preposition phrase bar "PB". Then it might combined with the complement such as "noun phrase" "المدرسة" in "إلى المدرسة" or a specifier such as "adverbial phrase" " قريبا فوق" in "الكتاب تقريبا فوق المكتب" to reach the maximal projection "PP" as in figure 6.

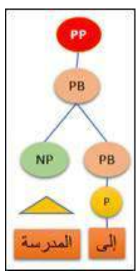


Fig.6. The tree diagram of maximal Prepositional Phrase in X-bar representation.

Moreover, **Adverbs** -either indicates time or manner- might have complements that completes the meaning such as "مقارنَةً" which needs a prepositional phrase such as "بالماضي" to be projected to the maximal projection of the adverbial phrase "AP" as in figure 7. On the other hand, there are adverbs that completes the meaning directly on its own such as "اليوم", "أمس", "مبتسمًا"; therefore, they will be projected directly to the maximal projection "AP" as shown in figure 8.

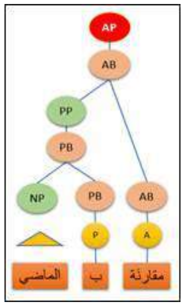


Fig.7. The tree diagram of maximal projection of Adverbial Phrase.

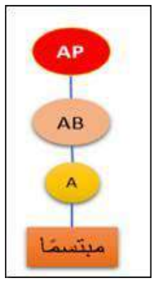


Fig.8. The tree diagram of directly projected Adverbial Phrase.

For the **VERBs**, the verb is projected to be a "VB". Complements of the verb differ according to the verb's transitivity; it may have no complements or have one or more complements. The parser plays an effective role in recognizing the verb's complements. For example, in the phrase "ذهب الولد", the "NP" "الولد" is predicted to be the subject of the verb as presented in figure 9. While the object of the verb can occur in the form of noun phrases, prepositional phrases or verbal phrases, for example, the complement occurred as "NP" "التفاحة" in the phrase "أكل محمد التفاحة" as presented in figure 9, the complement occurred as "PP" "إلى المدرسة" as in "ذهب إلى المدرسة" as presented in figure 10, and finally the complement occurred as "VP" in the phrase "بدأت الأم تصنع الطعام".Verbs can also have adjuncts, for example, the adverbial phrase "بسرعة" in "ذهب بسرعة".

During the analysis of the verb phrase, verb bars "VB" are directly formed, if the noun phrase "NP" precedes the verb the VB will be maximal projected to be "VP" immediately. In addition, there is another extra projection happens, only if the verb is an auxiliary verb, consequently, "VP" is projected to be an inflectional phrase "IP". Figure 11 shows the hierarchy of a fully analyzed verb phrase while connecting the tree diagram of the arguments of the verb phrase.

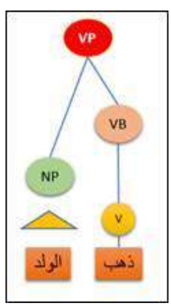


Fig.9. The tree diagram of verb phrase in the presence of the subject.

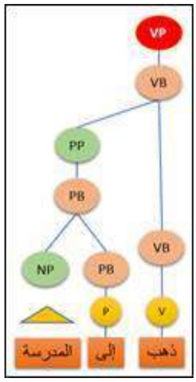


Fig.10. The tree diagram of verb phrase with complement.

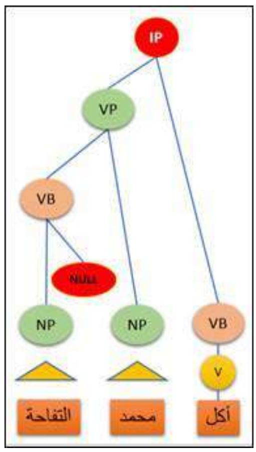


Fig.11. The tree diagram of a complete verbal phrase.

For more illustration, the sentence in (1) is an example to show how the parser works and figure 12 shows the tree diagram for it:

Sentence (1):

"واليوم انضم سيرف إلى وكالة ناسا الفضائية للعمل على مشروع إطلاق شبكة إنترنت في الفضاء الخارجي"

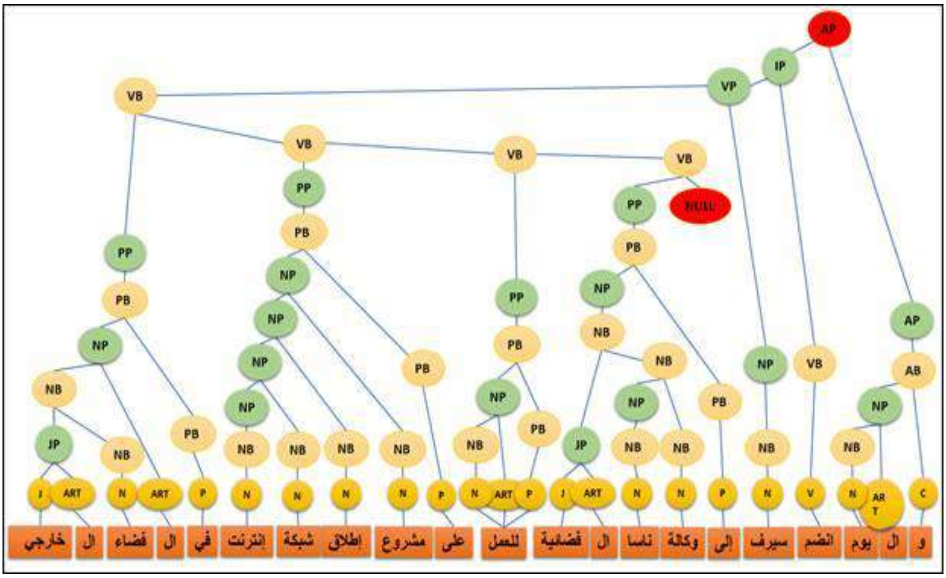


Fig.13. Tree diagram of sentence 1.

D. Deletion

After the parser completes forming the trees' diagram for each sentence in the document. Only one thing left now to start summarizing is the deletion of the insignificant constituents (does not share the meaning of the sentences). Automatic text summarization, aims to reduce the size of a text while preserving its information content. A summarizer is a system that produces a condensed representation of its input's for user consumption [3]. Deletion or omission stage happens by dispensing with certain un-important phrases or clauses in the inputted sentence or text. It is a crucial stage preparing the output to the generation stage.

As mentioned before there is an extraction has done in the normalization stage for some sentences before parsing stage. In addition, after parsing we investigate un-necessary phrases in order to be deleted. There are two types of phrases in the text, the **essential phrases** and the **optional phrases** [2]. The essential phrases are irreplaceable and convey the important information, while the optional phrases have no remarkable meaning and could be omitted such as prepositional phrases, adverb phrases, relative clauses and sometimes complementizer phrase CP.

For the PP, the deletion of it depends on their role. If the PP plays the role as an argument of a verb or an adjective or a noun (object or goal), in this case, the PP is very important and it will not be essential and could not deleted. Otherwise, it can be deleted if it is not add. Sometimes, adverb phrases have no impact in the meaning either indicating time or manner. For the relative clauses, it is better to be omitted because they are treated as the parenthetical expressions that are omitted in the normalization stage.

When Deletion was tested on the above mentioned sentence in (1), only essential phrases were kept during the summarization, while the optional ones were omitted as shown in sentence in (2). Figure (14) shows the omitted phrases and the kept ones for the sentence in (1) during the summarization process.

Sentence (2):

”واليوم انضم سيرف إلى وكالة ناسا الفضائية للعمل على مشروع إطلاق شبكة إنترنت في الفضاء الخارجي“

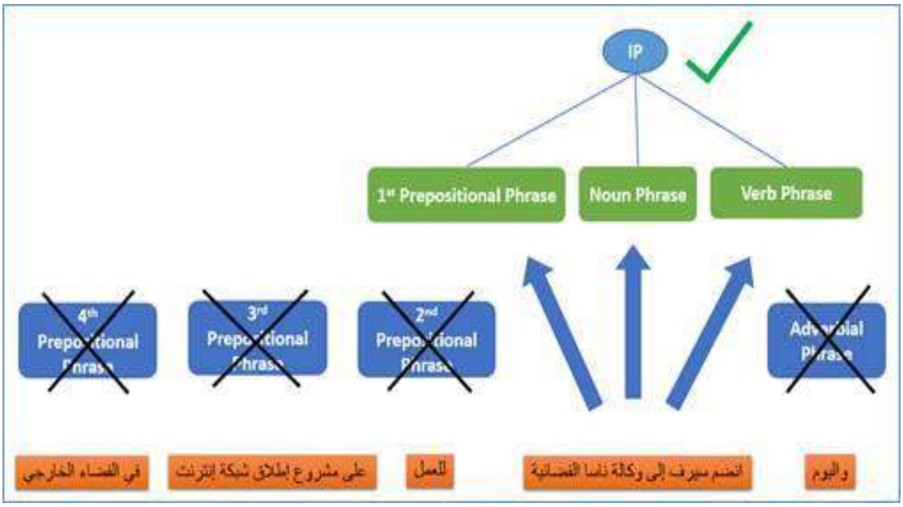


Fig.14. The deletion Process for sentence (1).

E. Generation

In the generation stage, there are a bund of settled rules built for handling the summarized trees after analysis in order to generate and list a well-formed Arabic summarized text. Sentence in (3) becomes the final generated condensed representation of the original sentence in (1). Sentence in (3) is the generated Arabic text of the sentence represented in figure 15:

Sentence (3):

"انضم سيرف إلى وكالة ناسا الفضائية"

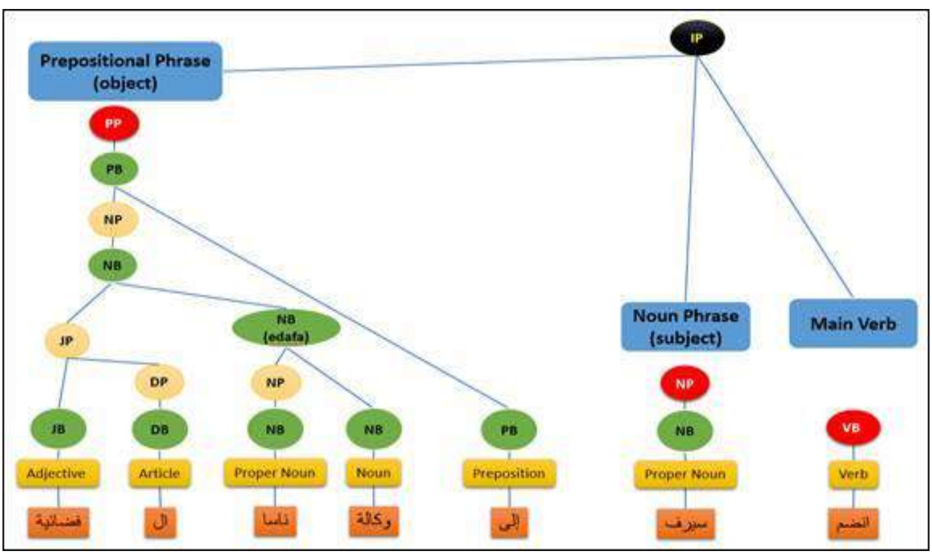


Fig.15. The tree diagram of sentence in (3).

One hundred new contexts were chosen to be automatically summarized. This data were classified into five stages (zero – one – two – three – four – five) according to their complexity. There are many factors that control this complexity, which represent in the type of the text, the sentence length, multiple boundaries, and the number of topic words the sentence may contain. The automatic summarizer in this stage works on the zero, one and two stages according to the established rules, the achieved studies and the results are surprising. The third and the fourth stages are not finished yet because both consist of somehow more sophisticated structures. The most advanced stage is the fifth stage, where all limitations of summarization such as the extra sentence length and the increasing multiple boundaries are encountered in. The proposed automatic system is still working on these difficult issues in order to handle the further work.

Moreover, an Arabic document with the title "الفضاء" was tested to be summarized as shown in figure 16. When applied to the summarizer, essential and optional constituents were detected. The developed rules were applied to segment and modify constituents, select adequate parts of speech, build trees, identifying boundaries and omitting the phrases that does not share meaning of the sentence. The most important stage in Summarization process is the Deletion as mentioned above.

بالنسبة لكوكب الأرض فإن الفضاء الخارجي هو المنطقة الواقعة على بعد 100 كيلومتر عاموديا على سطح الأرض، والتي
سيكون التنفس فيها صعبا جدا لقلة الأكسجين، حيث يكون الغلاف الجوي للأرض قد اختفى تقريبا . هنالك ثلاثة أقسام أساسية من
العلم تهتم بدراسة الفضاء أو السماء، وهي : علم الفلك، وعلم الفيزياء الفلكية، وعلم الكونيات، ولكل واحد منها مجاله المميز،
والذي يهتم بجانب معين من الكون . في هذا المقال سنتحدث عن بعض المعلومات العامة جدا عن الفضاء- وخاصة الأجسام
السماوية الأكثر شهرة - وبعض الأمور الغامضة في هذا الكون. النجوم هي كرات غازية مضيئة، وتتكون بشكل أساسي من
الهيدروجين والهيليوم) أخف عنصرين في الجدول الدوري(، وتحدث فيها اندماجات نووية تولد طاقة وعناصر جديدة مثل
الكربون، والأكسجين، والمغنيسيوم، ... ، وحتى الحديد، لكن الاندماجات النووية الحاصلة داخل النجوم لا تمتلك الطاقة الكافية
لدمج عنصر الحديد وتشكيل عناصر أثقل منها ، فالحديد هو أثقل عنصر يتم تشكيله داخل النجوم، لكن العناصر الأخرى الأثقل
منه يتم تشكيلها في انفجار المستعر الأعظم) بالإنجليزية(Super Nova : ،كالذهب مثلاً ، لذلك لو كان في يدكِ خاتمٌ من الذهب
فيجب أن تعرفي قيمة هذا الخاتم الحقيقية حيث تمّت صناعته في واحدٍ من أعظم مصانع الكون . ومن الأمثلةعلى النجوم وهو
أقرب نجم لنا الشمس . و تتكون الشمس من نفس العناصر الموجودة في الأرض ولكن بنسب مختلفة ؛ حيث إننا نعرف ذلك عن
طريق تحليلنا لطيف الشمس ، الأمر الذي أكدلنا أن الشمس، والأرض، وجميع كواكب المجموعة الشمسية قد تشكلوا من نفس
الغيمة السديمية، ونسب المواد الموجودة في الشمس هي كالآتي :71 %هيدروجين، 27 %هيليوم، 2 %عناصر أثقل.

Fig.16. Original Document.

The summarization of the document in figure (16) will be achieved after going through different stages: the deletion in normalization stage according to the surrounding punctuations as follows:

- Relative Clauses

The clause "والتي سيكون التنفس فيها صعب جدا لقلة الأكسجين" reveals an extra un-important meaning.

- Key-worded expressions

There are constituents that are remarked by some key words that are considered in the settled bund of applied rules; those constituents are deleted whenever they appear in the context. Phrases such as: " وهي: علم الفلك، وعلم الفيزياء الفلكية، وعلم الكونيات، ولكل واحد منها مجاله المميز، والذي يهتم بجانب معين من الكون", " حيث إننا نعرف ذلك عن طريق تحليلنا لطيف الشمس", " لكن الاندماجات النووية الحاصلة داخل النجوم لا تمتلك الطاقة الكافية لدمج عنصر الحديد وتشكيل عناصر أثقل منها", " في هذا المقال سنتحدث عن بعض المعلومات العامة جدا عن الفضاء -وخاصة الأجسام السماوية الأكثر شهرة- وبعض الأمور الغامضة في هذا الكون" match this case of deletion.

- Parenthesis constituents

Constituents occur between parenthesis are deleted, such as "(أخف عنصرين في الجدول الدوري)".

Then, after parsing each sentence in the document some deletion decision are made such as the following:

a) Adverbial phrases

The word "عموديا" has no effective meaning, so it will be omitted.

b) Prepositional phrases

The PP “على سطح الأرض”, “من العلم”, “لنا” and “داخل النجوم” are deleted because they are not considered as essential arguments.

Figure 17 shows the final generated output after deletion. In figure 16, the original document counts 240 words, while in figure 17 the automated summarized text counts 91 words.

بالنسبة لكوكب الأرض فإن الفضاء الخارجي هو المنطقة الواقعة على بعد 100 كيلومتر. هنالك ثلاثة أقسام أساسية تهتم بدراسة الفضاء أو السماء. النجوم هي كرات غازية مضيئة، وتتكون بشكل أساسي من الهيدروجين والهيليوم، وتحدث فيها اندماجات نووية تولد طاقة وعناصر جديدة، فالحديد هو أثقل عنصر يتم تشكيله. ومن الأمثلة وهو أقرب نجم الشمس. وتتكون الشمس من نفس العناصر الموجودة في الأرض، الأمر الذي أكد أن الشمس، والأرض، وجميع كواكب المجموعة الشمسية قد تشكلوا من نفس الغيمة السديمية، ونسب المواد الموجودة في الشمس هي كالآتي:
71% هيدروجين، 27% هيليوم، 2% عناصر أثقل.

Fig.17. Automated Summarized Document.

4 CONCLUSION

As there are no summarization systems that directly depend on syntactic analysis which helps in understanding the meaning of the sentences, we are trying to build a summarization system that attempts to understand and identify the essential arguments and the optional ones in the sentences. we have developed a shallow parser as a try to add a new method that helps in automatic summarization, this method is considered promising as it tries to overcome the limitation faced the existing summarization systems that are based on statistical methods. Some examples of the existing summarization systems have been mentioned. Our rule based summarization system is presented.

REFERENCES

[1] J Larocca Neto and A. Freitas and A. A. Kaestner “Automatic Text Summarization using a Machine Learning Approach”, Pontifical Catholic University of Parana (PUCPR) Rua Imaculada Conceicao, 1155.

[2] S Alansary. (2018) “Automatic Arabic Text Summarization: A pilot study”, Bibliotheca Alexandrina, Phonetics and Linguistics Department, Faculty of Arts, Alexandria University - Alexandria, Egypt.

[3] Kasture, N. R., Yargal, N., Singh, N. N., Kulkarni, N., & Mathur, V. (2014). A survey on methods of abstractive text summarization. Int. J. Res. Merg. Sci. Technol, 1(6), 53-57.

[4] S. Alansary, “Alserag: An Automatic Diacritization System for Arabic”. The 2nd International Conference on Advanced Intelligent Systems and Informatics (AISI’16), Cairo, Egypt, 2016.

[5] S. Alansary, (2016, December). Improving Alserag Arabic Diacritization Grammar through Syntactic Analysis. In 16th international conference on language engineering, Cairo, Egypt.2016.

[6] K. H. Nofal, “Collocations in English and Arabic: A comparative study”. English Language and Literature Studies, Department of English, Language Centre Philadelphia University, Jordan, 2012.

[7] Mitchell, T.F., 1965. Linguistic going–on: Collocations and other lexical matters arising on the syntagamatic record. In J. R. Firth, Eds. Principles of firthian linguistics. London: Longman.

[8] J hammerton, M Osborne, S Armstrong and W Daelemans “Introduction to Special Issue on Machine Learning Approaches to Shallow Parsing, journal of Machine Learning Research 2 (2002) 551-558.

[9] Wolfgang Wahlster, editor. Verbmobil: Foundations of Speech-to Speech Translation. Springer, 2000.

[10] Michael John Collins. A new statistical parser based on bigram lexical dependencies. In 34th Annual Meeting of the Association for Computational Linguistics. University of California, Santa Cruz, California, USA, June 1996.

[11] Sabine Buchholz and Walter Daelemans. Complex Answers: A Case Study using a WWW Question Answering System. Natural Language Engineering, 2001.

[12] R. Srihari and W. Li. Information extraction supported question answering. In Proceedings of TREC 8, 1999.

[13] T. Sekimizu, H. Park, and J. Tsujii. Identifying the interaction between genes and gene products based on frequently seen verbs in medline abstracts, 1998.

[14] S. Abney. Parsing by chunks. In Principle-Based Parsing, pages 257{278. Kluwer Academic Publishers, Dordrecht, 1991.

[15] Dalarna “X-bar Theory and Standard Arabic” Fall Term, 2006\7 Said Tamadla.

BIOGRAPHY

Dr. Sameh Alansary: *Director of Arabic Computational Linguistic Center at Bibliotheca Alexandrina, Alexandria, Egypt.*

He is professor of computational linguistics in the Department of Phonetics and Linguistics and the head of Phonetics and Linguistics Department, Faculty of Arts, Alexandria University. He obtained his MA in Building Arabic Lexical Databases in 1996, and his PhD from Nijmegen University, the Netherlands in building a formal grammar for parsing Arabic structures in 2002. His main areas of interest are concerned with corpus work, morphological analysis and generation, and building formal grammars.

He is also the head of Arabic Computational Linguistics Center in Bibliotheca Alexandrina. He is supervising and managing the Universal Networking Language project in Library of Alexandria since 1-6-2005 till now.

Dr. Alansary is the co-founder of the Arabic Language Technology Center (ALTEC), an NGO aims at providing Arabic Language resources and building a road map for Arabic Language Technology in Egypt and in the Middle East. He has many scientific works in Arabic Natural Language Processing published in international conferences and periodicals, and a member in many scientific organizations: (1) Egyptian Society of Language Engineering, Cairo, (2) Arabic Linguistic Society - USA, (3) Association of Computational Linguistics - USA – Europe, (4) Universal Networking Language foundation, United Nations, Geneva, Switzerland.

TRANSLATED ABSTRACT

التحليل النحوي السطحي من أجل التلخيص الآلي للنصوص العربية

سامح الأنصاري

مكتبة الإسكندرية، الشاطبي، الإسكندرية، مصر
قسم الصوتيات واللسانيات، كلية الآداب، جامعة الإسكندرية، الشاطبي، الإسكندرية، مصر
sameh.alansary@bibalex.org

ملخص:

في هذه الورقة، نقترح طريقة تلخيص من وجهة نظر التحليل النحوي السطحي ويتضمن نظام التلخيص مراحل مختلفة: مرحلة استخراج الجمل، مرحلة التحليل النحوي، ومرحلة التوليد .وأخيرا، ستناقش عملية تلخيص التقييم.

الكلمات المفتاحية: التلخيص الآلي (TS)، التحليل الآلي للنص، أجزاء الكلمات، التحليل السطحي والتلخيص التجريدي.

L'Application du Formalisme des Fonctions Lexicales sur la Langue Arabe

Racha MOHAMMAD SALEM
Département de Langue et de Littérature Françaises, Faculté des Lettres, Université d'Alexandrie
salem.racha@live.com

Résumé : ***Les recherches qui traitent de la formalisation de la langue arabe ont commencé dès les années quatre-vingt-dix du siècle précédent. Les linguistes arabes modernes cherchent à créer un système formel rigoureux qui soit inspiré des systèmes de formalisation des langues européennes et qui soit apte à gérer la langue arabe avec toutes ses richesses. Au cours des pages suivantes, nous proposerons une méthode de formalisation de l'arabe selon le système des fonctions lexicales (système d'encodage de la théorie Sens-Texte). Ce dernier sera appliqué sur trois phénomènes linguistiques : le cliché, l'hyperbole et le diminutif.***

Mots-clés : ***Fonctions lexicales, linguistique arabe, mathématisation de langue, théorie Sens-Texte, traitement automatique***
de langue naturelle.

1 INTRODUCTION

Le mouvement de mathématisation des langues naturelles commençait dans les années trente avec la tendance de rendre la machine capable de gérer les langues humaines (1). « Mathématiser » consiste à représenter les règles linguistiques sous forme de règles mathématiques. Mais la question à laquelle les créateurs des systèmes de mathématisation devraient penser est la suivante : une formule mathématique est-elle censée refléter le signifiant du signe linguistique ou son signifié ? Autrement dit, est-ce qu'il s'agit de mathématiser la forme ou le fond ? Même si la réponse paraît claire et indiscutable, la réalité est que la plupart des langages de formalisation n'ont pas réussi à exprimer le contenu linguistique des signes dont une langue naturelle se compose. La preuve est que jusqu'à nos jours les applications du TALN (traduction automatique, analyse automatique de texte, etc.) ne sont pas très performantes. De nouveaux défis qui sont intraitables par les systèmes d'encodage surgissent continuellement.

En comparaison avec le mouvement de mathématisation des langues européennes, la mathématisation de la langue arabe s'est attardée vu la complexité de sa linguistique. Dans le présent article, nous présenterons un formalisme qui était appliqué aux langues anglaise et française et dont les créateurs ont pu démontrer sa capacité de déchiffrer les secrets des langues humaines. Il s'agit du formalisme des fonctions lexicales (FL) représentant le système d'encodage de la théorie Sens-Texte (TST). Son atout est qu'il garantit l'équilibre entre les deux niveaux : sémantique et formel. Dans la première partie de l'article, nous expliquerons en détail le formalisme des FL, puis nous l'appliquerons à trois phénomènes linguistiques arabes (le cliché, l'hyperbole et le diminutif) dans le but de démontrer sa capacité de comprendre les différents aspects de la langue arabe.

Cette recherche s'ajoute aux études qui ont été commencées au sein du laboratoire ATILF (Université de Lorraine, France) (2) en 2014 et qui concernent la création du Réseau Lexical de l'Arabe (RL-ar) à l'image du Réseau Lexical du Français (RL-fr) (3). Il est à noter que les réseaux lexicaux représentent la nouvelle génération des dictionnaires électroniques. La sémantique lexicale constitue le cadre linguistique selon lequel un réseau lexical est créé. En plus de la définition du lexème (**L**), l'article lexicographique d'un RL énumère les relations paradigmatiques et syntagmatiques liant (**L**) aux autres lexèmes ainsi que la représentation formelle de chacune de ces relations.

2 LA MATHÉMATISATION DES LANGUES NATURELLES

A. Deux Tendances Majeures

La mathématisation des langues naturelles a commencé avec le besoin de faire comprendre à la machine la langue humaine pour qu'elle puisse assumer les tâches exécutées par l'homme – surtout la traduction automatique – dans un laps de temps court et avec des prix réduits. Effectivement, il existe deux tendances ou deux méthodes de représentation formelle des langues humaines: la logico-mathématisation et l'automatisation-mathématique. La première qui est apparue dans les années trente se penche plutôt vers la mathématique que la linguistique. Cependant, la deuxième qui est surgie pendant les années cinquante afin de combler les lacunes de la première tendance, conserve l'équilibre entre les mathématiques et la linguistique.

En général, le processus d'encodage ne consiste pas seulement à la transformation des lexèmes en symboles ou en équations mathématiques, car cette méthode prive l'unité lexicale d'une partie importante de son sens : A l'encontre de la conception saussurienne, les théories contemporaines définissent le signe linguistique par le rapport liant entre le signifiant, le signifié et le syntactique (4). Ce triplet souligne que l'accès au sens complet d'un lexème s'effectue sur deux niveaux : A partir de son signifié ou son sens propre et de son syntactique ou son sens contextuel (5). Ce qui veut dire qu'un système de mathématisation fiable doit prendre en considération la combinatoire lexico-syntaxique du lexème afin de surmonter les problèmes linguistiques résultant de la pluralité de sens, à titre d'exemple. [Racha SALEM, 2017]

C'est pour cette raison que l'automatisation-mathématique paraît la plus logique pour les linguistes contemporains. Cette dernière prend son élan avec Noam Chomsky – le père de la grammaire générative – qui souligne que la formalisation d'une langue naturelle doit passer par l'analyse linguistique. Car il s'agit d'une formalisation de sens plutôt que de forme.

B. Une Grammaire Particulière

Pour aboutir à ses fins, Chomsky fonde la grammaire générative. A travers un nombre de règles dites « de réécriture », il analyse la structure syntaxique de la phrase pour dégager son sens profond, puis il formalise cette structure. La phrase « la fille mange les bonbons », par exemple, est représentée en :

1) S → NP – VP
2) NP → Det – N
3) VP → V – NP
4) NP → Det – N

La première ligne représente la structure générale de la phrase : le (NP) est le groupe nominal et le (VP) est le groupe verbal. La deuxième ligne détaille la structure syntaxique du groupe nominal alors que la troisième détaille celle du groupe verbal. La dernière ligne représente la structure syntaxique du groupe nominal qui est créé au sein du groupe verbal.

Donc, Chomsky lance la première tentative de formalisation des liens lexico-syntaxiques unissant les lexèmes à l'intérieur d'une phrase. Mais la pratique a montré que l'analyse syntaxique est insuffisante surtout en cas d'ambiguïté. C'est ce qui a poussé les linguistes à se recourir à la sémantique et à la déterminer comme étape introductive à l'étude syntaxique de la phrase. Le mathématicien et le linguiste russe Igor Mel'čuk et ses collègues ont travaillé sur l'importance de la sémantique dans le déchiffrement du sens contextuel.

3 UN MODÈLE D'ANALYSE LINGUISTIQUE MULTISTRATAL

Comme nous l'avons souligné au début, la TST diffère des autres théories de la linguistique informatique parce qu'elle garde l'équilibre entre l'étude linguistique de la langue naturelle et sa formalisation. A l'encontre de Chomsky, Mel'čuk précise qu'un système d'encodage fiable prend en considération les liens sémantico-syntaxiques liant le lexème aux différentes composantes de son réseau sémantique. C'est pourquoi le point de départ de la TST est l'analyse sémantique. Celle-ci est secondée par l'analyse syntaxique, puis morphologique et enfin phonologique du lexème. A l'issu de ce modèle, un aperçu complet de chaque lexème se tisse où toutes ses nuances de sens et ses dérivations syntaxiques seront déterminées. Par conséquent, le lexème ne sera pas codé par un seul symbole, mais chaque constituant de cet aperçu en procura un.

4 UN FORMALISME RIGOUREUX

Pour couvrir toutes les nuances de sens et les dérivations syntaxiques, le formalisme de la TST étudie les liens paradigmatiques [(6)] et syntagmatiques [(7)] qui font connecter les unités lexicales d'une langue naturelle. Les fondateurs de la théorie commencent leur projet par l'élaboration d'un inventaire de règles linguistiques qu'ils appellent « le système de paraphrasage » et qui groupe toutes les relations sémantiques et syntaxiques universelles. A travers une quarantaine de règles, le système détermine pour chaque lexème tous les équivalents ainsi que les dérivés qui pourraient le remplacer dans les différents contextes tout en conservant le sens principal. Ces règles qui sont formalisées par la suite sous forme de fonction mathématique : [**F** (x) = y] où (x) représente l'argument de la fonction et (y) est sa valeur.

Le formalisme des FL est réparti en deux rubriques principales dont la première groupe les FL paradigmatiques telles que la synonymie, l'antonymie, la conversion, la métaphore, la nominalisation, la verbalisation, le nom typique de l'actant et le nom du point d'origine.

Syn (أنار) = أضاء

Anti (قَبِلَ) = رفض

Conv (باع ل) = اشترى من

Figur (دخان) = ساتر من ~

S₀ (سافر) = سفر

V₀ (حديث) = تحدث

Sᵢ (تحدث) = متحدث

Germ (غضب) = شرارة ال~

La deuxième rubrique est celle des fonctions syntagmatiques, nous citons à titre d'exemple : l'intensification, le laudatif et les verbes supports.

Magn (صورة) = ضخمة

Bon (اختيار) = موفق

Oper (أمر) = أعطى

Real (طائرة) = قاد

Parfois, la même FL est divisée en sous-catégories afin de gérer les micros nuances de sens au sein de la même relation linguistique. Prenons l'exemple de la synonymie qui est divisée en synonymie exacte (الترادف الكامل) et approximative (شبه الترادف).

Syn (حزن) ≡ غم (synonymie exacte)

Syn (أشرق) ≅ سطع (synonymie approximative)

A son tour, la synonymie approximative comporte trois sous-catégories comme les exemples suivants le montrent :

F (x) = y

- Synonyme moins spécifique : lorsque le sens de (y) est inclus dans (x)

Syn ⊏ (ديك) = طائر

- Synonyme plus spécifique : lorsque le sens de (y) inclut celui de (x)

Syn ⊃ (قتل) = اغتيال

- Synonyme à intersection : lorsque le sens de (y) et celui de (x) partagent certains traits lexicaux.

Syn ∩ (يلعب) = يمزح

Le formalisme des FL offre d'autres options pour couvrir les unités lexicales ayant un sens composé : Premièrement, la fonction lexicale composée. Elle comporte deux FL qui se combinent facilement sur le plan syntaxique, telles les fonctions **(Anti)** qui dénote « l'opposition » et **(Magn)** qui dénote « l'intensification ». La FL composée **(AntiMagn)** dénote « la diminution / l'atténuation ».

AntiMagn (حرارة) = منخفضة

Deuxièmement, la configuration de fonctions lexicales. Il s'agit d'une unité lexicale qui se compose de plusieurs sémantèmes dont chacun est modélisé par une FL. Ce groupe de FL est réuni dans une configuration. Prenons l'exemple de la collocation : « انفجر بالبكاء » ou « fondre en larmes ». Le verbe « انفجر » dénote à la fois « بدأ في البكاء » (commencer à) et « بكى كثيرا » (pleurer beaucoup). La collocation est donc modélisée par une chaîne de FL :

[**IncepReal** $_1$ (بكى) + **Magn** (بكى)] [8]

Le premier sens est formalisé par une FL verbale composée et le deuxième par une FL d'intensification. La fonction **(Incep)** exprime le sens de « بدأ / commencer » et la fonction **(Real)** exprime le sens de « أنجز / produire / exécuter ». Ainsi, **(IncepReal** $_1$**)** modélise le sens « بدأ في البكاء / commencer à pleurer ».

Troisièmement, la valeur fusionnée. C'est le cas d'un lexème qui exprime le même sens que celui résultant de la combinatoire de deux lexèmes. Prenons l'exemple du lexème « غفا » qui exprime « être au début du sommeil », c'est un sommeil léger. Donc, « غفا » reflète le sens exprimé par la collocation « بدأ في النوم / commencer à dormir ». La valeur fusionnée est marquée par le symbole « // ».

Incep (نام) = // غفا

Reste à mentionner que lors de la création du système des fonctions lexicales, Igor Mel'čuk et ses collègues l'ont voulu un système universel qui s'adapte à toutes les langues humaines. C'est pour cette raison qu'ils ont distingué les FL standard des FL non-standard. Les premières modélisent les relations linguistiques communes à toutes les langues alors que les deuxièmes formalisent les relations spéciales de chaque langue. Les FL standard sont invariables tandis que les FL non-standard sont variables.

5 TROIS PHÉNOMÈNES LINGUISTIQUES FACE AUX FONCTIONS LEXICALES

Dans la dernière partie du présent article, nous appliquons le formalisme des FL sur trois phénomènes linguistiques arabes. Ces derniers ont été sélectionnés précisément car ils diffèrent de leurs équivalents en langue française.

A. Les Clichés Linguistiques (التعبيرات المسكوكة)

Un cliché linguistique est un syntagme non libre de type pragmatique qui est dit dans une situation communicative déterminée. Dans les dictionnaires traditionnels, le cliché est cité sous l'entrée qui représente une de ses composantes (souvent celle portant le contenu sémantique principal du cliché). Examinons cet exemple : le cliché « بالرفاء و البنين » est cité sous l'entrée de la racine [9] « رفأ » qui exprime le sens « السكينة » [10]. Ce cliché se dit dans le contexte du mariage pour féliciter les nouveaux mariés et leur souhaiter une vie tranquille en espérant qu'ils auront d'enfants très bientôt.

Dans un réseau lexical, le même cliché est représenté de deux manières différentes. Il fait partie intégrante des champs sémantiques de ses composantes. Etant donné que chaque composante exprime une partie du sens du cliché, donc, il doit apparaître dans leurs champs sémantiques. De l'autre côté, en général, le sens d'un cliché est exprimé par un autre lexème, alors celui-là est mentionné dans le champ sémantique de celui-ci comme étant son synonyme. Reprenons le cliché « بالرفاء و البنين », il se compose de deux lexèmes : « رفاء » et « بنين ». De même, il est le synonyme du lexème « مبروك / félicitations». Ce qui veut dire que linguistiquement il sera mentionné trois fois. Formellement, il sera mentionné une seule fois sous « مبروك » qui est son synonyme.

Syn « مبروك » = « بالرفاء و البنين »

Un autre cliché qui est fréquemment employé dans la presse sportive : « الساحرة المستديرة » ou « la charmante ronde » selon la traduction littérale. C'est une métaphore qui désigne l'amour excessif des gens pour le football. Le sens de ce cliché est composé vu qu'il est constitué du lexème « ساحرة » exprimant l'admiration de beaucoup de gens pour le football et le lexème « مستديرة » désignant la forme ronde du ballon. Dans le réseau lexical, le cliché
« الساحرة المستديرة » sera-t-il lié aux champs sémantiques des deux lexèmes dont il se compose ? En fait, ce cliché diffère du précédent, c'est un cas spécial puisqu'il reflète un sens figuré et non un sens propre. C'est alors qu'il sera plutôt mentionné dans le champ sémantique du lexème « football ». La FL qui est utilisée en cas des sens figurés est **(Figur)**.

Figur (كرة القدم) = الساحرة المستديرة

B. L'Hyperbole (صيغة المبالغة)

Parmi les formules d'intensification arabes, nous avons sélectionné celles qui sont utilisées pour caractériser les humains. La linguistique arabe précise cinq schèmes de ce type d'hyperbole :

الوزن	مثال
فعّال	فتّاح / سفّاح
مفعال	مقدام
فعول	شكور
فعيل	سميع / عليم
فَعِل	حذِر

Chaque schème exprime le fait de répéter une action plusieurs fois. Ci-dessous quelques exemples :

- هذا هو قطّاع الطرق.

Le lexème souligné est une exagération dérivée du lexème « قاطع » ou « bandit » en langue française. Il est utilisé pour montrer que ce bandit s'est habitué à voler et à commettre des crimes. Ainsi, « قطّاع » est paraphrasé par
« يقطع الطرق كثيرا ». Selon le formalisme des FL, **(Magn)** est la fonction qui modélise les adjectifs et les adverbes connotant l'intensification.

Magn (قاطع) = قطّاع

Prenons aussi l'exemple de « مقدام » qui est employé pour designer l'homme très courageux :

Magn (مُقدِم) = مقدام

C. Le Diminutif (صيغة التصغير)

Dernièrement, les tournures du diminutif. Elles sont utilisées dans le but de :

- valoriser ou dévaloriser soit une personne, soit une chose, soit une place ;
- marquer la diminution d'une quantité ou d'un espace ;
- désigner la courte durée ;
- désigner le rapprochement.

Parmi les schèmes selon lesquels le diminutif est construit : « فَعيل », « فُعَيعِل » et « فُعَيعيل ». Examinons les exemples suivants (11) :

- سويسرا دويلة بارعة في صناعة الساعات.
- هذا قليم سيئ الخط.
- يجري في الوادي نهير يسقي المزارع.
- نذهب إلى المسجد قبيل الصلاة.

Dans la première phrase, le diminutif connote la valorisation. « دويلة » est dérivé de « دولة / pays ». Pourtant, dans la deuxième phrase, le lexème « قليم » qui est dérivé de « قلم / crayon » exprime la dévalorisation. Quant à la troisième phrase, le lexème souligné est le diminutif du lexème « نهر / fleuve ». Il reflète le sens « petit fleuve ». Enfin, le diminutif dans la quatrième phase exprime la petite durée existant entre l'acte d'aller à la mosquée et la prière. « قبيل » est dérivé de « قبل / avant », il est paraphrasé par « juste avant ».

De ce qui précède, il paraît clair que le diminutif représente un phénomène linguistique particulier vu les valeurs sémantiques contradictoires qu'il exprime. Ce qui par conséquent justifie l'importance de l'analyse linguistique qui doit précéder tout processus de formalisation. Car, dans ce cas, ce n'est pas le phénomène qui sera mathématisé – comme l'hyperbole – mais la valeur sémantique exprimée par la forme lexicale. Ce qui veut dire que chaque valeur sémantique sera formalisée par une FL reflétant son sens. La valorisation sera formalisée par la FL **(Pos)** qui exprime l'évaluation positive alors que la dévalorisation sera représentée formellement par la FL composée **(AntiPos)**.

Pos (دولة) = دويلة

AntiPos (قلم) = قليم

En ce qui concerne la valeur diminutive, Anne-Laure Jousse y a accordé, dans sa thèse, la FL **(Dimin)**. [JOUSSE, 2010]

Dimin (نهر) = نهير

Reste la valeur exprimée dans la quatrième phrase, Jousse a parlé de la FL composée **($\mathbf{AntiMagn_{temps}}$)** qui connote la courte durée par opposition à **($\mathbf{Magn_{temps}}$)** qui désigne la longue durée.

6 CONCLUSION

Pour conclure, le présent article a comme objectif principal de jeter la lumière sur le formalisme des fonctions lexicales en tant qu'un système de mathématisation rigoureux qui est capable d'assimiler les différents phénomènes linguistiques d'une langue naturelle. Les principales catégories du formalisme des FL ont été explicitées en premier lieu. Puis, des exemples qui cernent les liens linguistiques fondamentaux modélisés par les FL ont été énumérés en deuxième lieu. De même, il était nécessaire de mettre en évidence les atouts du formalisme de la TST : l'analyse sémantique représentant le socle de ce système permet de comprendre les nuances de sens ainsi que le sens composé des unités lexicales.

A travers l'étude de trois phénomènes arabes (le cliché, l'hyperbole et le diminutif), nous avons démontré que les fonctions lexicales sont applicables sur la langue arabe. Mais les études sont encore peu nombreuses, beaucoup de champs ne sont pas encore découverts. Alors, nous espérons que cette brève étude encouragera les chercheurs à se lancer dans ce domaine riche et d'étudier avec profondeur le formalisme des FL.

NOTES

([1]) « Gérer une langue naturelle par la machine » ne signifie pas seulement le fait que la machine comprenne la langue humaine mais aussi qu'elle puisse réagir à l'égard des ordres donnés par l'utilisateur.

([2]) http://www.atilf.fr/

([3]) Le laboratoire ATILF a élaboré le projet RELIEF (REssource Lexicale Informatisée d'Envergure sur le Français). Pour des extraits clarifiants, veuillez visiter le lien suivant : https://spiderlex.atilf.fr/fr/id/35415

([4]) Le terme « syntactique » a été utilisé pour la première fois par Mel'čuk afin de désigner la combinatoire lexico-syntaxique d'un signe linguistique.

([5]) Nous visons par « le sens contextuel » le sens de l'unité lexicale au sein du contexte. Autrement dit, le sens qu'elle acquiert de sa combinatoire avec d'autres unités lexicales.

([6]) Les liens paradigmatiques représentent les relations sémantico-syntaxiques de base d'une langue naturelle.

([7]) Les liens syntagmatiques concernent la constitution des collocations ou des syntagmes semi-figés qu'une unité lexicale peut former avec les autres unités lexicales d'une langue naturelle.

([8]) Le chiffre employé dans la FL **(IncepReal)** dénote le premier actant du verbe. Dans ce cas, c'est le sujet ou la personne qui fond en larmes.

([9]) Nous rappelons que l'entrée dans un dictionnaire arabe est une racine et non un lexème. Ce qui fait que sous la même entrée nous trouvons tous les lexèmes qui sont dérivés de la même racine.

([10]) Selon le dictionnaire arabe *المكنز الكبير*, version numérisée, 2000, p. 459.

([11]) Les exemples sont extraits du site https://www.tunisia-sat.com/forums/threads/1712077

RÉFÉRENCES

[1] A. Clas, I. Mel'čuk et A. Polguère, *Introduction à la lexicologie explicative et combinatoire*. Editions Duculot, 1995.

[2] A. Polguère, *Liste de fonctions lexicales fréquemment utilisées dans le Réseau Lexical du Français (RL-fr)*, 2014.

[3] A.-L. Jousse, *Modèle de structuration des relations lexicales fondé sur le formalisme des fonctions lexicales*, Thèse de doctorat, Université Paris 7, 2010.

[4] I. Mel'čuk et J. Milicévic, *Introduction à la linguistique*, volume 1. Hermann, 2014.

[5] J.-B. Grize, *Langues logico-mathématiques et langues naturelles*, in Revue française de pédagogie, vol.23, 1973. Version électronique in https://www.persee.fr/doc/rfp_0556-7807_1973_num_23_1_1831 , consulté le 12/12/2019.

[6] R. Salem, *Modélisation de la polysémie : approche contrastive arabe – français basée sur la Théorie Sens – Texte*. Thèse de doctorat. Université d'Alexandrie, 2017.

[7] S. Auroux, *Mathématisation de la linguistique et nature du langage*, in Histoire Epistémologie Langage, tome 31, 2009. Version électronique in https://www.persee.fr/doc/hel_0750-8069_2009_num_31_1_3105 , consulté le 12/12/2019.

[8] *أنواع الترادف في اللغة العربية* , منتدى مجمع اللغة العربية على الشبكة العالمية http://www.m-a-arabia.com/vb/showthread.php?t=26627 , consulté le 17/1/2020

[9] L. Moussaoui, *دراسة دلالية تقابلية للتعابير المسكوكة بين العربية و الفرنسية* , https://www.academia.edu/13850490/%D8%AF%D8%B1%D8%A7%D8%B3%D9%80%D9%80%D9%80%D8%A9_%D8%AF%D9%84%D8%A7%D9%84%D9%8A%D8%A9_%D8%AA%D9%82%D8%A7%D8%A8%D9%84%D9%8A%D8%A9_%D9%84%D9%84%D8%AA%D8%B9%D8%A7%D8%A8%D9%8A%D9%80%D9%80%D8%B1_%D8%A7%D9%84%D9%85%D8%B3%D9%83%D9%88%D9%83%D8%A9_%D8%A8%D9%8A%D9%86_%D8%A7%D9%84%D8%B9%D8%B1%D8%A8%D9%8A%D8%A9_%D9%88%D8%A7%D9%84%D9%81%D8%B1%D9%86%D8%B3%D9%8A%D8%A9_A_Semantic_and_comparative_study_of_fixed_expressions_between_Arabic_and_French , consulté le 5/1/2020.

السيرة الذاتية

رشا محمد سالم

مدرس بقسم اللغة الفرنسية، كلية الآداب، جامعة الإسكندرية. تخصص لغويات حاسوبية. حاصلة على درجة الدكتوراه عام 2017 و موضوعها "المعالجة الآلية للألفاظ متعددة المعاني : دراسة مقارنة بين اللغتين العربية و الفرنسية". أثناء تواجدها في فرنسا في مهمة علمية عام 2016 ، شاركت في العديد من ورش العمل و الدورات التدريبية الخاصة بعلم المعاني الحاسوبية و التي نظمها معمل ATILF التابع لجامعة لورين. تم نشر بحثين لها في مجال الترجمة الآلية.

BIOGRAPHY

Lecturer, Department of French Language, Faculty of Arts, Alexandria University. Computer Linguistics. She holds a PhD in 2017 with the topic "Automated word processing with multiple meanings: a comparative study between Arabic and French". While in France on a scientific mission in 2016, she participated in many workshops and training courses on computer meanings organized by the ATILF laboratory of the University of Lorraine. Two papers have been published in the field of machine translation.

Application of Lexical Functions Formalism on the Arabic Language

Racha MOHAMMAD SALEM
Department of French Language and Literature, Faculty of Letters, University of Alexandria
salem.racha@live.com

Abstract: Research on the formalization of the Arabic language began in the nineties of the previous century. Modern Arab linguists seek to create a rigorous formal system which is inspired by the formalization systems of European languages and which is capable of managing the Arabic language with all its riches. On the following pages, we will propose a method of formalizing Arabic according to the lexical function system (encoding system of the Meaning-Text theory). The latter will be applied to three linguistic phenomena: cliche, hyperbole and diminutive.

Keywords*:* Lexical functions, Arabic linguistics, language mathematization, Meaning-Text theory, automatic processing
natural language.

تطبيق اللغة الحاسوبية القائمة على الوظائف المعجمية على اللغة العربية

رشا محمد سالم
مدرس بقسم اللغة الفرنسية و آدابها، كلية الآداب، جامعة الإسكندرية
salem.racha@live.com

الملخص:
بدأ البحث في حوسبة اللغة العربية في تسعينيات القرن الماضي. يسعى اللغويون العرب المحدثون إلى إنشاء نظام حاسوبي قوي مستوحى من أنظمة حوسبة اللغات الأوروبية ، يكون قادار على إستيعاب اللغة العربية بكل ثرواتها. في الصفحات التالية ، سنقترح حوسبة اللغة العربية وفقًا لنظام الوظائف المعجمية (نظام الحوسبة الخاص بنظرية المعنى- النص). سيتم تطبيق هذا الأخير على ثلاث ظواهر لغوية عربية : التعبيرات المسكوكة، صيغ المبالغة، صيغ التصغير.

الكلمات المفتاحية : الوظائف المعجمية ، اللغويات العربية ، لغة حوسبة ، نظرية المعنى- النص ، المعالجة الآلية للغة الطبيعية.

Quelles Contraintes pour Traduire la Morphologie et la Syntaxe ?

Asmaa Gaafar Abdel-Rassoul

Faculte de Lettres, Universite de Menoufia

gmasmaa@yahoo.com

***Résumé*— Cette recherche vise à étudier l'importance de la morphologie et la syntaxe sur La traduction. Nous allons distinguer la morphologie lexicale de la morphologie dérivationnelle ainsi que le micro-syntaxe de la macro-syntaxe. Par conséquent, notre problématique consiste à répondre aux questions suivantes en appliquant la théorie interprétative dépendant de l' ISIT à la Sorbonne : les deux sont-elles nécessaires pour le texte ? Devons-nous les traduire littéralement d'une langue à l'autre ? Alors, quelles sont les répercussions de les traduire littéralement sur le texte ?**

***Keywords :* syntaxe, morphologie, traduire littéralement.**

1 INTRODUCTION

Bien entendu, la tâche de **la morphologie** est basée sur le morphème qui se réalise au niveau de deux échelles : en premier lieu, **le morphème zéro ou la morphologie lexicale** qui s'intéresse à l'étude de la racine et de l'étymologie de l'unité lexicale. En second lieu, **les morphème (s) ou la morphologie dérivationnelle** qui consiste à étudier les changements intervenus aux unités minimales à travers quelques procédés comme la flexion, la dérivation et la composition. Par ailleurs, la syntaxe est caractérisée par deux notions fondamentales : premièrement, **le micro-syntaxe** qui porte sur l'étude des clauses soit séparément, soit dans une phrase simple. Deuxièmement, **le macro-syntaxe** qui a pour objectif d'analyser la période à travers l'enchaînement des clauses dans une phrase complexe en formant implicitement un sens logique, ce qui nous incite à dire qu'il y a une étroite relation entre le macro-syntaxe et la sémantique.

2 LA RATIONALISATION

La rationalisation que nous visons dans notre étude, c'est toute modification au sein du texte sur le plan structural pour conserver un équilibre communicationnel entre le traducteur et le destinataire comme dans le cas de l'orthographe qui est un outil pour une bonne organisation du texte et par conséquent une bonne compréhension.

Selon Podeur :

> **«La rationalisation correspond à la simplification de l'organisation morpho-syntaxique d'un énoncé et à la modification de la ponctuation» [1].**

À cette période-là où l'œuvre a été traduite, l'utilisation de la ponctuation n'a pas été connue. Le décalage temporel joue un rôle très important à ce sujet. Nous ne pouvons pas dire que c'est une lacune de la part des traducteurs. Cela est probablement dû à la nature de l'imprimerie qui n'était pas développée. **Zaki Pacha** [2] est le premier qui a écrit un livre parlant de la ponctuation qui est très proche de celle qui a été utilisée dans les journaux étrangers et il a souligné qu'il était utile de les imiter. **Zaki Pacha** a aussi indiqué, à ce sujet, que les Turcs ont également utilisé la ponctuation dans leurs journaux mobiles. En revanche, nous trouvons que **El-Hamouz** [3] a dit en bref que les anciens grammairiens arabes ont connu cette ponctuation mais par des signes qui diffèrent de ceux que nous utilisons aujourd'hui.

Examinons quelques exemples où le traducteur n'a pas utilisé la ponctuation que dans les plus rares cas :

Ex1 : «Mon pauvre enfant… nous n'avons qu'un parti à prendre. **Tu es robuste, brave, intelligent** ; […]. Le recrutement t'atteindra l'an prochain ; devance le moment, fais-toi soldat, tu pourras du moins choisir ton arme…**»**. SUE (Eugène), *L'Orgueil*, p. 3.

«"قال لهُ" وقد بحثت ونقبت فما وجدت لك سوى سبيلين إما انك تلج أبواب الجندية وتتهياء لمراقيها قبل ان تدهمك بالقرعة وهى تليق بك **لأنك قوى باسل وذكى»**. حضرة ديمترى أفندي خلاط، *عزة النفس*، الأهرام، عدد1161، 1881.

Évidemment, le traducteur, dans cet exemple, n'a pas pu retransmettre littéralement la ponctuation de la phrase source car la ponctuation arabe suit autrement le découpage de la phrase. Par la suite, le traducteur a pris en conscience la spécificité de la ponctuation de la langue arabe en suivant les normes réceptrices afin d'unifier, dans sa version, entre l'ordre discursif et le rythme thématique exprimés dans le texte de départ.

Ex2 : **«**Ne m'avouez-vous pas vos peines ? Aimez-vous quelqu'un **?** **»**. DUMAS, *Vicomte de Bragelonne **II***, p. 107.

«قال لماذا تكتمين عنى أسرارك اتحبين**.**». نجيب الحداد، *عود على بدء*، الأهرام 4475، 1892.

Quant à cet exemple, le traducteur a amalgamé les deux phrases françaises en une seule arabe avec un point final. Nous pouvons aussi ajouter, à ce sujet, que la version du traducteur n'a aucun effet négatif sur le sens de la phrase cible vu que cette forme arabe «**اتحبين**» porte en soi la forme d'interrogation selon la chaîne syntaxique arabe. Cette forme «اتحبين» suit la flexion des formes dérivées de l'arabe de [4] «أتفعلين».Enfin de compte, le sens est retransmis clairement.

Ex3 : «Vos revenus se montent à la somme de *trois millions cent vingt mille francs* environ, ce qui vous fait à peu près *huit mille francs* par jour. Rien que cela, –, a ajouté le notaire en riant, – aussi êtes-vous ***LA PLUS RICHE HÉRITIÈRE DE FRANCE*»**. SUE (Eugène), *L'Orgueil*, p. 52.

«وقال لى ان دخلك يا حضـــرة الفتاة بالغ زهاء ثلاثة ملايين ومائة وعشـــرين الف فرنك بموجب الحســـاب المدقق فيكون دخلك اليومى نحو ثمانية الاف فرنك وبناءً عليهِ انت **اثرى وريثة فى فرنســـا**». حضـــرة ديمترى أفندى خلاط، *عزة النفس*، الاهرام، عدد1199، 1201، 1881 .

Dans l'exemple ci-dessus, l'auteur a voulu focaliser l'attention du lecteur sur le fond de la phrase en ayant recours au phénomène du majuscule tout en amplifiant les lettres. En contrepartie, l'arabe ne possède pas le phénomène du majuscule, mais quand même le sens est retransmis vu que le traducteur, avec une intelligence professionnelle, a opté pour la forme connue par «l'élatif» [5], pour remplir la même fonction de la phrase source en insistant sur la tonalité et la version rythmique afin de piquer la curiosité du lecteur.

Ex4 : «————Oh ! ma mère… ma mère ! —— murmura Gerald avec un accent d'ineffable reconnaissance, en tombant aux genoux d'Herminie et couvrant ses mains de larmes et de baisers**»**. SUE (Eugène), *L'Orgueil*, p. 136.

«فانكب جرال على اقدام ارمنا وطفق يبوس يديها ويغسـلهما بدموعه ويقول لامه بلسـان الامتنان -**لاعدمتك يا امى»**. حضـرة ديمترى أفندى خلاط، *عزة النفس*، الاهرام، عدد 1336، 1882 .

Dans cet exemple, nous pouvons constater qu'il y a une coordination sémantique entre la répétition et l'orthographe. En plus, nous trouvons que la répétition, représentée dans le syntagme nominal «ma mère» et le point d'exclamation (!), reflète un effet très fort. Mais, quand il s'agit de la ponctuation arabe et la ponctuation française, il n'y a aucune similitude, ce qui constitue une pierre d'achoppement d'une part, et ce qui explique le choix de cette locution «**لاعدمتك يا أمى**» de la part du traducteur d'autre part, vu que cette locution bien connue, embrasse le sens et l'effet du texte source car l'énoncé français montre que Gerald a entendu une nouvelle presque impossible.

Avant d'aborder le problème de la **hamza** dans l'exemple ci-dessous, nous pouvons donner quelques observations : en feuilletant ***Al-Ahram*** à cette époque-là, nous avons remarqué que la hamza n'a pas été écrite que dans les plus rares cas. C'était un trait caractéristique de l'époque du roman.

Ex5 : «— Mademoiselle ! Oh ! ne m'appelez pas ainsi, — s'écria mademoiselle de Beaumesnil, **— ne suis-je donc plus** votre Ernestine, l'orpheline à qui vous avez promis votre amitié… parce que vous la croyiez malheureuse ?... ». SUE (Eugène), *L'Orgueil*, p. 121.

«**الست** صديقتكِ ارنسة الست تلك الفتاة اليتيمة التى حنَّ قلبكِ عليها ووعدتِها بمودتك». حضرة ديمترى أفندى خلاط، *عزة النفس*، الاهرام، عدد1311، 1882.

Il est à signaler que «**la hamza**» est un alphabet indispensable dans l'écriture de la langue arabe car à travers laquelle, nous pouvons mettre la main sur le sens exact. Quant à l'exemple ci-dessus, le traducteur n'a pas écrit le syntagme arabe «الست» avec la hamza comme c'était la coutume de cette période-là parce qu'il comptait sur le contexte qui déchiffrait, à son tour, le sens voulu.

3 LES TEMPS VERBAUX

De prime abord, nous travaillons sur deux systèmes temporels différents. L'arabe qui appartient aux langues sémitiques et le français qui appartient aux langues indo-européennes. En outre, l'arabe connaît deux formes accomplie et inaccomplie, tandis que le français connaît aussi l'accompli et l'inaccompli mais en ajoutant que ce dernier comporte plusieurs formes contrairement à l'arabe.

Chaïret a indiqué que :

> **«*Le système verbal de l'arabe, comme celui des langues sémitiques d'une manière générale, a la réputation de reposer sur l'opposition aspectuelle accompli-inaccompli. Les distinctions temporelles n'y seraient assurées que par l'environnement contextuel et éventuellement par des éléments auxiliaires*» [6].**

Examinons quelques exemples en la matière :

Ex1 : «Que l'on juge de l'étonnement du marquis et de Gerald. Tous deux **arrivaient** pâles… effarés… comme des gens qui accouraient sauver quelqu'un d'un grand danger…**»**. SUE (Eugène), *L'Orgueil*, p. 145.

«فليتصـور القاري مقدار انذهال دى ملفور وجرال اللذين **اقبلا** مصـفرين منزعجين خوفاً على الفتاتين من خطر جسيم». ديمترى أفندى خلاط، *عزة النفس*، الأهرام، عدد 1353، 1882.

Selon la règle arabe, «l'imparfait» en français se traduit par «**كان+المضارع** ». Par conséquent, la traduction devrait être «**كانوا يقبلون**». Le traducteur a opté pour le substantif «**اقبلا**» sous cette forme pour avoir un niveau de langage soutenu du point de vue poéticité. Par la suite, le traducteur a pu garder la même valeur morpho-syntaxique évoquée par le texte source.

Ex2 : «Je **resterai** toute la journée dans ma chambre**»**. SUE (Eugène), *L'Orgueil*, p. 71.

«ولا اخرج من غرفتي». ديمترى أفندي خلاط، *عزة النفس*، الأهرام، عدد1217، 1881.

Quand le traducteur aborde les temps verbaux, il prend en considération leurs contextes qui précisent la traduction pertinente des temps verbaux. Il est à noter que chaque temps fait une partie intégrante de sa langue.

Dans l'exemple ci-dessus, le traducteur a restitué le futur simple par le présent. Il est à signaler que dans ce contexte le présent donne le sens du futur à travers la formule de négation «**لا**». Le traducteur a choisi un autre temps verbal qui justifie sa compétence langagière et qui véhicule l'effet de sens particulier évoqué dans le texte source.

> ***«La traduction des temps verbaux s'appuie forcément sur des équivalences susceptibles de rendre un effet de sens particulier révélé dans le texte source»[7].***

Ex3 : «Je **me moquerais** fort de la noblesse en général**»**. SUE (Eugène), *L'Orgueil*, p. 129.

«فانى **اضحك** على الحسب وأهله**»**. ديمترى أفندي خلاط، *عزة النفس*، **الأهرام**، عدد1323، 1882.

La correspondance morpho-syntaxique totale est presque inaccessible de deux cultures assez distanciées l'une de l'autre comme notre cas. Chaque temps a ses particularités et sa fonction qui se diffèrent d'une langue à l'autre.

L'existence de l'adjectif «fort» a neutralisé la formule du conditionnel et son effet vu que la notion de doute - l'essence du conditionnel - se contredit avec le sens voulu de l'adjectif «**fort**». Il est à signaler que le conditionnel neutre équivaut au présent. Donc, le traducteur a capté cette modulation.

> ***«Dans toutes les langues, le système verbal diffère passablement, et dans un contexte bilingue qui est celui de la traduction, de nombreux problèmes d'équivalence se posent»[8].***

4 LA TRANSPOSITION

La transposition est une sorte de la traduction libre travaillant sur le signifiant. En d'autres termes, c'est une traduction linguistique à travers laquelle le traducteur adopte des changements aux catégories grammaticales. Par conséquent, **Oustinoff** [9] l'appelle «une récatégorisation».

Selon Forges et Braun, la transposition consiste à ***«remplacer une partie de discours par une autre, sans changer le sens du message»*[10].**

Ex1 : «Ernestine **sourit tristement»**. SUE (Eugène), *L'Orgueil*, p.121.

«فابتسمت ارنسة **ابتسامة الحزن»**. ديمترى أفندي خلاط، *عزة النفس*، الأهرام، عدد 1311، 1882.

Le traducteur a eu recours à la transposition pour éviter toute littéralité qui pourrait porter atteinte au sens. Il a opté, avec un savoir-faire, pour le complément absolu outre le nom, ce qui nous a donnés un sens qui s'accorde avec le génie de la langue arabe.

Oustinoff a mentionné à ce propos :

***«Chaque fois que la traduction "directe" ou "littérale" aboutit à un énoncé équivalent sur le plan linguistique et stylistique, on le maintiendra ; dans le cas inverse, il faudra recourir à la traduction oblique»*[11].**

Ex2 : «Désespérés ! ... mais pourquoi cela ? ...[…] **——** s'écria inconsidérément Ernestine […]**»**. SUE (Eugène), *L'Orgueil*, p. 119.

«قالت ارنسة لمَ انت مستمرة على **القنوط** يا ارمنا». ديمترى أفندى خلاط، *عزة النفس*، الأهرام، عدد 1309، 1882.

Quant à cet exemple, le traducteur a évité le calque syntaxique de la phrase source, qui serait, à titre d'exemple : **«يائسة ! لمَ هذا اليأس»**. Le traducteur a sollicité un changement grammatical - **«désespérés»**, un adjectif remplacé en arabe par le syntagme arabe «**القنوط**» - qui convient avec le tissu morpho-syntaxique arabe en reproduisant cette phrase.

D'ailleurs, l'ajout de l'adjectif «**مستمرة**» donne un excès d'éclaircissements au sens voulu afin de souligner la frustration durable d'Herminie.

Oustinoff a aussi ajouté à ce sujet :

«La traduction doit donner l'impression que l'original a été écrit directement en français : la visée est "cibliste"»[12].

5 LE LANGAGE ET LA PENSÉE

En traitant ce point très délicat, nous pouvons mettre l'accent sur le rapport indissociable entre le signifiant et la pensée. À travers leur enchaînement découle deux notions fondamentales : **l'intertextuelle et l'interdiscursive** qui montrent que le traducteur ne traite pas le texte en s'appuyant sur les signes linguistiques seulement, mais à travers une sorte de corrélation entre ces signes et la culture. En d'autres termes, le traducteur essaye dans ce cas, de restituer le texte conformément au contexte social voulu de la part de l'auteur. En somme, la notion du génie de la langue en général fait allusion à ne pas traiter les langues comme des stéréotypes, mais plutôt le traducteur doit bien manipuler l'ensemble **référentiel/inférentiel**.

Voici quelques exemples à cet égard :

Ex1 : «Le roi, […], descendant de cheval au moment où l'on ouvrait la portière du carrosse, il lui avait offert la main. […] En voyant le roi entrer bravement dans **le bois** avec La Vallière, […]**»**. DUMAS, *Vicomte de Bragelonne II*, p. 18.

«وترجل الملك عن جواده واقبل حتى اخذ بيد لويزا فانزلها من المركبة وســـار بها إلى **احد جوانب الغابة»**. نجيب الحداد، عود *على بدء*، الأهرام، 4467، 1892.

Le fait de retransmettre l'article défini «**le bois**» en un article indéfini «**احد جوانب الغابة**»، dans **le premier exemple**, montre que le traducteur a sollicité sa pleine visualisation de la situation pour une bonne restitution de la situation. Il est évident que le roi en entrant dans le bois avec La Vallière, ils se sont assis dans une place précise. Dans cette perspective, nous pouvons souligner que cette version a jeté la lumière sur la fonctionnalité suprême de la traduction qui donne la priorité aux éléments suprasegmentaux.

Ex2 : «Herminie, jusqu'alors craintive, accablée releva orgueilleusement **la tête»**. SUE (Eugène), *L'Orgueil*, p. 135.

«فلما سمعت ارمنا عبارة الأميرة المهينة رفعت **راسها** بعزة وتوردت وجناتها». ديمترى أفندي خلاط، *عزة النفس*، الأهرام، عدد 1330، 1882.

Il en est de même pour **le deuxième exemple**, la traduction du syntagme français **«la tête»** par «**رأسها**» est plus précise car il parle d'une chose particularisante : de sa tête.

« L'arabe préfère les formules concrètes et personnalisées»[13].

Ex3: « Votre médecin ne vous a-t-il pas déclaré devant moi que, sans les moyens héroïques auxquels il venait de recourir, **vous risquiez de perdre votre fils d'une fièvre cérébrale ?** ». SUE (Eugène), *L'Orgueil*, p. 127.

« قال لكِ انهُ مصاب بحمى فى الدماغ كادت ان . . . **(لا سمح الله)** لولا اجراء الوسائط الفعالة لازالتها ». حضرة ديمترى أفندى خلاط، عزة النفس، الاهرام، عدد 1320، 1882.

Concernant **le troisième exemple** : le traducteur a omis volontairement cet énoncé «**vous risquiez de perdre votre fils d'une fièvre cérébrale**» car il fait appel à l'esprit une mauvaise augure. Le traducteur l'a remplacé par les points de suspension **(…)** et il a prédit le bon augure en disant «**لا سمح الله**». Ce changement textuel est dû à la dissemblance culturelle. Donc, le traducteur, en exploitant les aspects cognitifs du texte de départ, a prêté une attention particulière à l'unité-texte qui donne la priorité à la cohérence des idées d'une part, et qui vise à garder la spécificité de la langue arabe d'autre part.

Ex4 : « - Elle veut cela ?... **Oh** ! la vaillante et noble fille ! - s'écria le marquis, après un moment de surprise ». SUE (Eugène), *L'Orgueil*, p. 123.

ـــ « فصاح الامير بتعجب ـ **عافاك الله** ايتها الفتاة الباسلة يا ارمنا ». حضرة ديمترى أفندى خلاط، عزة النفس، الأهرام، عدد 1313، 1882.

Quant à cet exemple, le traducteur, en restituant «**Oh !**» par «**عافاك الله**»، il a pu atteindre la dimension esthétique de la langue arabe. Par la suite, il a voulu garder la force émotive de l'énoncé source.

Jean et Brisset ont mentionné à ce sujet :

«Le traducteur, jusqu'alors invisible, acquiert le statut explicite d'un spécialiste de la communication interculturelle. On exige qu'il soit capable de

***déterminer les moyens de médiation les plus fonctionnels, c'est-à-dire les mieux adaptés aux objectifs de la communication dans un contexte socioculturel donné*»[14].**

6 LE DISCOURS RAPPORTÉ

Nombreuses sont les formes du **discours rapporté** : **le discours direct, le discours indirect, le discours indirect libre et le discours narrativisé**. Il est à signaler que le sens rapporté est leur production commune. **Le discours direct** est le plus fidèle car le discours cité a la primauté sur le discours citant. Par contre, **le discours indirect** est le contraire de la première forme car il est connu sous la forme d'un récit et la voix du narrateur s'impose tout au long du roman au détriment des personnages. Autrement dit, le disours citant prend le pas sur le discours cité. En plus, **le discours indirect libre** fait un amalgame de deux premiers discours. La polyphonie est l'une des caractéristiques de ce discours car le narrateur participe au dialogue ou plus particulièrement au récit avec les personnages. Dernièrement, **le discours narrativisé** dont la tâche principale est la concision.

Examinons quelques exemples en la matière :

Ex1 : «Celui-ci, à peine entré, dit à la portière :

- Dans quelques instants une dame viendra demander mademoiselle Herminie… vous l'introduirez.

- Oui, monsieur, - répondit madame Moufflon en se retirant». SUE (Eugène), *L'Orgueil*, p. 134.

«ولما صار في وسط المحل قال للحاجبة ستقدم بعد ساعة سيدة مصونة وتلتمس مقابلة السيدة ارمنا فادخليها فأجابت الحاجبة بالإيجاب وانصرفت». ديمترى أفندي خلاط، *عزة النفس*، الأهرام، عدد1329، 1882.

Ex2 : «À ce moment, le valet de pied, qui avait disparu avec la voiture, revint sur ses pas, avisa à travers la grille les personnages rassemblés sous la tonnelle, s'approcha, et mettant la main à son chapeau :

—— Messieurs, pourriez-vous, s'il vous plaît, me dire si ce jardin dépend de la maison numéro 7 ? ». SUE (Eugène), *L'Orgueil*, p. 8.

«فنزل المجرى بعد ما نقب في عنوان رقعةٍ بيده كان يبحث عن نمرة المحل المقصود ثم دنا من مقام الجلساء ورفع قبعتهُ قائلاً هل هذه الحديقة تابعة لبيت نمرة 7». ديمترى أفندى خلاط، *عزة النفس*، الأهرام، عدد 1164، 1881.

Ex3 : «À la voix du commandant Bernard, madame Barbançon arriva en hâte, s'excusa auprès de son maître, et dit au domestique qui attendait :

– Vous avez une lettre pour moi… mon garçon ? et de quelle part ? **».** SUE (Eugène), *L'Orgueil*, p. 8.

«فنادي القبطان المدبرة بالحضور فاتت مسرعة وخاطبت المجرى». ديمترى أفندي خلاط، *عزة النفس*، الأهرام، عدد 1164، 1881.

Ex4 : «Puis, s'interrompant, l'ancienne sage-femme poussa une exclamation, comme si une idée subite lui eût traversé l'esprit, et elle dit à son maître :

- Monsieur…

- Eh bien ! ...

- Voulez-vous venir un instant avec moi dans le jardin ? J'ai à vous parler en secret, dans le plus profond secret». SUE (Eugène), *L'Orgueil*, p. 9.

«(...) ثم التمست منهُ ان يتنحى لتستشيرهُ بامر سرى جال في خاطرها». ديمترى أفندي خلاط، *عزة النفس*، الأهرام، عدد1164، 1881.

Bien entendu, chaque traducteur traduit selon sa formation méthodologique et son bagage cognitif qui doit être assez fort et varié. Le traducteur peut certainement emprunter quelques structures semblables mais en conservant le sens logico-sémantique du texte original.

D'ailleurs, le traducteur a eu recours au changement en traduisant le style direct en style indirect. Ce qui nous intéresse dans ce contexte est le sens logique des phrases et la retransmission du vouloir-dire de l'auteur.

Selon Rudy : ***«Traduire un texte, c'est transposer "transférer" un contenu informationnel d'un système linguistique A à un système linguistique B, chacun possédant ses propres règles. Le contenu informationnel, en d'autres termes le sens logico-sémantique, doit rester stable au cours de l'opération pour que la traduction soit la plus fidèle possible»*[15].**

Le traducteur n'était pas dans l'obligation de traduire le style direct par un autre indirect car le style direct est le style le plus préférable au sein de la chaîne syntaxique arabe, mais par rapport à la période pendant laquelle le roman a été publié, le style indirect était le style préférable. En même temps, le traducteur n'a pas illustré l'esprit du texte source en omettant la scène théâtrale existante dans les quatre exemples en question. Donc, il y a un acte de violence concernant la forme car le traducteur a effacé toute trace de pleine saturation avec le texte original.

7 LE CHANGEMENT D'UN ACTE DE PAROLE

L'acte de parole fait une partie intégrante de la pragmatique car le langage avec la pragmatique se transforme en action qui nécessite par suite une réaction de la part d'autrui. D'ailleurs, il y a trois types de l'acte de parole:

les actes locutoire, illocutoire et perlocutoire. Premièrement, **l'acte locutoire** consiste au sens même de l'énoncé. Deuxièmement, **l'acte illocutoire** est le but de la personne qui parle. Dernièrement, **l'acte perlocutoire** est l'influence de la personne sur l'autrui.

Voici quelques exemples à ce propos :

Ex1 : «-Vous m'avez dit tout à l'heure, madame, et très sagement, qu'il ne fallait plaisanter, ni avec la noblesse, ni avec la religion, **n'est-ce pas ?**

- Oui, monsieur le marquis.

- Vous avouerez qu'il ne faut pas non plus plaisanter avec le mariage ? **».** SUE (Eugène), *L'Orgueil*, p. 129.

«قد قلتِ لي الآن يا سـيدتي عدم جواز الهزء بالحسـب والدين **وانا أقول لكِ** انهُ لا يجوز الهزء بالزواج». ديمترى أفندي خلاط، *عزة النفس*، الأهرام عدد 1323، 1882.

Il est à noter que «**n'est-ce pas ? Oui, monsieur le marquis**» est remplacé par «**وانا اقول لكِ**» car à travers cet énoncé, l'auteur n'a pas voulu désigner le doute mais plutôt la confirmation vu que la négation de la négation exprime la confirmation. En plus, le traducteur a eu recours à cette modulation pour mettre l'accent sur l'unité de pensée révélée dans le texte source.

Ex2 : «- Comte, dit-elle, ménagez-moi. Vous voyez que je souffre, **et vous n'avez aucune pitié**». DUMAS, *Vicomte de Bragelonne II*, p. 107.

«وقالت انك تقتلني يا كونت **أفلا تشفق عليَّ**». نجيب الحداد، *عود على بدء*، الأهرام، 4475، 1892.

Quant au **deuxième exemple**, le traducteur a modulé la phrase source qui est en négation «**et vous n'avez aucune pitié**» en une phrase interrogative «**أفلا تشفق عليَّ**» car Madame «Beaumesnil» a voulu s'interroger sur la pitié du comte. Le traducteur a mis l'accent, dans ce contexte, sur l'unité de traduction en faisant abstraction aux unités simples. C'est la raison pour laquelle, le traducteur a donné la priorité à la traduction oblique en restituant cet énoncé. Par la suite, le traducteur, par son intelligence professionnelle, a pu prendre en considération l'acte de langage voulu et l'a traduit selon l'intention de l'auteur.
Autrement dit, le traducteur n'a totalement pas changé le texte, mais il s'est contenté de faire un changement sur le plan structural seulement, ce qui lui a aidé à garder la même allure du texte source.

Keightley a montré à ce propos :

> ***«Une traduction doit impérativement transmettre les mêmes informations que l'original mais il faut en même temps veiller au bon fonctionnement pragmatique dans la situation destinée»[16].***

8 LES NOMS DE NOMBRE

Dans la langue arabe, **les noms de nombre** sont différents des noms qui les suivent. Plus précisément, les noms de nombre de (1 jusqu'à 10) ne suivent pas les genres du nom qui les suivent.

> ***«Le nom de nombre ne devrait pas être influencé par le genre du nom compté, […]. Mais suivant la doctrine des grammairiens arabes, il a été nécessaire, pour confirmer la qualité de substantif à ces noms de nombre, de leur donner le genre inverse de celui qu'a le nom compté au sing. ; c'est-à-dire que les noms de nombre de 3 à 10, […], accompagnent les noms comptés féminins et que les noms de nombre prennent la désinence : quand ils sont construits avec des noms masculins»[17].***

Ex1 : «Une vieille ménagerie, […], était depuis **dix ans**, au service du commandant Bernard». SUE (Eugène), *L'Orgueil*, p. 1.

«(...) مع خادمة امينة على امتعتهِ صـادقة الود لهُ كرَّت عليها بخدمتِه **عشرة سنين**». ديمترى أفندى خلاط، *عزة النفس*، الاهرام عدد1881،1161.

Dans **l'exemple N.1**, le traducteur a littéralement retransmis le nom de nombre «dix» en suivant le genre féminin du nom compté «années». Ce qui contredit les règles des grammairiens arabes. Le traducteur aurait dû le traduire par : **«عشر سنين».**

Ex2 : «Moi, qui ai **sept maisons** sur le pavé de Paris, je n'ai pas seulement de tapis dans mon salon». SUE (Eugène), *L'Orgueil*, p. 61.

«حال كونى املك **سبع منازل** في شوارع باريز وما عندي بساط مفروش في مسكنى». ديمترى أفندى خلاط، *عزة النفس*، الأهرام، عدد 1207، 1881.

Dans cet exemple, le traducteur a suivi la même méthode de l'exemple précédent en traduisant le nom de nombre «sept» en suivant le genre masculin du nom compté «**la maison**» qui est un nom masculin en arabe contrairement au français. Par conséquent, Le traducteur aurait dû le traduire par : «**سبعة منازل**».

Ex3 : «Puis elle compta dix pas de la fenêtre à son lit, et écrivit encore : "**Dix pas**"». DUMAS, *Vicomte de Bragelonne II*, p. 236.

«ثم قاست من النافذة الى سريرها وكتبت "**عشر اقدام**"». نجيب الحداد، عود *على بدء*، الأهرام،4498، 1892.

Dans cet exemple, le traducteur n'a pas réussi à retransporter le nom de nombre : le nom de nombre «**dix**» ne suit pas le genre masculin du nom compté «**pas**». Il est à noter que le substantif «**القدم**» - dans la langue arabe - comme une unité de mesure est considérée comme un nom masculin. À l'inverse, lorsque le substantif «**القدم**» signifie - dans la langue arabe - un membre du corps humain, «**القدم**» constitue un nom féminin [18]. Par la suite, Le traducteur aurait dû le traduire par : «**وكتبت عشرة أقدام**».

9 CONCLUSIONS

En guise de conclusion, nous pouvons conclure que toute traduction valable devrait dépasser inévitablement toutes barrières de nature morpho-syntaxique pour mieux manipuler le rapport explicite/implicite pour être à l'abri de toute opacité culturelle.

Références

[1] J. PODEUR, «La Traductologie entre description et évaluation», in *Repères Dorif Traduction*, médiation, interprétation, volet N.1, June 2013, in www.dorif.it/ezine/ezine_printarticle.php?dorif..

[2]زكى باشا (أحمد)، *الترقيم وعلاماته فى اللغة العربية*، قدم له واعتنى بنشره عبد الفتاح أبو غدة، الطبعة الأولى، المطبعة الأميرية، مصر، 1912، ص 16. (نسخة مصورة).

«ورأيت من المفيد استعمال العلامات الإفرنجية، (...) وفوق ذلك قد استخدمها الأتراك فى (...) جرائدهم السيارة (...)». (Traduit par nos soins)

[3]الحموز (عبد الفتاح أحمد)، *فن الترقيم فى العربية أصوله وعلاماته*، دار عمار، عمان، الأردن، 1991، ص 20.

«وبعدّ فلعلّك تَّتفق معى فى أنَّ المصنَّفين والناسخين القدامى لم يتناسوا تلك العلامات والرموز التوضيحيّة التى لابدُ منها فى الكتابات المختلفة، (...)، وليس بخافٍ أنَّ ما مرَّ يُعد دليلاً بيناً وإشارةً ساطعةً إلى أن للعرب علاماتٍ وأماراتٍ كتلك التى تطالعنا فى الكتابات فى عصرنا، على الرغم من أنَّ هناك اختلافاً فى الأشكال والرموز (...)». (Traduit par nos soins).

[4] R. BLACHÈRE, M. GAUDEFROY-DEMOMBYNES, *Grammaire de l'arabe classique*, Maisonneuve et Larose, France, 1975, pp. 49-73.

[5] *Ibid*., p. 97.

[6] M. CHAÏRET, *Linguistique contrastive et traduction, Fonctionnement du système verbal en arabe et en français*, N. spécial, Ophrys, Paris, 1996, p. 7.

[7] L. AVENDAÑO ANGUITA, «Perspective et temps verbaux : problèmes de traduction», in *CLE des langues*, Avril 2010. **In** **cle**.ens-lyon.fr/.../com.univ.collaboratif.utils.LectureFichiergw?ID... (La dernière consultation était le 15-3-2015).

[8] J. MOESCHLER, C. GRISOT, B. CARTONI, «Jusqu'où les temps verbaux sont-ils procéduraux ?», in *Nouveaux cahiers de linguistique française*, N. 30, 2012, pp. 119-139.

[9] M. OUSTINOFF, *La Traduction, Que sais-je ?,* Presses Universitaires de France, 2003, p. 75.

[10] G. FORGES, A. BRAUN, *Didactique des langues, traductologie et communication*, De Bock Université, Paris, 1998, p. 15.

[11] M. OUSTINOFF, *Op.cit.,*, p. 72.

[12] *Loc.cit.*

[13] C. HECHAÏMÉ, *La Traduction par les textes*, 2e édition revue, Dar el-Machreq éditeurs, Beyrouth, p. 69.

[14] M. Y. JEAN, A. BRISSET, «La Notion de culture dans les manuels de traduction : domaines allemand, anglais, coréen et français », in *META*, V. 51, N. 2, Juin 2006, pp. 389-409.

[15] R.LOOCK, «"Parce qu'en plus il faut traduire la syntaxe ?!" : Contraintes et stratégies dans la traduction de la structuration d'un texte», éd. D'Amélio, *Actes du colloque international "la forme comme paradigme du traduire", Mons, CIPS*, 2009, pp. 173-190.

[16] S. KEIGHTLEY, *Changements syntaxiques, modulations et adaptations dans un texte médical,* un mémoire, Linnæus University, 2013, pp. 17-18.

[17] R. BLACHÈRE, M.GAUDEFROY-DEMOMBYNES, *Op.cit.*, p. 368.

[18] Thaqafat.com › إضاءات (La dernière consultation était le 27-3-2015).
«القَدَمُ: ما يَطَأُ بِهِ الإنْسانُ الأَرْضَ، وَجَمْعُها "أَقْدامٌ" (إذا قُصِدَ بِها وَحْدَةُ القِياسِ المَعْرُوفةُ فَإنَّها تُذَكَّرُ)».

What Constraints to Translate Morphology and Syntax ?

Asmaa Gafar Abdel-Rasoul

Faculty of Literatures

gmasmaa@yahoo.com

***Abstract*—The aim of this research is to study the importance of morphology and syntax in translation. We will distinguish the lexical morphology from the derivative morphology as well as the micro-syntax of the macro-syntax. Therefore, our problem is to answer the following questions by applying the interpretative theory dependent on ISIT in the Sorbonne: are both necessary for the text ? Are both necessary for the text ? Do we have to translate them literally from one language to another ? So what are the implications of literally translating them into the text ?**

Keywords: syntax, morphology, literally translation.

ما المعايير المُتبعة لترجمة القواعد والتراكيب النحوية ؟

أسماء جعفر عبد الرسول

كلية الآداب- جامعة المنوفية

gmasmaa@yahoo.com

الملخص **ـــ الهدف من هذا البحث هو دراسة أهمية القواعد والتراكيب النحوية وبناء الجملة في الترجمة. وسنميز المورفولوجيا المعجمية التى تهتم بدراسة أصل الكلمات عن المورفولوجيا الاشتقاقية التى تؤثر على بناء الجملة وكذلك بناء الجملة البسيطة من تلك المركبة. ولذلك، فإن إشكالية هذه الدراسة تكمن فى الإجابة على الأسئلة التالية من خلال النظرية التفسيرية التابعة لمدرسة المترجمين العليا بالسوربون : هل كلاهما ضروري للنص ؟ هل يجب علينا ترجمتهما حرفيا من لغة إلى أخرى ؟ ما هي إذن الآثار المترتبة على ترجمتهما حرفيا على النص ؟**

الكلمات المفتاحية : القواعد، التراكيب النحوية، الترجمة الحرفية.

السيرة الذاتية :

أسماء جعفر عبد الرسول

حاصلة على ليسانس آداب وتربية ثم على ليسانس آداب. حصلت على الماجستير فى الترجمة وتعمل حالياً مدرس مساعد بكلية الآداب، جامعة المنوفية. لقد شاركت بالبحث المعنون «الاختلاف الثقافى» فى المؤتمر الدولى للترجمة الذى أقيم فى المجلس الأعلى للثقافة والمجلس القومى للترجمة فى نوفمبر 2016. ومن جانب آخر، شاركت بالبحث المعنون «حركة الترجمة وتأثيرها على الأدب العربى» فى ملتقى العلاقات الثقافية الفرنسية-المصرية والذى أقيم فى المجلس الأعلى للثقافة يومى 21 و22 مايو 2017. ولها بحث منشور فى مجلة كلية الآداب، جامعة المنوفية، تحت عنوان «إعادة قراءة الروايات الفرنسية المترجمة إلى العربية فى الصحافة المصرية فى الفترة من 1881 حتى 1893. دراسة فى ترجمة الثقافة». وشاركت فى المؤتمر الدولى الثانى للغات الأوروبية بكلية الآداب، جامعة المنوفية المنعقد فى الفترة من 3 إلى 5 ديسمبر 2017، بالبحث المعنون «المترجم والوساطة الثقافية». وشاركت أيضاً فى المؤتمر السابع عشر لهندسة اللغة والذى أقيم فى جامعة مصر الدولية يومى السادس والسابع من ديسمبر 2017 بالبحث المعنون «دور السياق فى صياغة المعنى فى الترجمة». وشاركت فى المؤتمر الدولى الثانى عن التراث العربى والإسلامى، الذى أقيم بمعهد المخطوطات العربية يومى 21 و22 فبراير 2018، بالبحث الموسوم «حركة الترجمة فى جريدة الأهرام فى الفترة من 1881 حتى 1893». وشاركت فى مؤتمر «اتجاهات معاصرة فى دراسات المستعربين» والذى أقيم بكلية الآداب، جامعة القاهرة فى الفترة من 3 إلى 5 أبريل 2018 بالبحث المعنون «التباعد الزمنى والترجمة». وشاركت فى المؤتمر العلمى السابع لكلية الآداب بقنا جامعة جنوب الوادي العربية والدراسات الإنسانية والاجتماعية المنعقد فى الفترة من 11 إلى 13 نوفمبر بالبحث الموسوم«العربية واللغات الأخرى : دراسة فى التعريب والتأصيل والاشتقاق». شاركت أيضا فى 2019 فى عدة مؤتمرات تابعة لجامعة قناة السويس. وشاركت أيضاً فى المنتدى السادس لشباب الباحثين المنعقد فى مارس 2019. ورشحت لدورة تدريبية لغوية وتربوية من الحكومة الفرنسية ومقرها مدينة فيشى بفرنسا فى أغسطس 2019. عضوة فى عدة نشاطات تابعة للجمعية المصرية لأساتذة اللغة الفرنسية، وقد حضرت الكثير من المؤتمرات والندوات واللقاءات على هامش هذه الجمعية. وقد حصلت على شهادة تفيد بإجادتها للمستوى اللغوى *Delf B2* من وزارة التربية والتعليم الفرنسية. وسجلت مؤخراً حلقتين على النيل تى فى *NILE TV* فى البرنامج المعنون «***Bibliotheca***».

Asmaa Gaafar Abdel-Rassoul

She holds a Bachelor of Arts and Education and then a Bachelor of Arts.She holds a master’s degree in translation and is currently an assistant lecturer at the Faculty of Arts, Menoufia University. She participated by the research entitled "Cultural Difference" in the International Conference on Translation held in the

Supreme Council of Culture and the National Council for Translation in November 2016.On the other hand, she participated by the research entitled "The movement of translation and its impact on Arabic literature" at the French-Egyptian Cultural Relations Forum, which was held at the Supreme Council of Culture on May 21 and 22, 2017.She has a research published in the Journal of the Faculty of Arts, Menoufia University, entitled "Re-reading French novels translated into Arabic in the Egyptian press in the period from 1881 to 1893. A study in translating culture". She participated in the Second International Conference of European Languages, Faculty of Arts, Menoufia University, held from 3rd to 5th December 2017, with research entitled "Translator and cultural mediation". She also participated in the 17th conference for language engineering, which was held at Misr International University (MIU) on the 6th and7th of December 2017, with research entitled "The role of context in formulating meaning in translation".
She participated in the second international conference on the Arab and Islamic heritage, which was held at the Institute of Arabic Manuscripts on February 21 and 22, 2018, with the research titled “The translation movement in Al-Ahram newspaper from 1881 to 1893. She participated in the conference "Contemporary Trends in Arabist Studies", which was held at the Faculty of Arts, Cairo University from 3 to 5 April 2018 with research entitled "Time Spacing and Translation". And participated in the 7th scientific conference of the Faculty of Arts in Qena, South Valley University, Arab and humanities and social studies held from 11 to 13 November with research titled «Arabic and other languages: a study in Arabization and derivation». In 2019 she also participated in several conferences affiliated with the University of the Suez Canal. She also participated in the sixth forum for young researchers held in March 2019.She was nominated for a linguistic and educational training course from the French government based in Vichy, France in August 2019. She is a member of several activities of the Egyptian Association of French Language professors and has attended many conferences, seminars and meetings on the sidelines of this association. She obtained a certificate of proficiency in the language level Delf B2 from the French Ministry of Education. And recently recorded two episodes on the NILE TV in the program entitled «**Bibliotheca**».

Artificial Intelligence and NLP

Chatbot System Architecture

Moataz Mohammed[*1] and Mostafa M. Aref[**2]

[*]*Computer Science Department, Faculty of computer and Information Sciences, Ain-shams University*

[**]*Computer Science Department, Faculty of Computer and Information Sciences, Ain-shams University*

[1]moataz.mohammed@cis.asu.edu.eg

[2]aref_99@yahoo.com

Abstract— **The conversational agents is one of the most interesting topics in computer science field in the recent decade. Which can be composite from more than one subject in this field, which you need to apply Natural Language Processing Concepts and some Artificial Intelligence Techniques such as Deep Learning methods to make decision about how should be the response. This paper is dedicated to discuss the system architecture for the conversational agent and explain each component in details.**

Keywords: **Conversational Agents, Chatbots, System agent, Dialog System, Natural Language Understanding, Natural Language Processing, Deep Learning, System Architecture**.

2 INTRODUCTION

A chatbot (conversational agent (CA), dialogue system) is a computer software that acts as an interface between human users and a software application, using spoken or written natural language as the primary means of communication. Examples for that Apple Siri, Google Assistant and Amazon Alexa. In the recent decade most of companies throughout the world use it as a service to improve the way of communicating with clients in any time and the response is instantly which there is no delay. That provide the client with the comfortableness. So this essential technology service should be improved continuously. In this paper we will discuss the CA system architecture. There are many variant system architectures which the developers can follow them to develop the chatbot(conversational agent) but any one of them must have three significant core components: Natural Language Understanding(NLU) is responsible for understanding the user's utterances meaning and put them in representational format, then The Dialog Manager(DM) which is the most important component about acting as a mediator which receive the representational format from the NLU and process them .After that, send the responses for the Natural Language Generator(NLG) which is the last core component in any Conversational Agent(CA) architecture. It takes the responses from the Dialogue Manager (DM) and check if there are more than one valid response taking the one with highest priority. Finally, produce the response in the final format which may be text or Speech.

Social chatbots' appeal lies not only in their ability to respond to users' diverse requests, but also in being able to establish an emotional connection with users. The latter is done by satisfying users' need for communication. How to make the social chatbot more Human Like about Emotional Quotient (EQ) **[1]**. Also it can make an examination on the influence of its responsiveness and embodiment on the answers. people give in response to sensitive and non-sensitive questions **[2]**. AlI and deep learning can help us build such chatbots that improve the lives of people who have busy schedule to easily keep a check on their health **[3]**. The access to large-scale data and real-world feedback can drive faster progress in research **[4]**.

3 LITERATURE REVIEW

When we made a deep reading to a survey papers, an articles and journals. We found many conversational agents each one with its own architecture. Like Amazon Alexa bot which take the data of the user that may be voice and make an Automatic Speech Recognition ASR about the Amazon ASR service. Then make a processing to the received data about some Amazon Web Services AWS and using Amazon DynamoDB to store the conversations and its state**[4]**. Google Assistant the most successful bot at all.Which receive the recordings from users, then sending these recordings to google's servers which works on making processing. It makes break down the voices into individual sounds then try to match every single sound with the most similar word's pronunciation one which is stored on google database. Then it identifies the required task through some matched key words. And other more agents so we mentioned just two examples. From all these different architectures we will talk in this paper about the most commonly generic one which include three main components Natural Language Understanding, Dialog Manager, and Natural Language Generator.

4 SYSTEM ARCHITECTURE

In this section we will discuss each component in the architecture in details .We will talk about the three core components of any chatbot (Conversational Agent) and its subs, First Component is the Natural Language Understanding(NLU) and its role in the system. After that, we will talk about the subcomponents of the NLU such as Topic Detection, Intent Analysis, and Entity Linking. Second Component is the Dialog Manager (DM) and its subs such as Rule-base, Knowledge-based, Neural Network Reply Generation and the Online Information Retrieval. Finally the Natural Language Generation (NLG) or Reply Generator and it subs like Content Filter and Engagement Ranking. the following figure1 **[5]** can be considered as a simulation for these components.

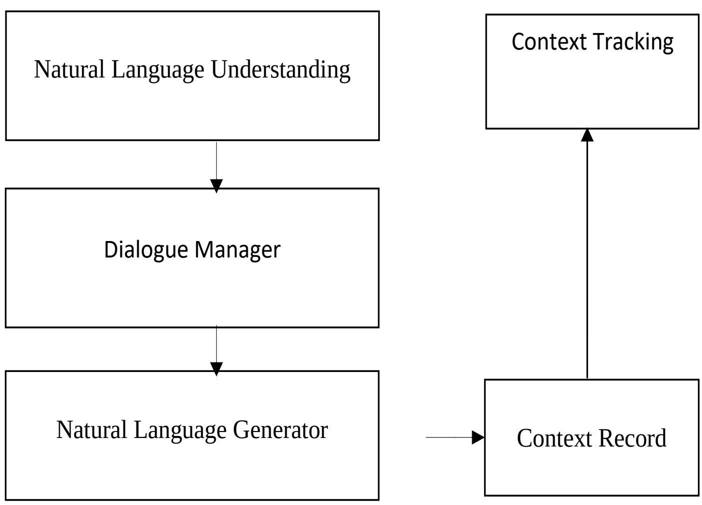


Figure 17: System Architecture of the Conversational Agent [e.g. chatbot]

A. Natural Language Understanding

Natural language understanding (NLU) is the first core component of the conversational agents which is responding about providing a semantic representation for user utterance **[6]** such as an in form of logic or class' intent, extracting the "meaning" of an utterance **[7]**. Parsing is the main task of - an NLU, which takes the string of words and provides a linguistic structure for the utterance. Implementation-dependent is the method which an NLU uses it to parse the input and can utilize context free grammars, pattern matching, or data-driven approaches as we can see in figure2. NLU outputs have to able to be tackled by a dialogue manager **[8]**.

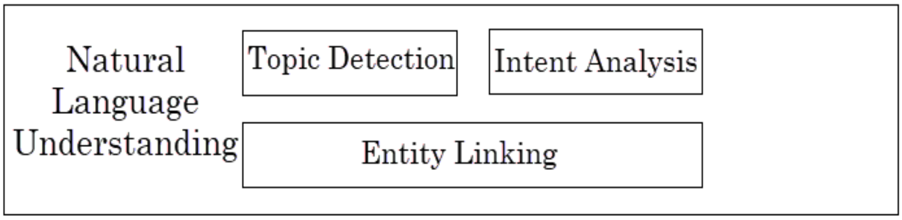


Figure 18: Natural Language Understanding [First Core Component]

1) *Topic Detection: from figure2. Topic Detection is One of the most important steps in developing the conversational agents NLU stage is to identify the Topic of the text (Conversation) for natural free-form interactions. Accurate tracking of the conversation topic can be a valuable signal for a system for dialog generation* **[9]**. *Examples on techniques can be used for topic classification are DANs and ADANs.*
2) *Intent Analysis: An intent is a group of utterances with similar meaning. Notice the following example of these two sentences: "I want to make a reservation in an Italian restaurant" and "I need a table in a pizzeria". They both have the same meaning. That means the intent analysis goal is to identify the similarities between words in meaning. If you can't understand a user, your bot will be useless whatever the effort you consume to develop the other components such as dialogue management. How we can Teach semantics to a machine. We can do this with word vectors algorithms such as Word2Vec or Glove.* [10]

3) *Entity Linking: the bottom component in figure2 consists of a Disambiguation model and Named Entity Recognition (NER) and a template selection model. NER [11] links entity mentioned in a text to a dictionary or knowledge base, local or remote such as the whole web, to make sense of an entity and know what it is. This is a significant step for allowing the chatbot to understand conversation topics and generate appropriate responses. As it links the words of the user's utterances with concepts and subtexts in the real word. Options are to use StanfordNLP, TAGME [12] or web mining via a search engine API. TAGME can take the input text and returns a list of entities with their concept titles from wikipedia, which in turn can be converted to nodes in the Wikidata knowledge graph.*

B. Context Tracking

When a sentence from a user appears, the chatbot obtains the most recent utterances of that user from the chat history database. The Stanford CoreNLP toolkit [13] can be used to resolve coreference. If a coreferent was identified, the pronouns and the mentions of entities in the new sentence will be replaced.

C. Dialogue Manager

Dialogue Manager is the second core component in any chatbot and we can differentiate the chatbots through this component which have many parts can be improved or adding some parts in the future if we discover that, it will serve the DM. DM receives a user input from the NLU and produce the system responses at a concept level to the natural language generator (NLG). the response which the DM will choose, depends on the strategy that has been chosen. Strategies are related to maintaining conversational state and the ability to model the dialogue structure beyond that of a single utterance **[6]**. the strategies are rule-based, knowledge-based, retrieval based, and generative. The rule-based strategies are backstory, intent templates, and entity-based templates ordered by their priorities. Because rule-based strategies encode human knowledge into the form of templates, they provide the most accurate responses. The system will adopt a template reply if input is recognized by one of these strategies. If there is no matching template for the input, the system can try to get an answer from a knowledge-based question answering (Q/A) engine **[5]**. Failing that, the input is handled by an ensemble of neural network models and information retrieval modules to create a general conversation output

.

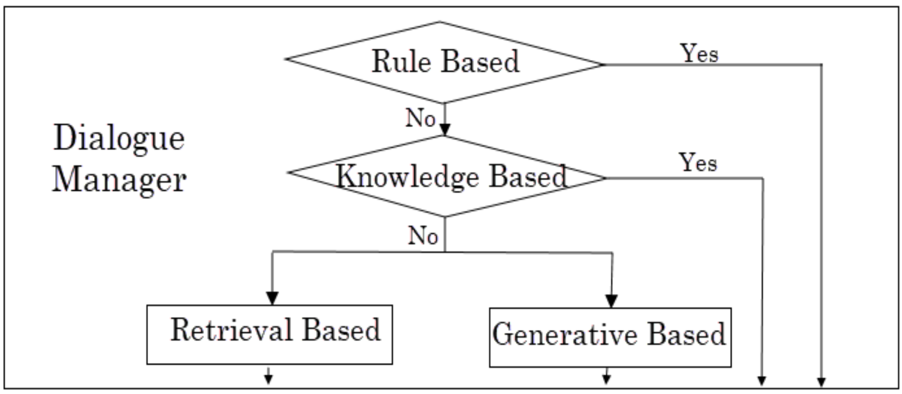


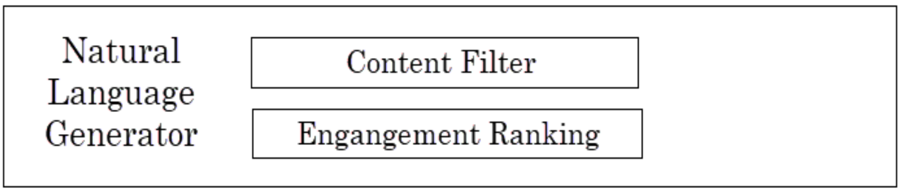
Figure 19: Dialogue Manager [Second Core Component]

1) *Online Information Retrieval: This module can be used when the rule-based and the knowledge-based is fail to provide responses ,it tries to provide more human-like, more concrete, and fresher responses compared to the rule-based templates and neural dialogue generation modules. The source of information for this module can be the most recent tweets provided by Twitter search API.**[14]** employed tweets as the source because they are usually short sentences closer to verbal language of most users compared to long written posts.*

2) *Neural Network Reply Generation: Generating sensitive context responses related to the conversation which helps make the user feel more engaging. This should be used for taking into account the recent utterances. Deep Learning approach and other Machine Learning techniques can be applied to develop.**[15]***

D. Natural Language Generator (Generator Reply)

The last main core component of any chatbot. It receives a communicative act from the Dialogue Manager (DM) and generates a matching textual representation. There are two functions that the NLG must perform: content planning (Content Filter | Engagement Ranking) and language Generation using just Text| Speech using Text to Speech). After going through one or more of strategies of the DM, the pipeline proceeds to the reply generator. This generator firstly will apply a content filter out incoherent or questionable candidates. If there are more than one valid response, a ranking process is used to reorder candidate utterances firstly according to the priority and then according to the engagement ranking. Finally, the chatbot will send the selected utterance as a text in the final output format or sending it to the Text to Speech to generate the final



output in Speech form. Simultaneously, all conversations are tracked in a history.

Figure 20: Natural Language Generator [Last Core Component]

5 CASE STUDY

KLM is a Dutch airline has an enormous number of flights around the world which was founded since 1919 so it was considered the world's oldest airline. KLM's owners thought about how they could improve their services for their clients. They found out that using the technology will be the key of success for this improvement. They are knowledge experts about having the information related to the area of study. They discovered that the social media is the most engagement way to the people. So they decided to ask the help from an AI experts about how they could benefit from the social media in the impact on the KLM airline services improvement.

A. Discussion

The main goal was to refining communication with their clients, making its client experience better by making it easier for people to talk to its agents via social media. KLM's owners studied their customers and knew that them spend a lot of time on social media like Facebook and Messenger. So, it can be the entry point and this about enabling the clients from sending any inquiry and may be reservation message privately to KLM's social media page. Because of the huge number of messages, the AI experts come with some bots such as KLM bot which provide as many as services like immediate reply on messages that can reach to 60000 message, Try before you fly with augmented reality, Tip advisor, and booking your favorite seat with more comfortability, privacy and more. You can use this bot through Facebook Messenger, Telegram, Twitter, and more other Social Media.

B. Evaluation

They made a survey at the end of each Messenger conversation, about asking the client some questions like "How you can feel comfortability with our Messenger bot?", "Are you like to recommend this bot for a friend?", and so on. And the answers were options to choose from them scaled from [1-10].
This Messenger bot achieved a breakthrough through both meeting and exceeding the expectations. Since January 2017, the airline has achieved:

1) Increasing in the interaction with client through the Messenger reached to 40%.
2) The success of online boarding booking was 15% about this Messenger bot.

6 Conclusion

The discussion was around the chatbot or CA. Firstly we talked briefly about what is the CA and then moved on the main purpose of the discussion, which is about what is the chatbot system architecture and its main components. After that, we discussed in details NLU and it subs, DM and it subs, then finally the NLG or reply generator. Notice that the DM is the most important part of any chatbot system because its components can be improved continuously or we can develop some alternative subs. If you are interested in chatbots, you can use this paper as a reference either in developing chatbot system or working on improving the Dialog Manager components.

REFERENCES

[1] Shum, H., He, X. & Li, D. From Eliza to XiaoIce: challenges and opportunities with social chatbots. Frontiers Inf Technol Electronic Eng. 19, 10–26 (2018).

[2] Schuetzler, Ryan M.; Grimes, G. Mark; Giboney , Justin Scott; and Nunamaker, J ay F . Jr., " The Influence of Conversational Agents on Socially Desirable Responding", "in *proc. of the 51St Hawaii International Conference on System Sciences", (HICSS),* University of Hawaii at Manoa, 2018

[3] S. Rai, A. Raut, A. Savaliya and R. Shankarmani, "Darwin: Convolutional Neural Network based Intelligent Health Assistant," Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2018, pp. 1367-1371.

[4] Ashwram, Roprasad, Ckhatri, Anuvenk, Raeferg, Qqliu, Jeffnunn, Behnam, Chengmc, Nashish, Kinr, Kateblan, Warticka, Yipan, Hasong, Skj, Ehwang, Pettigru "*Conversational AI: The Science Behind the Alexa Prize*" 1st Proceedings of Alexa Prize (Alexa Prize 2017), 2018

[5] Boris Galitsky, *Developing Enterprise Chatbots, Learning Linguistic Structures, Springer Nature Switzerland AG 2019*

[6] Jurafsky D., Martin J. H., Speech and language processing (Pearson International), 2nd edn. Pearson/Prentice Hall, Upper Saddle River. ISBN 978-0-13-504196-3, 2009

[7] Skantze G., "Error handling in spoken dialogue systems-managing uncertainty, grounding and miscommunication". Doctoral thesis in Speech Communication. KTH Royal Institute of Technology. Stockholm, Sweden,2007

[8] Lee C, Jung S, Kim K, Lee D, Lee G. G., "Recent approaches to dialog management for spoken dialog systems". *Journal of Computing Science and Engineering* (JCSE), 4(1):1–22,2010

[9] C. Khatri *et al*., "Contextual Topic Modeling For Dialog Systems, "*IEEE Spoken Language Technology Workshop (SLT)*, Athens, Greece, 2018, pp. 892-899. doi: 10.1109/SLT.2018.8639552, 2018

[10] Botfront.(2016), How intent classification works in NLU, https://botfront.io/blog/how-intent-classification-works-in-nlu#teaching-semantics-to-a-machine, (accessed 9 Jan 2020).

[11] Haptik Open source chatbot NER(2013) : https://haptik.ai/tech/open-sourcing-chatbot-ner/(accessed 12 January 2020)

[12] Ferragina P., Scaiella U. Tagme: "on-the-fly annotation of short text fragments (by Wikipedia entities)". In: Proceedings of the 19th ACM *international conference on information and knowledge management (CIKM). ACM*, New York, pp 1625–1628, 2010

[13] Manning CD., Surdeanu M., Bauer J, Finkel J, Bethard SJ, McClosky, “in proc. 52nd The Stanford CoreNLP
natural language processing toolkit*”. Annual Meeting of the Association for Computational Linguistics: System Demonstrations (ACLSD)*, pp 55–60, Baltimore, Maryland USA, June 23–24,2014
[14] Liu H, Lin T, Sun H, Lin W, Chang C-W, Zhong T, Rudnicky A. RubyStar: “A non-task oriented mixture model dialog system”. *First Alexa Prise comptions proceedings*, 2017
[15] *A. Sordoni, M. Galley, M. Auli, C. Brockett, Y. Ji, M. Mitchell, J.-Y. Nie, J. Gao, B. Dolan. A Neural Network Approach to Context-Sensitive Generation of Conversational Responses. In Proc. of NAACL-HLT. Pages 196-205,2015*

BIOGRAPHY

Moataz Mohammed is a teaching assistant of Computer Science department, Ain Shams University, Cairo, Egypt. Master candidate student in NLP and DL fields, B.Sc. of Computer Science Dept., Faculty of Computer and Information Sciences, in 2017, Ain Shams University, Cairo, Egypt.

Prof. Mostafa M. Aref is a professor of Computer Science and Chairman of Computer Science department, Ain Shams University, Cairo, Egypt. Ph.D. of Engineering Science in System Theory and Engineering, June 1988, University of Toledo, Toledo, Ohio. M.Sc. of Computer Science, October 1983, University of Saskatchewan, Saskatoon, Sask. Canada. B.Sc. of Electrical Engineering - Computer and Automatic Control section, in June 1979, Electrical Engineering Dept., Ain Shams University, Cairo, EGYPT.

بنية نظام المحادثة الآلية

مصطفى محمود عارف[2], معتز محمد[1]
*جامعة عين شمس, كلية الحاسبات والمعلومات, *قسم علوم الحاسب*
*جامعة عين شمس, كلية الحاسبات والمعلومات, **قسم علوم الحاسب*
[1]moataz.mohammed@cis.asu.edu.eg
[2]aref_99@yahoo.com

***نبذه مختصرة* - في واحدة من أكثر المجالات اهتماما في الفترة الأخيرة ألا وهي المحادثات الآلية حيث تستطيع ان تفتح البحث في أكثر من مجال في مجالات علوم الحاسب حيث انها تحتاج الى فهم كيف للحاسب ان يتعامل ويفهم لغة الإنسان وكيف يمكننا ان نجعل هذا الحاسب ان يطبق خصائص الذكاء الاصطناعي والتعلم العميق في إمكانية اخذ القرار والرد المناسب. في هذه الورقة البحثية سوف نناقش معا كيف يكون الهيكل البنائي لبناء أي محادثة آلية وسيتم هذا عن طريق مناقشة تفصيلية لكل مكون فيه.**

الكلمات الرئيسة:-
وكيل المحادثة الآلية, المحادثات الآلية, نظام وكيل المحادثة الآلية, نظام الحوار, فهم اللغات الطبيعية, معالجة اللغات الطبيعية, التعلم العميق, بنية النظام

Arabic Optical Character Recognition Using Sequence to Sequence Models

Mohamed Sobhi*[1], Yasser Hifny**[2], Saleh Mesbah***[1]

[1]*Arab Academy for Science, Technology and Maritime Transport, Alexandria, Egypt.*

[2]*University of Helwan, Cairo, Egypt.*

* mohamed.sobhi@bibalex.org
** yhifny@fci.helwan.edu.eg
*** saleh.mesbah@aast.edu

Abstract—Optical character recognition (OCR) software is used to convert scanned documents into text. Arabic OCR is an active area of research where high accuracy is demanding. This paper focuses on building a model for converting images that contain Arabic text into their corresponding text using a deep learning approach. This model does not require any knowledge of the underlying language and it is simply trained end-to-end on different datasets. It combines several standard neural components from vision and natural language processing. Features are extracted from images using Convolutional Neural Networks (CNNs) where the features are arranged in a grid. Each row is then encoded using a Recurrent Neural Networks (RNNs). An RNN decoder with a visual attention mechanism is used to generate the output text. Our preliminary experiments show that the presented approach is effective. The obtained accuracy is in the range of 97.5% - 99.1%.

Keywords: **Sequence to Sequence Model, Arabic OCR, Convolutional Neural Networks (CNNs), Recurrent Neural Network (RNN), Attention Mechanism, Long Short Term Memory (LSTM) .**

1 INTRODUCTION

Print disability represents an obstacle against some people concerning gaining information from printed material in an identical method that makes them rely upon alternative ways to access these printed data using special format (i.e. braille, large print, audio, digital text). Print disabilities involve Visual impairments, learning disabilities, and physical disabilities that prevent the ability to read a book in some way. Print disabilities guidelines suggest presenting information in various modalities, e.g. through vision, hearing or touch. These guidelines present an accessible format for users. For example, visual texts enable users to enlarge texts, amplify sounds, and click for supporting information like definitions and images.

To satisfy these requirements, there must be a high-quality application of optical character recognition OCR in order to analyze the picture of the page, find the letters and words and generate text file paragraphs and pages because the image cannot transfer into Braille or synthetic voice output directly [1].

Arabic OCR has many challenges such as the variations of Arabic characters according to their position in the word where every character may have two or four different forms. Thus, the number of classes will increase to be recognized from 28 to 100. Moreover the cursive nature of Arabic writing does not allow direct implementation of many algorithms formulated for other [2].

This paper presents a model that is built to transfer the Arabic text images to their corresponding texts depending on the recent deep-learning techniques. The model uses different Arabic datasets. Section 2 reviews the related work of OCR. Section 3 explains the used model for supporting the current proposed work. Section 4 describes the used datasets and how these data sets are preprocessed to be used in the current model. Section 5 details the experiments used for building the model. Section 6 reviews the results of the proposed work and finally. Finally, Section 7 concludes the paper.

2 RELATED WORK

This section provides the appropriate background on previous work on image to text generation. Recently, numerous methods have been proposed for generating the text from image. Many of these methods are based on recurrent neural networks (RNN) and inspired by the successful use of sequence to sequence Models with neural networks.

A synthetic text generation engine to train three different models for scene text recognition is proposed: 90K-way dictionary encoding, character sequence encoding and bag-of-N-grams encoding. The experiments showed that the synthetic dataset is realistic and sufficient that it can efficiently replace real data [3].
A method to eliminate the need of pre-processing stages and the use of multiple neural networks in OCR by training a single neural network is proposed [4]. The network is used to detect and recognize text relying on semi-supervised learning. The proposed network was firstly evaluated on SVHN dataset, then it was evaluated on ICDAR 2013, SVT and IIIT5K datasets with accuracies 90.3%, 79.8% and 86% respectively. Finally, it was evaluated on FSNS dataset with 97% accuracy for character recognition and 71.8% accuracy for words.
A framework for lexicon-free OCR is proposed [5]. The proposed method mainly uses: recursive CNNs for image feature extraction, RNNs for character-level language modeling with soft-attention mechanism. However, the proposed method did not use LSTM. The framework was evaluated on ICDAR 2003, ICDAR 2013, Street View Text and IIIT5K datasets with accuracies 88.7%, 90%, 80.7% and 78.4% respectively for unconstrained text recognition.
In [6], they demonstrated the convenience of sequence to sequence learning in OCR by proposing a recurrent encoder-decoder framework. Attention based model was not used, but the paper shows the effect of LTSM in the sequence to sequence learning. The proposed model outperforms LSTM with CTC output layer. However, it requires more memory space.
In [7], they proposed a method to identify multiple scripts at the same text-line level. The method is based on sequence learning model with LSTM capabilities, where a 1D-LSTM architecture is used. The model was evaluated on an English-Greek data and it gives 98.186% accuracy.
In [8], they presented WYGIYS (What You Get Is What You See) model for OCR of presentational markup. The model is based on convolutional networks with visual attention. They introduced a new dataset IM2LATEX-100K, besides a synthetic dataset of webpages paired with HTML, for training and evaluating the model. The model gives 75% accuracy.
In [9], they proposed a model based on sequence-to-sequence architecture and a convolutional network to recognize handwritten text. The proposed model mainly consists of: A convolutional network for features extraction, an LSTM-based RNN for encoding and another LSTM-based RNN with attention for decoding. The model was evaluated on IAM and RIMES databases on which it gives 12.7% and 6.8% for word error rate, respectively.
In [10], they introduced ASTER (Attentional Scene Text Recognizer with Flexible Rectification) to recognize text with distortions and irregular layout. ASTER is an end-to-end model that includes both a rectification network and recognition network. The rectification network is based on Spatial Transformer Network, while the recognition network is based on attentional sequence-to-sequence model. The proposed model is trained on two synthetic datasets: Synth90K and Synth-Text, and is evaluated on: IIIT5K, SVT, ICDAR 2003, ICDAR 2013, SVTP and CUTE with accuracies 92.67%, 91.16%, 93.72%, 90.74%, 78.76% and 76.39%, respectively.
In [11], they proposed a system of three-neural network models for Arabic OCR. The first network is a three-layer neural network to recognize the text font size, then the text is normalized to 18pt font size to be fed to the second two networks. The second network model is a multi-channel neural network with three-window input. Finally, the third network is a two-layer convolutional network with two max pooling layers. The whole pipeline was evaluated on APTI dataset and gives an accuracy of 94.8%.

3 MODEL

This model is mainly based on combining a number of standard neural components for vision and natural language processing. Firstly, it uses a Convolutional Neural Network (CNN) for features extraction from the image, which are then arranged in a grid. Secondly, it uses two Recurrent Neural Networks (RNN) for as an encoder-decoder. The first RNN encodes each row in the feature grid, then the second RNN, which is enhanced with a visual attention mechanism, decodes the encoded features. The full structure is illustrated in Figure1.

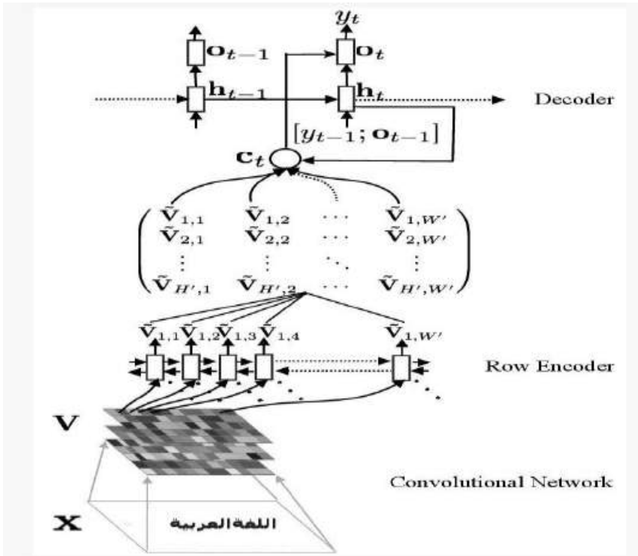


Figure 1. The full structure [12]

Model Structure:

- A CNN is used to extract a feature map V from the input image.
- Arrange the features in a grid.
- An RNN is used to encode spatial layout information for each row in the feature map.
- An RNN is used to decode with a visual attention mechanism to produce final outputs.

Convolutional Network: A Convolutional Neural Network (CNN, or ConvNet) is a sort of deep learning, feedforward Artificial Neural Network used extensively in analyzing visual image recognition tasks. It is mainly composed of an input layer, an output layer, and multiple hidden layers: convolutional layer, pooling layer, loss layer (drop out), fully connected layer (dense layer – Multi-layer Perceptron). [13] The visual feature extraction in this model is achieved through a multi-layer convolutional neural network interleaved with max-pooling layers. This network is considered a standard architecture nowadays. The convolution layer neurons execute a dot product between the image pixels and a filter. Then a Rectified Linear Unit (ReLu) and pooling layer follow each convolution layer. The ReLu is a non-linear activation function performed on each element to detect the negative values and threshold them to zero, such as max(0,x). For example: the output of ReLu (2, -3) will be (2,0). The pooling layer mainly reduces the dimensions of the image as the computation moves towards successive layers, and its output is then the input for down-sampling. The image-dimensions reduction results in reducing the number of parameters which helps in controlling the problem of overfitting. The CNN is designed such that pooling layers are interleaved in-between successive convolutional layers. The pooling layer acts on each features map independent of others and resizes it by performing a MAX operation. For example, in Figure 2, an image of dimensions 4*4 will be reduced by a pooling layer having a filter of dimensions 2*2 and stride 2. The output image from this layer will be 2*2, thus it has been reduced to 50% of its previous size. The ***max*** is performed to get the greatest number from the numbers that are in the filter's window [14].

$$\max(2, 1, 0, 3) = 3$$

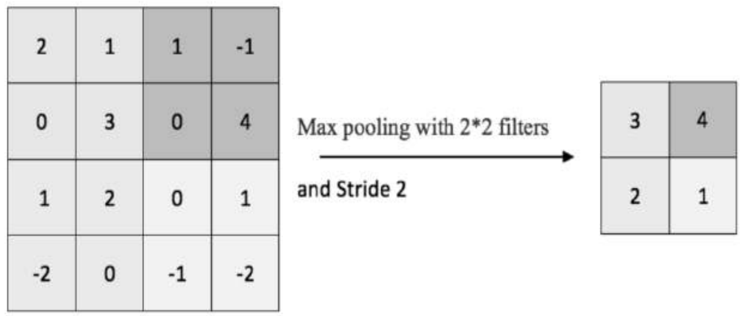


Figure 2: Pooling layer operation [14]

In this model, as we use visual attention model, we have to preserve the locality of CNN features, so final fully-connected layers are not used. The CNN accepts the raw input $R^{H \times W}$ and outputs a grid of features V of size $D \times H^1 \times W^1$, where D is the number of channels, H^1 is the reduced-sized grid height and W^1 is its width.

Row Encoder: Unlike image captioning, where the CNN features are used as they are, in OCR the encoder has to localize the relative positions within the image. This is achieved through running RNNs over each row of the CNN features.

A recurrent neural network (RNN) is a neural network characterized by the concept of 'internal memory', which differentiate it from other feedforward networks. The RNN internal memory is created by cyclic connections through its units. RNN is composed of three input layers (X_{t-1}, X_t, X_{t+1}), three hidden layers (h_{t-1}, h_t, h_{t+1}), three output layers (y_{t-1}, y_t, y_{t+1}), and weight matrices (W, U, V). The RNN accepts one input at time, for example, at time t the network is fed with input X_t, then it goes through the hidden layers for output prediction. The hidden layers represent the core of the network, the reason behind their importance is that they keep track of previous inputs. Each hidden layer accepts input from its previous hidden layer, in addition to its input unit, to predict the output. The weigh matrix of the hidden layer must be squared to maintain the same number of inputs as there are outputs.

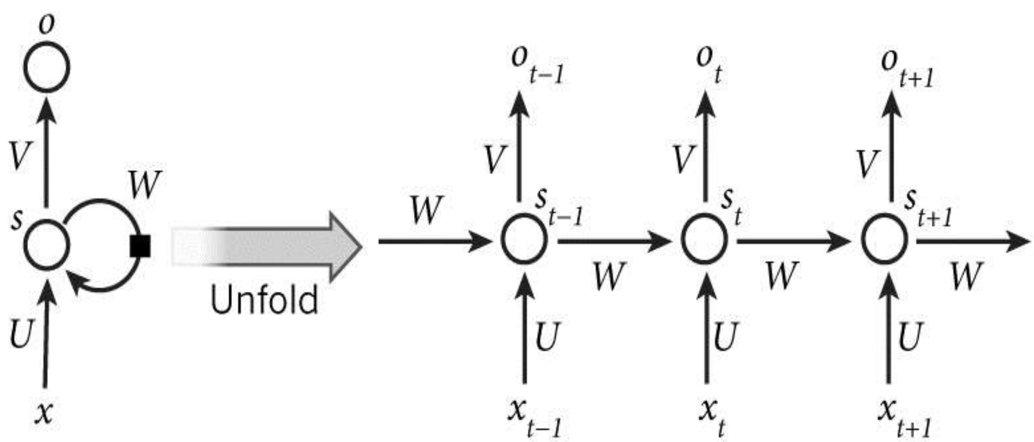


Figure 3: RNN Architecture [15]

The weight matrices (W, U, V) are initially set randomly. As for the first hidden layer (h_{t-1}), it is initialized by the dot product performed between its current input X_{t-1} at time t-1 and its weight matrix U, since it has no prior hidden layer and thus no contributions for the output prediction. The activation function such as sigmoid function then accepts the resulted dot product to generate values for the first hidden layer. Generally, data is processed in RNN by accepting an input X_t at time t, multiplying it with the weight matrix U and then passing it through hidden layer h_t. The hidden layer h_t also accepts another input from its prior hidden layer h_{t-1}, which has been parameterized with the weight matrix W. These two inputs contribute in predicting the output y_t, where a dot product is performed on the hidden layer h_t and the weight matrix V. This flow is repeated until all layers are covered to predict the final output.

Among the different variants of RNN, long short-term memory networks (LSTMs) [16] have shown effectiveness in the majority of NLP tasks. Thus, in this model the experiments use RNN with LSTM networks.

The long-short-term memory (LSTM) neural network is an extended version of simple RNN, where the past and current memories are related through linear dependence, instead of the non-linear connection between the past and current layers' activity. Furthermore, and most importantly, an LSTM has an introduced forget gate to adjust each of the past memory elements to contribute to the current memory cell.

Generally, each layer of the LSTM network accepts three inputs: X_t, h_{t-1} and C_{t-1}, which denotes the input at time t, the output of the previous LSTM unit, and the memory of the previous unity, respectively. The memory unit represents the most important unit in LSTMs, as it differentiates them from RNNs. The current layer has output h_t and the current unit has memory C_t. LSTMs have shown their effectiveness in captioning long term temporal dependencies. This obviously helped in improving the state-of-the-art for many difficult problems including: handwriting recognition and generation, language modeling and translation, acoustic modeling of speech, speech synthesis, protein secondary structure prediction and analysis of audio and video data among others. [17]

In this model, the new feature grid $\mathbf{V}$ is created from $\tilde{\mathbf{V}}$ by running an RNN across each row of that input. Recursively for all rows $h \in \{1, \ldots, H\}$ and columns $w \in \{1, \ldots, W\}$, the new features are defined as

$$\mathbf{V}_{hw} = \mathbf{RNN}\,(\mathbf{V}_{h,w-1}, \mathbf{V}_{hw}) \quad (1)$$

In order to capture the sequential order information in vertical direction, i use a trainable initial hidden state $\mathbf{V}_{h,0}$ for each row, which we refer to as positional embeddings.

This model uses an attention-based encoder-decoder composed of a pair of recurrent neural networks (RNNs). As for the encoder, it maps the variable-length input sequence to a vector. The decoder then maps the vector back to a variable length output sequence. To achieve maximum conditional probability of an output sequence given an input sequence, the two networks are trained jointly. [18]

Decoder Only considering the grid **V**, the decoder predicts the tokens of output text $\{y_t\}$. The decoder is trained as a conditional language model in order to predict the probability of the following token taking in consideration the history and annotations. The mentioned language model is described on top of a decoder RNN,

$$p(y_{t+1}|y_1, \ldots, y_t, \mathbf{V}) = \mathrm{softmax}(\mathbf{W}^{out}\,\mathbf{o}_t) \quad (2)$$

Where $\mathbf{o}_t = \tanh(\mathbf{W}^c[\mathbf{h}_t; \mathbf{c}_t])$ and $\mathbf{W}^{out}$, $\mathbf{W}^c$ are learned linear transformations. The vector h_t is used to summarize the decoding history: $\mathbf{h}_t = \mathbf{RNN}(\mathbf{h}_{t-1}, [y_{t-1}; \mathbf{o}_{t-1}])$. The context vector $\mathbf{c}_t$ is used to capture the context information from the annotation grid.

Attention Model: This model follows the mechanism of highly resolving a certain image region and fading the surrounds to focus on the specified region. Using the attention method eases the long-term dependencies modeling by adjusting the focal point over the period. This mechanism arises direct dependence between the model states at different time, which consequently introduces a hidden state h_t at each time step. [19]

4 DATA DESCRIPTION AND PREPROCESSING

The used datasets in the current work was selected from two sources. The first one is the LDC's Arabic Treebank PART 3 (ATB3). It is about 280,000 words that is select from "An-Nahar" Lebanese News. The second one is the Holy Quran without the Verses Numbers. Two versions of each dataset are used; diacritized and undiacritized with different formats.

Before beginning the experiments, each dataset is preprocessed as following:

1. A python script is used to separate the texts into lines where the maximum number of characters per each line is 80 characters by spaces. Each line is generated in Traditional Arabic with different sizes.
2. Another python script is used to generate the corresponding image of each line with different Dots Per Inch (DPI) to have different image qualities. A random name is generated to each image file and it is linked to its correspondent line.
3. A vocab list of the set of the date (unique characters) is generated.
4. Spaces in each line is marked by "<SP>" and a space is added after each character to be used in the proposed model.

5 EXPERIMENTS

For training the system, 70-80% of the images are selected from the data sets randomly and 10-15% of the images are selected for testing and validating the system. The training data sets are used in four experiments using Torch [20] based on the Open NMT (Neural Machine Translation) system [21]. These experiments are run on a 24GB NVIDIA Tesla K80 GPU [22].

In the first experiment 5197 images of the holy Quran without diacritics are used; 4158 images for training, 520 images for validation and 519 images for testing. The second experiment uses 8940 images of the holy Quran without diacritics; 6258 images for training, 1341 images for validation and 1341 images for testing. The third experiment aims at training the model on different format of the diacritized holy Quran as table 1 shows. Equal amount of images number for training, validation and testing are used in each adopted format. In the run time process, both data sets are combined in training, validation and testing.

TABLE VIII
EXPERIMENT 3 ON THE HOLY QURAN WITH DIFFERENT FORMATS

Experiment	Description	Total	Train	Val	Test
3	Holy Quran with diacritics and different format	8940	7152	894	894
3.1	Bold Italic Underline	2980	2384	298	298
3.2	Bold Italic Underline Strikethrough	2980	2384	298	298
3.3	Bold Italic Underline Double Strikethrough	2980	2384	298	298

For conducting the fourth experiment the LDC data set is used and it is dealt with twice. At first, the undiacritized LDC dataset with different format is used. Then, diacritized LDC data set with different format is used. In the run time process, both data sets are combined in training, validation and testing. As the total of the images is 39454, and Table 2 shows the division of these images.

6 RESULTS

The performance measures used to evaluate the system is the Word Error Rate (WER); the percentage of words has at least one error. A python script is developed to calculate the WER of the system depending on the Levenshtein distance [23] to calculate the insertion, deletion and substitution percentage. In the first experiment the rate yielded 99.92% of accuracy. In the second experiment the rate yielded 99.02 of accuracy and the third one yielded 98.55% of the rate (part one 97.49% - part two 99.58% - part three 98.57). The main problem in the second and third experiments' results is the insertion of diacritics for undiacritized words (Surah's name) which arise the need to conduct the fourth experiment that combines between diacritized and undiacritized words. As a result, the fourth experiment yielded to an accuracy of 97.56%. Some sample test images and their prediction are shown in Figure 4

TABLE II
EXPERIMENT 4 ON LDC WITH DIFFERENT FORMATS

Experiment	Description	Total	Train	Val	Test
4	LDC	39454	31564	3946	3944
4.1	without diacritics Regular	1640	1312	164	164
4.2	without diacritics Bold	1640	1312	164	164
4.3	without diacritics Italic	1640	1312	164	164
4.4	without diacritics Underline	1640	1312	164	164
4.5	without diacritics Double Strikethrough	1640	1312	164	164
4.6	without diacritics Bold Italic	1640	1312	164	164
4.7	without diacritics Bold Underline	1640	1312	164	164
4.8	without diacritics Bold Double Strikethrough	1640	1312	164	164
4.9	without diacritics Italic Underline	1640	1312	164	164
4.10	without diacritics Italic Double Strikethrough	1640	1312	164	164
4.11	without diacritics Underline Double Strikethrough	1640	1312	164	164
4.12	without diacritics Full Format	1687	1350	169	168
4.13	with diacritics Regular	1640	1312	164	164
4.14	with diacritics Bold	1640	1312	164	164
4.15	with diacritics Italic	1640	1312	164	164
4.16	with diacritics Underline	1640	1312	164	164
4.17	with diacritics Double Strikethrough	1640	1312	164	164
4.18	with diacritics Bold Italic	1640	1312	164	164
4.19	with diacritics Bold Underline	1640	1312	164	164
4.20	with diacritics Bold Double Strikethrough	1640	1312	164	164
4.21	with diacritics Italic Underline	1640	1312	164	164
4.22	with diacritics Italic Double Strikethrough	1640	1312	164	164
4.23	with diacritics Underline Double Strikethrough	1640	1312	164	164
4.24	with diacritics Full Format	1687	1350	169	168

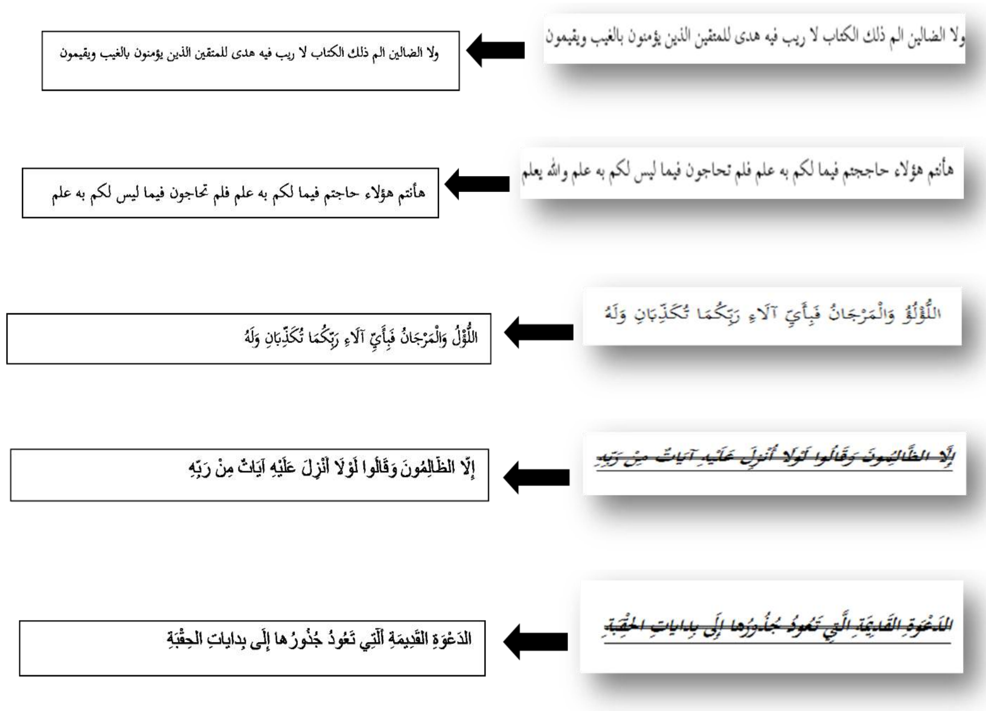


Figure 4. Samples of some scanned images that contain Arabic text into their corresponding text.

7 CONCLUSION

This paper introduces a deep learning framework for converting the scanned images that contain Arabic text into their corresponding text. This model does not require any knowledge of the underlying language. It is simply trained end-to-end on different datasets. Convolutional Neural Networks (CNNs) are used to extract salient features from images and an RNN decoder with a visual attention mechanism is used to generate the output text. The preliminary experiments show that the presented approach is effective. The obtained accuracy is from 97.5% to 99.1%. This work can be extended further in numerous ways, such as training this model to identify other fonts.

ACKNOWLEDGEMENT

The authors would like to thank Bibliotheca Alexandrina for granting the permission to use its computational resources provided by the High-Performance Computing (HPC). (hpc.bibalex.org)

REFERENCES

[1] M. Hersh and M. A. Johnson, Assistive Technology for Visually Impaired and Blind People, (2008).

[2] A. M. Zeki and M. S. Zakaria, "Challenges in Recognizing Arabic Characters," , *(*2004).

[3] M. Jaderberg, K. Simonyan, A. Vedaldi and A. Zisserman, "Synthetic Data and Artificial Neural Networks for Natural Scene Text Recognition," *arXiv preprint arXiv:1406.2227, (*2014).

[4] C. Bartz, H. Yang and C. Meinel, "STN-OCR: A single Neural Network for Text Detection and Text Recognition," *arXiv preprint arXiv:1707.08831, (*2017).

[5] C.-Y. Lee and S. Osindero, "Recursive Recurrent Nets with Attention Modeling for OCR in the Wild," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016).

[6] D. K. Sahu and M. Sukhwani, "Sequence to Sequence Learning for Optical Character Recognition," *arXiv preprint arXiv:1511.04176, (*2015).

[7] A. Ul-Hasan, M. Z. Afzal, F. Shafait, M. Liwicki and T. M. Breuel, "A sequence learning approach for multiple script identification," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, (2015).

[8] Y. Deng, A. Kanervisto and A. M. Rush, "What You Get Is What You See: A Visual Markup Decompiler.," *arXiv: Computer Vision and Pattern Recognition, (*2016).

[9] J. Sueiras, V. Ruíz, Á. Sánchez and J. F. Vélez, "Offline continuous handwriting recognition using sequence to sequence neural networks," *Neurocomputing,* vol. 289, pp. 119-128, (2018).

[10] B. Shi, M. Yang, X. Wang, P. Lyu, C. Yao and X. Bai, "ASTER: An Attentional Scene Text Recognizer with Flexible Rectification," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 41, no. 9, pp. 2035-2048, (2019).

[11] M. A. Radwan, M. I. Khalil and H. M. Abbas, "Neural Networks Pipeline for Offline Machine Printed Arabic OCR," *Neural Processing Letters,* vol. 48, no. 2, pp. 769-787, (2018).

[12] Y. Deng, A. Kanervisto, J. Ling and A. M. Rush, "Image-to-Markup Generation with Coarse-to-Fine Attention," in *International Conference on Machine Learning*, (2017).

[13] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," *Intelligent Signal Processing,* pp. 306-351, (2001).

[14] A. Karpathy, "Stanford university cs231n: Convolutional neural networks for visual recognition," 2018. [Online]. Available: http://cs231n. stanford. edu/syllabus. html/.

[15] D. Britz, "Recurrent neural networks tutorial, part 1–introduction to rnns.," (2015). [Online]. Available: http://www. wildml. com/2015/09/recurrent-neural-networkstutorial-part-1-introduction-to-rnns/. [Accessed 1 February 2019].

[16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation,* vol. 9, no. 8, pp. 1735-1780, (1997).

[17] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink and J. Schmidhuber, "LSTM: A Search Space Odyssey," *IEEE Transactions on Neural Networks,* vol. 28, no. 10, pp. 2222-2232, (2017).

[18] K. Cho, B. v. Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk and Y. Bengio, "Learning Phrase Representations using RNN Encoder--Decoder for Statistical Machine Translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (2014).

[19] C. Raffel and D. P. W. Ellis, "Feed-Forward Networks with Attention Can Solve Some Long-Term Memory Problems," *arXiv preprint arXiv:1512.08756, (*2015).

[20] R. Collobert, K. Kavukcuoglu and C. Farabet, "Torch7: A Matlab-like Environment for Machine Learning," in *BigLearn, NIPS Workshop*, (2011).

[21] G. Klein, Y. Kim, Y. Deng, J. Crego, J. Senellart and A. M. Rush, "OpenNMT: Open-source Toolkit for Neural Machine Translation," *arXiv preprint arXiv:1709.03815, (*2017).

[22] "nvidia Web Site," [Online]. Available: https://www.nvidia.com/en-gb/data-center/tesla-k80/. [Accessed 1 February 2019].

[23] K. U. Schulz and S. Mihov, "Fast string correction with Levenshtein automata," *International Journal on Document Analysis and Recognition,* vol. 5, no. 1, pp. 67-85, (2002).

Biography

Mohamed Sobhi

Sobhi is an Advanced Specialist in Assistive Technology Training at the Bibliotheca Alexandrina, he graduated from the faculty of Arts (Arabic Section) in 2004 and completed his premaster in linguistics in 2005. He also holds a master degree in Phonology with an overall grade: Excellent. In 2016 he completed a diploma in Information system and is currently a postgraduate researcher at the college of computing and Information Technology at the Arab Academy for Science, Technology and Maritime Transport.

Dr. Yasser Hifny

He is an associate professor at the University of Helwan, Egypt. Dr. Hifny received his M.Sc. and B.Sc. degrees in electronics and communication engineering in 1995 and 2000 respectively, from the University of Cairo, Egypt. In 2006, he received the PhD degree from the University of Sheffield, U.K. In 2000, he joined Research and Development International (RDI), where he worked on research in text-to-speech (ArabTalk), limited domain speech compression, and speech verification (HAFS) projects. In addition, he was a postdoctoral fellow in the Human Language Technologies Group at IBM T.J. Watson Research Center in 2007.

Dr. SALEH MESBAH ELKAFFAS

Dr. Saleh Mesbah Elkaffas is an Associate Professor at the Department of Information Systems, College of Computing and Information Technology, Arab Academy for Science, Technology and Maritime Transport (AAST). He is currently the director of the Remote Sensing and Spatial Studies Unit, College of Engineering and Technology, AAST. He received his B.Sc. in Electronics and Telecommunications Engineering and M.Sc. in information content in remote sensing. He received his Ph.D. degree under the channel system supervision from the Center for Remote Sensing, Imperial College, UK, and Alexandria University. He has over 35 years of experience in the field of Remote Sensing, GIS, Computing and Information Technology. Over the span of his career, Dr. Saleh Mesbah has published a number of papers and book chapters in the fields of Remote Sensing, GIS, Digital Image Processing, Advances in Intelligent Systems and Computing, Applied Geomatics, Sensing and Imaging and Environmental Management. His research interests include Satellite Remote Sensing, Digital Image Processing, Geographic Information Systems (GIS), Global Positioning Systems (GPS), Information Technology and Environmental Manageme

محمد صبحي *[1] ، ياسر حفني **[2] ، صالح مصباح ***[1]

1 الأكاديمية العربية للعلوم والتكنولوجيا والنقل البحري ، الإسكندرية ، مصر.

2 جامعة حلوان ، القاهرة ، مصر.

mohamed.sobhi@bibalex.org *

yhifny@fci.helwan.edu.eg **

saleh.mesbah@aast.edu ***

الخلاصــة — يسـتخدم برنامج التعرف الضـوئي على الأحرف (OCR) لتحويل المسـتندات الممسـوحة ضـوئيًا إلى نص. تعد التعرف الضـــوئي على الحروف العربية مجالًا نشـــطًا للبحث حيث تتطلب الدقة العالية. توقدم هذه الورقة بناء نموذج لتحويل الصور عربي الى النص المقابل لها باستخدام نهج التعلم العميق. لا يتطلب هذا النموذج أي معرفة باللغة الأساسية ويتم تدريبه ببسـاطة من طرف إلى طرف على مجموعات بيانات مختلفة. فهو يجمع بين العديد من المكونات العصـبية القياسـية من الرؤية ومعالجة اللغة الطبيعية. يتم اسـتخراج الميزات من الصـور باسـتخدام الشـبكات العصـبية التلافيفية (CNNs) حيث يتم ترتيب الميزات في شبكة. ثم يتم تشفير كل صف باستخدام الشبكات العصبية المتكررة (RNNs). يتم استخدام وحدة فك ترميز RNN مع آلية الاهتمام البصـري لإنشـاء نص الإخراج. تظهر تجاربنا الأولية أن النهج المقدم فعال. الدقة التي تم الحصـول عليها هي في حدود 97.5 ٪ الى 99.1 ٪.

الكلمات المفتاحية: **نموذج التسلسل إلى التسلسل ، التعرف الضوئي على الحروف العربية ، الشبكات العصبية التلافيفية (CNN) ، الشبكة العصبية المتكررة (RNN) ، الاهتمام البصري.**

Whale Swarm Algorithm Methodology for Text Mining

Amr M Sauber[*1], Passent M. El-Kafrawy[*,**2], Amr. F. Shawish[*3]

[*]*Faculty of Science, Menoufia University, Egypt*
[1]amr@science.menofia.edu
[2]basant.elkafrawi@science.menofia.edu.eg
[3]*amr_shawish@yahoo.com*

[**]*School of Information Technology and Computer Science, Nile University, Egypt*
[2]pelkafrawy@nu.edu.eg

Abstract: **growing nature-inspired meta-heuristic algorithms are used to resolve real-world optimization issues, as they have some benefits over the classical techniques of numerical optimization. This paper explains the methodology of nature-inspired meta-heuristic called Whale Swarm Algorithm (WSA) for feature optimization, which is inspired by the whales' behavior of communicating with every different other via ultrasound for hunting. Text mining is used in every field for business intelligence, social media analysis, sentiment analysis, biomedical analysis, software process analysis and even for security analysis. This paper discusses different application of WSA as an optimization algorithm specifically the automation of understanding of Arabic text into ontology construction in the state-of-the-art literature. The key issues which are involved in the WSA enhancement models are also discussed here. This paper presents an up-to-date review over the uses of WSA in different fields to improvement of swarm optimization applications with focus on ontology learning from Arabic text.**

Key words: **optimization, swarm intelligence, whale swarm algorithm, whale optimization algorithm, ontology, Arabic text mining, concepts mapping.**

1 INTRODUCTION

Optimization is utilized in different applications. In the manufacturing of a new device, in a new artificial intelligence method, in big data application or in deep learning network, optimization is the most vital phase of any application. In order to develop a device with optimum sizes utilizing minimal power, to train a network, in order to limit the desired between the desired output and actual output values, optimization is desired. Text Mining has become an important research area. Text Mining is the discovery by computer of new, previously unknown information, by automatically extracting information from different written resources. Text mining can work with unstructured or semi-structured data sets such as emails, full-text documents and HTML files etc. Text Mining is widely used in field of Natural Language Processing and Multilingual Aspects. The Data Mining Optimization Ontology (DMOP) has been developed to support informed decision-making at various choice points of the data mining process. An evolutionary approach that combines information extraction technology and genetic algorithms can produce a new, integrated model for text mining. Text mining discovers unseen patterns in textual databases. The major uses of a text mining tool are for: Text Analytics, Text Processing, Classification/Categorization, Sentiment Analysis and Knowledge Discovery. Genetic Algorithms are the algorithms used to solve optimization problems. These algorithms are search based algorithm used to generate useful solutions for search problems [1].

Nature-inspired algorithms are becoming vigorous in solving numerical optimization issues, specially the NP-hard issues for example, the travelling salesman issue [3], vehicle routing [4], classification issues [5], routing issue of wireless sensor networks (WSN) [6] and multiprocessor scheduling issue [7], etc. These real world optimization issues often likely come with more global or local optima of a given mathematical model. Whereas, a point-by-point classical technique of numerical optimization is utilized for this task, the classical technique has to try repeatedly for locating various optimal solutions in every iteration [8], which takes a lot of time and work. Thus, using nature-inspired meta-heuristic algorithms to solve these issues has become an important research topic, as they are simple to execute and can converge to the global optima with high probability. In this research, we shall discuss a new nature-inspired meta-heuristic called Whale Swarm Algorithm (WSA) or Whale Optimization Algorithm (WOA) for function optimization; rely on the whales' behavior of communicating with every other through ultrasound for hunting. Hence, a brief overview of the nature-inspired meta-heuristic algorithms is explained. GA and WSA where merged in a new methodology, G-WSA, to extract concepts in text mining. The approach where utilized to construct Arabic ontologies from Arabic text [35]. The concepts where automatically extracted by optimizing the identification of related concepts and their relationships to parent concepts.

In this paper after the introduction in section 2 the natural phenomena of Whales Swarm will be retrieved, then detailed definition of the WSA is explained with the algorithm details and the mathematical background is defined in sections 3 and 4 consequently. In section 5 applications of WOA is listed with related state of the art publications. Finally, a discussion and conclusion in sections 6 and 7.

2 BACKGROUND

A. Whale Hunting Behavior

Social animals which live in groups inside the sea are called Whales. They make various sounds to demonstrate their movement, sustaining and mating designs. Whales decide nourishment azimuth and stay in contact with one another from enormous separations by ultrasound.

For example, pregnant females will assemble with other female whales and calves in order to improve protection abilities. Also, sperm whales are frequently seen in gatherings of somewhere in the range of 15 to 20 population, as shown in figure. 1. The whale sounds are delightful tunes in the sea and their sound range is exceptionally wide. As of not long ago, researchers have found 34 types of whale sounds, for example, whistling, squeaking, moaning, yearning, thundering, chattering, clicking, humming, churring, talking, trumpeting, clopping, etc. These sounds made by whales can frequently be connected to significant capacities, for example, their relocation, encouraging and mating designs. What's more, whales decide food azimuth and keep in contact with one another from an extraordinary separation by the ultrasound which are past the extent of human hearing [2].

Figure 1: The swarm of sperm whales.

B. Swarm Behavior

Swarming, or swarm behavior is an aggregate swarm behavior shown by creatures of comparative size which total together, maybe processing about a similar spot or maybe moving as a group or relocating toward some path. As a term, swarming is connected especially to mealy bug, likewise can be connected to some other creature that shows swarm conduct. The term rushing is typically used to allude explicitly to swarm conduct in fowls, grouping to allude to swarm conduct in quadrupeds, shoaling or tutoring to allude to swarm conduct in fish. Phytoplankton likewise assembles in tremendous swarms called blossoms, despite the fact that these living beings are green growth and are not self-impelled the manner. The term swarm is connected likewise

to lifeless substances which display parallel practices, as in a robot swarm, a seismic tremor swarm, or a swarm of stars. From an increasingly theoretical perspective, swarm conduct is the aggregate movement of an enormous number of self-pushed elements. From the viewpoint of the numerical modeler, it is a developing conduct emerging from straightforward principles that are traced by people and does not include any focal coordination. Swarm conduct was first reproduced on a PC in 1986 with the reproduction program boids. This program recreates basic operators that are permitted to move as per a lot of fundamental standards. The model was initially intended to impersonate the rushing conduct of winged animals; however it tends to be connected additionally to trained fish and other swarming elements [2].
The boids PC program, made by Craig Reynolds in 1986. Many consequent and current models use minor departure from these guidelines, frequently executing them by methods for concentric "zones" around every animal. In the "zone of aversion", near the creature, the central creature will look to remove itself from its neighbors to maintain a strategic distance from impact. Second, in the "zone of alignment", the central creature will try to adjust its heading of movement to its neighbors. Third, "zone of fascination", which reaches out as far away from the central creature as it can detect, and the central creature will try to move towards a neighbor.

C. Swarm Intelligence

Swarm Intelligence (SI) is a sort of man-made consciousness that intends in order to mimic the conduct of swarms or social mealy bug. Swarm alludes to any inexactly organized accumulation of cooperating specialists. Actually swarms are viewed as decentralized self-sorted out frameworks. Swarm knowledge has a multidisciplinary character its investigation gives bits of knowledge that can enable people to oversee complex frameworks. There is no reasonable definition for swarm insight. Developing conduct, self-sorted out conduct and aggregate knowledge are the related terms. Shockingly swarm insight framework can act in a planned manner with no facilitator or outside controller.

3 WHALE SWARM ALGORITHM

A. Overview of Whale Swarm Algorithm

Whale swarm algorithm is developed for comprehending work enhancement issue, have romanticized a few chasing principles of whale. At the point when a whale has discovered food source, it will make sounds to tell different whale's close-by of the quality and amount of food. So every whale will get loads of warnings from the neighbors and after that transition to the legitimate spot to discover food dependent on these warnings. The behavior of whales connecting with one another through sound for hunting fill with us to build up another meta-heuristic algorithm for function optimization issues as shown figure 2.

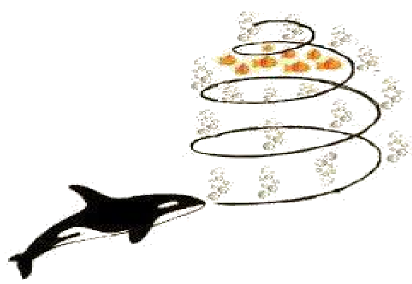


Figure 2: Method of WOA for accessing best solution

B. Whale Swarm Algorithm Rules

There are four rules are very important in building successful whale optimization algorithm as follows:

1. All the whales should connect with every other through ultrasound in the search region;
2. Every whale has a certain degree of computing capability in order to calculate the distance to other whales
3. The quality and quantity of food found through every whale is assigned to its fitness.
4. The motion of a whale is guided through the nearest one amongst the whales that are better (judged by fitness) than itself.

The flow chart of whale Swarm Algorithm is explained in figure 3 as follows:

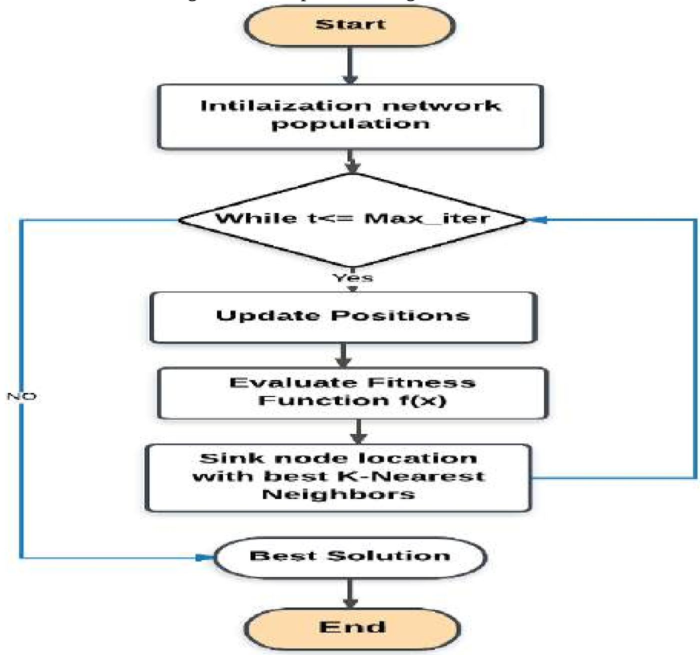


Figure 3. The flow chart of WSA technique

4 THE MATHEMATICAL MODEL OF WOA

The mathematical model for whale optimization algorithm is presented in details. The following functions describe the behavior of WOA to achieve encircling prey, feeding of spiral bubble-net maneuver, and search for prey.

The WOA algorithm supposes that the current best candidate solution is the target prey or is close to the optimum. After the best search agent is defined, the different search agents will hence attempt to update their positions in the direction of the best search agent. This conduct is represented with the aid of the following equations [9]

$$\vec{D} = |\vec{C}\,.\,\vec{X}^*(t) - \vec{X}(t)| \qquad (1)$$

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A}\,.\,\vec{D} \qquad (2)$$

Where t suggests the current iteration, A and (C) are coefficient vectors, X* is the position vector of the best solution acquired so far, X is the position vector, | | is the absolute value, and • is an element-by-element multiplication. It is well worth citing here that X∗ should be updated in every new release if there is a better solution.

The vectors $\vec{A}$ and $\vec{C}$ are calculated as follows:

$$\vec{A} = 2\vec{a}\,.\,\vec{r} - \vec{a} \qquad (3)$$

$$\vec{C} = 2.\vec{r} \qquad (4)$$

Where (a) is linearly reduced from 2 to 0 over the direction of iterations (in each exploration and exploitation stages) and r is a random vector in [0, 1].

For a 2D issue, the position (X, Y) of a search agent can be up to date according to the position of the current best record (X *, Y*). Various places round the best agent can be done with respect to the current position through using adjusting the value of A and (C) vector. It should be mentioned that with the aid of

defining the random vector (r) it is viable to reach any function in the search space defined among the key-points. Therefore, Eq. (2) permits any search agent to update its position in the neighborhood of the current best solution and simulates encircling the prey.
The same concept can be prolonged in order to a search space with n dimensions, and the search agents will go in hyper-cubes round the best solution got yet. As above-mentioned in the previous part, the humpback whales as well attack the prey with the bubble-net technique.
Then, Calculates the distance between the whale placed at (X, Y) and prey placed at (X*, Y*). A spiral equation is then created between the position of whale and prey to mimic the helix-shaped motion of humpback whales as follows:

$$\vec{X}(t+1) = \vec{D'} \, . \, ebl. \cos(2\pi l)+ \vec{X}^*(t) \qquad (5)$$

Where (D') = |X ^* (t) - (X() t)| shows the distance of the i-th whale to the prey (best solution acquired yet), b is a regular for defining the shape of the logarithmic spiral, l is a random quantity in [−1, 1], and * is an element-by-element multiplication.
Humpback whales swim round the prey inside a shrinking circle and among a spiral-shaped path with each other. To mannequin this simultaneous behavior, we suggest that there is a probability of 50% to select between either the shrinking encircling mechanism or the spiral model to update the position of whales throughout optimization. The mathematical model is as follows:

$$\vec{X}(t+1) = \begin{cases} \vec{X}^*(t) - \vec{A} \, . \, \vec{D} & \text{if } P < 0.5 \\ \vec{D'} \, . \, e^{bl} \, . \cos(2\pi l) + \vec{X}^*(t) & \text{if } P \geq 0.5 \end{cases} \qquad (6)$$

Where p is a random number in [0, 1].
Humpback whales search randomly in accordance to the position of every other. Thus, we utilize A with the random values greater than 1 or less than −1 to force search agent to get about far away from a reference whale. In contrast to the exploitation stage, we update the position of a search agent in the exploration phase in accordance to a randomly chosen search agent instead of the best search agent found yet. This mechanism and | A | > 1 confirm exploration and permit the WOA algorithm to execute a global search. The mathematical model is as follows:

$$\vec{D} = |\vec{C} \, . \, \overrightarrow{X_{rand}} - \vec{X}(t)| \qquad (7)$$

$$\vec{X}(t+1) = \overrightarrow{X_{rand}} - \vec{A} \, . \, \vec{D} \qquad (8)$$

where $\overrightarrow{X_{rand}}$ is a random position vector (a random whale) chosen from the current population.

5 USES OF WOA

There are many uses for whale swarm optimization algorithm in various areas of life. Stochastic nature-inspired meta-heuristic algorithms have validated their power on the closing two decades in dealing with global optimization issues bobbing up often in engineering. Whale optimization algorithm is a recent swarm based meta-heuristic. It copies the pattern of spiral bubble net hunting pattern of humpback. Mohit Verma and Amit Kumar (2018) [10] offered a brief survey on the whale optimization algorithm with focus on its several applications over single objective and multi-objective optimization issues. First of all, the goal of single objective optimization is to attempt to uncover the exceptional solution that corresponds to the minimal or the most well worth of one objective that lumps all definitely various objectives into one. WOA has been utilized for the various single objective optimization issue. Table I lists some of the application of WOA in solving single-objective optimization issues.

C. Engineering Applications

TABLE I: APPLICATION OF WOA IN SINGLE-OBJECTIVE OPTIMIZATION

S. No.	Application name	Reference
1.	Bearing fault diagnosis using a WOA	[11]
2.	Welded beam design	[12]

3.	The Workflow Planning of Construction Sites using WOA	[13]
4.	Optimal siting of capacitors in radial distribution network using WOA	[14]
5.	A hybrid WOA and pattern search technique for optimal power flow problem	[15]
6.	Combined Emission Constrained Economic Dispatch with Valve Point Effect Loading Problem Solution using WOA	[16]
7.	An emission constraint environment dispatch problem solution with microgrid using WOA	[17]

On the other hand, there exists no most suitable solution in case of multi-objective optimization with contradictory objectives [18]. The co-occurrence of definitely distinctive targets leads to a collection of trade-off solutions popularly known as non-dominated or Pareto-optimal solutions. Multi-Objective Whale Optimization Algorithms (MOWOA) has been utilized for the different issues with multi objective optimization. Table II lists some of the applications of WOA in solving multi objective optimization issues.

TABLE II: APPLICATION OF WOA IN MULTI-OBJECTIVE OPTIMIZATION

S No	Application	Reference
1.	Economic and Emission Dispatch using WOA	[19]
2.	Multi-Objective Optimal Vehicle Fuel Consumption based on WOA	[20]
3.	Multi-objective optimal mobile robot path planning base on WOA	[21]
4.	An Ameliorative WOA for Multi-Objective Optimal Allocation of Water Resources	[22]
5.	A MOWOA for Solving Engineering Design Problems	[23]
6.	WOA for combined heat and power economic dispatch	[24]

D. Feature Selection Applications

Choosing of relevant benefits of a dataset is vital in high dimensional datasets to keep away from the curse of dimensionality. Feature Selection is carried out to decrease overfitting, to enhance accuracy and to decrease the training time of the algorithms. Bing Zeng, Liang Gao and Xinyu Li used WSA for feature subset selection [2]. P. Anuradha and Dr. Vasantha Kalyani David (2019) [25] focused on choosing a features subset utilizing Whale Swarm Algorithm (WSA) where Logistic Regression (LR), Random Forest (RF) and k-Nearest Neighbor (KNN) are utilized as fitness functions. These WSA-LR, WSA-RF and WSA-KNN combinations generate different feature subsets for different number of iterations. Then training and trying out is accomplished on the dataset with the subset of selected features using LR, RF, Support Vector Classifier (SVC) and Gaussian Naive Bayes (GNB) and prediction accuracies generated are analyzed.

In High dimensional datasets, Feature Selection pursuits at decreasing the redundant and irrelevant features. The refined dataset with only the relevant features would enhance the learning accuracy and decrease the learning time [25]. The features which are utilized to train the machine learning model highly influence the efficiency of the model. Irrelevant features can bring down the efficiency of the model. Feature Selection can be greatly labeled into filter technique, wrapper method and Embedded method. In Filter technique, different statistical tests are utilized to choose the features that rely on their correlation with the outcome or dependent variable. In wrapper technique, a subset of features is utilized to train a model. Based on the efficiency of the model, features will be added or deleted to/from the subset. In embedded technique, both the benefits of filter and wrapper techniques are combined. The embedded technique algorithm operates using subset selection, train a model and additionally execute a penalization function to limit overfitting.

The pseudo code of finding a whale's better and nearest whale [2]:

Input: The whale swarm n, a whale u.

Output: The better and nearest whale u.

1: begin

```
Define an integer variable v initialized with 0;
Define a float variable temp initialized with infinity;
for i=1 to n do
if i≠u then
    if f(whale i )
        if dist(whale i , whale u )
            v=i;
            temp=dist(whale i , whale u );
        end if
    end if
  end if
end for
return whale v;
end
```

The authors experimented three various classifiers for the fitness function namely, Logistic Regression, Random Forest and K-Nearest Neighbors and the subsets are obtained from various numbers of iterations. The dataset rely on the chosen subsets are then utilized for classification. The classification accuracy of four various classifiers namely, Random forest, Logistic Regression, Support Vector Classification and Gaussian Naive Bayes are compared. Among these WSA with Logistic Regression as the fitness function (WSA-LR) gives a subset of eight features on an average and the accuracy of Random Forest Classifier is found to be 85.7% which is better than the other classifiers. In future, other classifiers can be tried as fitness function in the WSA, and the prediction accuracy can be compared among other classifiers [26].

E. Clustering Applications

Clustering is a powerful method in data-mining, which entails identifying homogeneous corporations of objects based totally on the values of attributes. Meta-heuristic algorithms for example, particle swarm optimization, artificial bee colony, genetic algorithm and differential evolution are now becoming powerful techniques for clustering.

Clustering is aggregating unlabeled objects into corporations with similarities between these objects. Such that the objects in the identical clusters are extra similar to every different object in distinct clusters in accordance to some predefined criteria [27] and [28]. A variety of algorithms have been proposed that take into account the nature of the data, the volume of the information and different enter parameters in order to cluster the data. The similarity standards in clustering are a range of in different researches. Most of the clustering troubles have exponential complexity in terms of the quantity of clusters.

Lately, Mirjalili and Lewis [29] described a new swarm based meta-heuristic optimization algorithm that mimicks the social behavior of humpback whales in searching. The algorithm is inspired through the bubble net hunting delineation. They have tested the WOA algorithm with 29 mathematical benchmark optimization issues and compared the efficiency of WOA algorithm with other traditional current heuristic algorithms for example, Particle Swarm Algorithm (PSO) [29], Differential Evolutional(DE) [31], Gravitational Search Algorithm (GSA) [32] and Fast Evolutionary Programming [33]. WOA was identified to be ample aggressive with different general and popular meta-heuristic techniques.

Jhila Nasiri1 and Farzin Modarres Khiyabani (2018) [34] proposed a new meta-heuristic clustering technique, the Whale Clustering Optimization Algorithm, primarily rely on the swarm foraging behavior of humpback

whales. After a detailed formula and explanation of its implementation, they compared the proposed algorithm with different existing well-known algorithms in clustering, including PSO, Artificial Bee Colony (ABC), GA, DE and k-Means. Proposed algorithm was once examined with the usage of one synthetic and seven real benchmark data units from the UCI computer mastering repository. Simulations exhibit that the proposed algorithm can efficiently be utilized for data clustering.
The consequences of their algorithm were contrasted with customary k-means clustering strategy and other popular stochastic algorithms such as PSO, ABC, DE, and genetic algorithm (GA) clustering. The Preliminary computational experience in terms of the intra-cluster distance function and standard deviation revealed that the whale optimization algorithm can successfully be applied to solve clustering issues. Furthermore, the results from the proposed algorithm was once effective, simple to execute and robust as compared with different strategies. There are some directions that can enhance the overall performance of the suggested algorithm in the future. The aggregate of WOA clustering algorithm with different clustering strategies and the usage of different fitness functions in clustering strategy should be considered in future researches.

F. Text Analysis

Rania M. Ghoniem et al. (2019) [35] proposed an optimized ontology learning from Arabic text. Ontology is a technique for extending web syntactic interoperability to semantic interoperability. Ontologies are exploited to signify massive information in such a way that permits machines to interpret its meaning, allowing it to be reused and shared. Their work was done in two phases. First, a text mining algorithm is proposed for extracting concepts and their semantic relations from text documents. The proposed algorithm calculated the *concept frequency weights* using the term frequency weights. Afterwards, they calculated the weights of thought similarity utilizing the facts of the ontology structure, involving (i) the concept's route distance, (ii) the concept's distribution layer, and (iii) the mutual mother or father concept's distribution layer. Then, feature mapping is carried out via assigning the concepts' similarities to the concept features. The second phase, a hybrid genetic-whale optimization algorithm was once proposed to optimize ontology learning from Arabic text. The operator of the G-WOA is a hybrid operator integrating GA's mutation, crossover, and selection processes with the WOA's procedures to achieve the stability between both exploitation and exploration, and to locate the solutions that showcase the best possible fitness. For estimating the overall performance of the ontology learning method, widespread comparisons are carried out utilizing extraordinary Arabic corpora and bio-inspired optimization algorithms. Moreover, two publicly accessible non-Arabic corpora are utilized to compare the performance of the proposed method with those of different languages. To validate the performance of the proposed G-WOA algorithm in mastering ontology from Arabic text, they compared the solution returned to those returned through the normal GA and WOA. The G-WOA starts off evolved to search for the fine answer via a set of iterations, which include embedding the genetic operators into the WOA architecture. Eventually, the algorithm returns the answer which recommends the great set of concepts/relations that can contribute to the ontology. The outputs revealed that the proposed genetic-whale optimization algorithm outperforms contrasting algorithms throughout all the Arabic corpora in precision, recall, and F-score measures. Moreover, the proposed approach outperforms the latest strategies of ontology learning from Arabic and non-Arabic texts in terms of these three measures.

This research contributes to todays Arabic ontology learning by the following: text mining algorithm is proposed specifically for extracting the ideas and their semantic relations from the Arabic documents. The extracted set of principles with the semantic relations constitutes the shape of the ontology. In this regard, the algorithm operated on the Arabic documents with the aid of calculating the concept frequency weights depending on the term frequency weights. Thereafter, they calculated the weights of concept similarity, using the information-driven from the ontology structure involving the concept's path distance, the concept's distribution layer, and the mutual parent concept's distribution layer. Eventually, it performs the mapping of aspects by means of assigning the notion similarity to the thinking features.
This study benefits from a prior knowledge (initial concept set obtained from the text mining algorithm) to create progressive solutions for the fine concept/relation set that can constitute the ontology. Proposed ontology learning strategy is applicable on different languages; it can be utilized to extract the most appropriate ontology structure from the non-Arabic texts. The proposed algorithm extracts standards and their semantic members of the family that constitute the ontology from every record of Arabic text, in three steps: term weighting, notion similarity weights, and feature mapping. Genetic algorithms (GA) were

embedded into the WOA algorithm in order to improve a wide variety of whales (search agents) in the form of chromosomes.

The evaluation to the model was composed of three experiments: (i) comparisons with different bio-inspired optimization algorithms existing in the literature involving Arabic ontology learning, (ii) comparisons with previous published approaches on Arabic ontology gaining knowledge from text, and (ii) comparisons with modern day on gaining knowledge of ontology from non-Arabic settings. Eventually, the proposed ontology getting to know strategy is relevant to the non-Arabic texts too. It achieved higher performance that outperformed the contemporary processes on gaining knowledge of ontology from Arabic and non-Arabic text [36].

6 DISCUSSION

Whale Optimization Algorithm (WOA) is a recent swarm intelligence based meta - heuristic optimization algorithm, which simulates the natural behavior of bubble - net hunting technique of humpback whales and has been correctly applied to clear up complicated optimization troubles in a huge variety of disciplines. Therefore, when applied to large size issues, its efficiency and performance degrades due to the huge computational work load required. Distributed computing is one of the effective ways to improve the scalability of WOA for resolving large - scale issues. Whale swarm algorithm can be applied to solve nearly any optimization issues. Whale Optimization Algorithm (WOA) is one of the newly proposed algorithms belonging to the class of swarm intelligence. The humpback whale is simulated so as to find optimum solutions to various optimization issues. There are nonetheless some shortcomings that are handy to fall into nearby top of the line or slow convergence, which want to be always improved and innovated. There are many complex optimization problems such as clustering, Hadoop MapReduce, Dynamic software rejuvenation in web…etc. So, in the future, we will improve the whale swarm algorithm to solve the previous issues.

7 CONCLUSIONS

New swarm intelligence based meta-heuristic called Whale Swarm Algorithm, inspired using the whales' behavior of communicating with every other through ultrasound for hunting, is explained in this paper. We showed the methodology and strategy of whale swarm algorithm in detail. Finally, we explained several uses of WSA in many problems such as Clustering, Prediction Diseases, Mechanical, Text Mining and Production Engineering. Semantic understanding of textual knowledge to find concepts of thought could be detected by the use of genetics and WOA. This will provide a new layer of knowledge understanding through an optimization mechanism for big data analytics. It is still required to find means of advanced processing of such optimization technique to enhance the performance of WOA and text mining of huge amount of documents.

REFERENCES

[1] S.M. Khalessizadeh, R.Zaefarian, World Academy of Science, Engineering and Technology, 2006, "Genetic Mining: Using Genetic Algorithm for Topic based on Concept Distribution"

[2] Bing Zeng. Liang Gao. Xinyu Li. 2017. Whale swarm algorithm for function optimization. LNCS, 10361:624-639

[3] M. Mahi, Ö.K. Baykan, H. Kodaz, A new hybrid method based on Particle Swarm Optimization, Ant Colony Optimization and 3-Opt algorithms for Traveling Salesman Problem, Applied Soft Computing, 30 (2015) 484-490.

[4] K.C. Tan, Y. Chew, L.H. Lee, A hybrid multi-objective evolutionary algorithm for solving truck and trailer vehicle routing problems, European Journal of Operational Research, 172 (2006) 855-885.

[5] S.N. Qasem, S.M. Shamsuddin, S.Z.M. Hashim, M. Darus, E. AlShammari, Memetic multi objective particle swarm optimization-based radial basis function network for classification problems, Information Sciences, 239 (2013) 165-190.

[6] B. Zeng, Y. Dong, An improved harmony search based energy efficient routing algorithm for wireless sensor networks, Applied Soft Computing, 41 (2016) 135-147.

[7] E.S. Hou, N. Ansari, H. Ren, A genetic algorithm for multiprocessor scheduling, Parallel and Distributed Systems, IEEE Transactions on, 5 (1994) 113-120.

[8] B.-Y. Qu, P. Suganthan, S. Das, A distance-based locally informed particle swarm model for multimodal optimization, Evolutionary Computation, IEEE Transactions on, 17 (2013) 387-402.
[9] Nasiri, J., & Khiyabani, F. M. (2018). A whale optimization algorithm (WOA) approach for clustering. Cogent Mathematics & Statistics, 5(1), 1483565.
[10] Mohit Verma, Amit Kumar," A Brief Review of Applications of Whale Optimization Algorithm to Mechanical and Production Engineering", in: International Journal of Pure and Applied Mathematics, 119(18), pp. 1953-1960, 2018.
[11] Zhang, X., Liu, Z., Miao, Q. and Wang, L., Bearing fault diagnosis using a whale optimization algorithm-optimized orthogonal matching pursuit with a combined time– frequency atom dictionary, Mech. Syst. Signal Process., vol. 107, pp. 29–42.
[12] Mirjalili, S. and Lewis, A., 2016. The Whale Optimization Algorithm, Adv. Eng. Softw., vol. 95, pp. 51–67.
[13] Rohani and Mohammad. 2016. The Workflow Planning of Construction Sites using whale optimization algorithm (WOA), Turkish Online Journal of Design Art and Communication 6: 2938-2950.
[14] Prakash, D.B. and Lakshminarayana, C., 2016. Optimal siting of capacitors in radial distribution network using Whale Optimization Algorithm. Alexandria Engineering Journal.
[15] Bentouati, B., Chaib, L. and Chettih, L., 2016. A hybrid whale algorithm and pattern search technique for optimal power flow problem, Modelling, Identification and Control ICMIC 8th International Conference on. IEEE 2016.
[16] Buch, and Hitartch., 2016. Combined Emission Constrained Economic Dispatch with Valve Point Effect Loading Problem Solution using Whale Optimization Algorithm.
[17] Trivedi, Indrajit N., 2016. An emission constraint environment dispatch problem solution with microgrid using Whale Optimization Algorithm, Power Systems Conference (NPSC), 2016 National IEEE, 2016.
[18] Jangir, P. and Jangir, N., 2017. Non-Dominated Sorting Whale Optimization Algorithm (NSWOA): A Multi-Objective Optimization Algorithm for Solving Engineering Design Problems, Global Journal of Researches in Engineering: F Electrical and Electronics Engineering Volume 17 Issue 4 Version 1.0.
[19] Faseela, C.K. and Vennila, H., 2018. Economic and Emission Dispatch using Whale Optimization Algorithm (WOA), 2018 in IJECE (Vol 8, No 3). 2018.
[20] Horng. andMong, Fong., 2016. A Multi-Objective Optimal Vehicle Fuel Consumption Based on Whale Optimization Algorithm, Advances in Intelligent Information Hiding and Multimedia Signal Processing: Proceeding of the Twelfth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2016, Kaohsiung, Taiwan, (Volume 2. Springer International Publishing). 2017.
[21] Dao, Thi-Kien, Tien-Szu Pan, and Jeng-Shyang Pan., 2016 A multi-objective optimal mobile robot path planning based on whale optimization algorithm. Signal Processing (ICSP), 2016 IEEE 13th International Conference on. IEEE, 2016.
[22] Yan, Z., Sha, J., Liu, B., Tian, W. and Lu, J., 2018. An Ameliorative Whale Optimization Algorithm for Multi-Objective Optimal Allocation of Water Resources in Handan, China, water, 10,87.
[23] Marler, R. T. and Arora, J.S., 2004. Survey of multi-objective optimization methods for engineering, Structural and Multidisciplinary Optimization, vol. 26, no. 6. pp. 369–395.
[24] Kalaipriyan, T., Amudhavel, J. and Sujatha, P., 2017. Whale Optimization Algorithm for combined heat and power economic dispatch, Advances and Applications in Mathematical Sciences Volume 17, Issue.
[25] Jie Cai. Jiawei Luo. Shulin Wang. Sheng Yang. 2018. Feature selection in machine learning: A new perspective. Neurocomputing, 300:70–79.
[26] Anuradha, P. and Dr. Vasantha Kalyani David," International Journal of Research and Analytical Reviews (IJRAR), 6(2), 2018.
[27] Elhag, A., & Ozcan, E. (2018). Data clustering using grouping hyper-heuristics. Evolutionary Computation in Combinatorial Optimization, LNCS, 10782, 101–115.
[28] Zhang, C., Ouyang, D., & Ning, J. (2010). An artificial bee colony approach for clustering. Experiments System Applications, 37, 4761–4767. doi: 10.1016/j. eswa.2009.11.003
[29] Mirjalili, S., & Lewi, A. (2016). The whale optimization algorithm. Advancement Engineering Softwares, 95, 51–67. doi:10.1016/j.advengsoft.2016.01.008
[30] Kenedy, J., & Eberhart, R. (1995). Particle swarm optimization. Proceeding of the 1995 IEEE international conference on neural network, 194–208.

[31] Storn, R., & Price, K. (1997). Differential evolution- a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization,11, 341–359. doi:10.1023/A:1008202821328
[32] Rashedi, E., Nezamabadi- Pour, H., & Saryazdi, S. (2009). GSA: A gravitational search algoritm. Information Sciences, 179, 2232–2248. doi:10.1016/j.ins.2009.03.004
[33] Yao, X., & Liu, Y. (1999). Evolutionary programing made faster. IEEE Transactions Evolution Computer, 3, 82– 102. doi:10.1109/4235.771163
[34] Jhila Nasiri and Farzin Modarres Khiyabani," A whale optimization algorithm (WOA) approach for clustering", in: APPLIED & INTERDISCIPLINARY MATHEMATICS, 2018.
[35] Mezghanni, I.B.; Gargouri, F. CrimAr: A Criminal Arabic Ontology for a Benchmark Based Evaluation. Procedia Comput. Sci. 2017, 112, 653–662. [CrossRef]
[36] Rania M. Ghoniem , Nawal Alhelwa and Khaled Shaalan," A Novel Hybrid Genetic-Whale Optimization Model for Ontology Learning from Arabic Text", In: the journal of algorithms,12(9),2019.

Creating and Implementing ArSL Corpus for Deaf Drivers

Samah A. Abbas[*1], Hassanin M. Al-Barhamtoshy [**2], Fahad M. Al-Otaibi [***3]

[*]*Management Information Systems Department, Faculty of Economics and Administration, King Abdulaziz University*
Jeddah, Saudi Arabia
[1]sabbas@kau.edu.sa

[**]*Information Technology Department, Faculty of Computing & Information Technology, King Abdulaziz University*
Jeddah, Saudi Arabia
[2]hassanin@kau.edu.sa

[***]*Information System Department, Faculty of Computing & Information Technology, King Abdulaziz University*
Jeddah, Saudi Arabia
[3]fmmalotaibi@kau.edu.sa

Abstract— **Sign languages are a general language that deaf people around the world use it for communication with others. However, normal people usually do not know the sign language and they do not have to learn their language for communicating with them in daily life. For supporting deaf people and facilitating their work, many technologies open up possibilities for overcoming such barriers, particularly through natural language processing (NLP), text understanding, machine translation, and sign language simulation. In this paper, we focus on the problem that faced the deaf community in Saudi Arabia as an important member of society who needs support in communicating with others, particularly in the field of work as a driver. Where they need a system that facilitates the process of communicating with passengers using NLP that helps translate Arabic sign language (ArSL) into voice and vice versa. In this paper, we discuss the linguistic background and automated ArSL. In addition, we provide a brief look at the studies that have used various techniques to identify the sign language in order to translate it into sound and vice versa. Moreover, we illustrate our corpus, data determination (deaf driver terminologies) and dataset creation and processing in order to implement our future proposed system. Therefore, the dataset evaluation will be presented and simulated.**

Keywords: **Arabic Sign Language, Speech Recognition, Sign Language Recognition, Natural Language Processing, Deaf Driver in Saudi Arabia.**

1 INTRODUCTION

Deaf people use sign language to communicate with their peers and normal people who know their sign language. Thus, sign language is the only way for communication with deaf although it is still getting less attention from normal people. Moreover, sign language is not a unified language between deaf around the world, where each country has its own sign language. For instance, American Sign Language (ASL), Indian Sign Language (ISL), Australian Sign language (ASL), British Sign Language (BSL) and Arabic Sign Language (ArSL). In particular, Arab countries who use ArSL like Gulf, Al-Sham, and some Arab countries in North Africa have some similarities and differences in sign language, the reason behind that is the differences in dialects. Therefore, deaf people face some communication problems with the community in many aspects while they are working or practicing their daily lives, such as health, education, and transportation. One of the solutions for these problems is using a sign language interpreter, which is a person that knows the sign language and can interpret it to normal people. However, this way is not an optimal way because of loss of privacy and working independence [1], [2]. So, researchers developed some technologies for providing a computer interpreter instead of using a human. For example, sign language recognition technique, machine translation (MT). In addition, there are some researchers and Arabic organizations like

The Arab League Educational, Cultural and Scientific Organization (ALECSO) provided their effort for unifying the ArSL by introducing the first dictionary in 1999 [3].

In Saudi Arabia, deaf people have some difficulties in communicating with others while they are driving a vehicle like a car and the deaf or normal person sitting as a passenger. They cannot use Arabic Sign Language (ArSL) at that time, and even if they can use ArSL, the normal passenger may not understand it. Also, a normal passenger cannot describe the needed location for the deaf driver. However, there are many mobile applications that can be used to facilitate communication such as, Tawasol, and Turjuman but still not translating in real-time [4], [5]. Moreover, normal passengers or drivers do not necessarily download the application of sign language translation just for using one time. In the deaf driver domain, Saudi Arabia faced a lack of technologies that can improve the communication between deaf drivers and passengers. Based on our knowledge, there is no research done yet that can introduce a solution in the deaf drivers' domain in Saudi Arabia.

This paper aims to overview the ArSL recognition and MT. In addition, we present our corpus (texts and videos) in the deaf driver domain. This paper is organized as follows: The second section illustrates the work related to ArSL and Speech Recognition Systems and the Machine Translation (MT) systems. The third section illustrates the ArSL linguistic background where the fourth is about ArSL design architecture. In the fifth section, we discuss our data collection processing and modules along videos of deaf driver corpus evaluation that required implementing the ArSL system in the context of deaf driving. Finally, we illustrate the future research directions.

2 LITERATURE REVIEW OF ArSL RECOGNITION AND MT SYSTEM

This literature review investigates the different techniques for ArSL recognition and speech recognition. Also, it focuses on the researches that illustrate some different systems by using a different method for translation from text or speech to ArSL and from ArSL to text or voice in Saudi Arabia.

In terms of ArSL recognition researches, many researchers conducted several experiments using different methods for translating ArSL to text or voice. Some of them focused on implementing systems for real-time communication [6]–[18]. In contrast, others dedicated their efforts to develop the best recognition system but not in real-time. They used several techniques of the ArSL recognition system to achieve a high percentage of accuracy [19]–[26]. For instance, Support Vector Machine (SVM) with a camera that gained 98.8 % [19] Principal Component Analysis (PCA), and Hidden Markov Models that achieved 99.9% accuracy [24]. The techniques used for ArSL recognition by some researchers divided into two methods. The first method is a wearable-based, which builds based on special equipment, such as smart gloves or glasses. This method supports real-time communication [7], [10], [17], [18], [27]. The second method is the vision-based (sometimes called a camera, image or video-based). This method is implemented based on recognition of hand gestures by a camera that captures the images and then segments to analyze (preprocessing). After that, some feature vectors are extracted for classifying in order to build the final model and measure the accuracy. In the classification stage, different classifications are used by researches for building their experiments. Some researchers focused their experiments on one classifier, for instance, Hidden Markov Models (HMM), Microsoft Kinect Device (sensor), a different method of the neural network, or a colored image [6], [12]–[14], [22], [24], [25]. While the researchers of [8], [9], [19]-[21], [23], [26] used two or more classifiers for ArSL recognition. Some of these researchers, which are [11], [15] implemented their system on mobile devices. All of these studies recorded different degrees of accuracy, where the goal of each study was to use the most method that achieves higher accuracy as a key performance indicator (KPI).

In the translation system from text or speech to ArSL, researchers conducted their experiments in building the translation system based on MT. Where the output of this system can be either video or virtual animation called the avatar [28]. One of the researchers used both video and avatar as ArSL output in order to compare them based on WER [29].

During translation, the translation system from the Arabic text is very complex due to the writing rules. The natural Arabic text should consider the grammatical rule, syntax, and semantic. Also, these considerations can support converting the text into the right meaning. Researches of [29]–[33] used various methods for analyzing the text. For example, they used semantic, pragmatic, syntactic, and morphological analysis.

To sum up the previous related work, we can say that the Arab deaf community, who use ArSL, has been the center of interest for researchers. The researchers dedicated their effort in improving the performance of recognition and building a correct system. The reason behind that is to facilitate communication with the deaf in the working environment and also in their daily life.

As we saw previously, there is a lack of research that supports and enhances the method of communication between the deaf driver or passenger. Whereas, a lot of research done focused on proposing the best techniques or methods in many sectors, such as education and health. In our proposed system, we intend to choose the best method or technique for recognizing ArSL to speech by using a video camera. Also, we intend to choose the best method or technique for recognition of speech and translation of the text into ArSL.

3 ArSL and Linguistic Background

ArSL has massive complexities in phonology, morphology, and structure, which is not like other sign languages. These complicities are explained below.

G. Phonology of ArSL

The "‘phonemes" are mental representations, which is just a way to empty what's inside the brain. The phonemes consist of four elements: 1) shape of hand. 2) Orientation of hand. 3) Position of hand from the body. 4) Direction of hand while moving [34]. In phonology, these four elements known as Manual Features (MFs). Particularly in Sign Language there are MFs and also Non-Manual Features (NMFs) are involved. The (NMFs) refers to the emotional parts of the body, for instance, lip motion, facial expression, shoulder, head, eyelids and eyebrows movement. Usually in ArSL , we use both MFs and NMFs for giving the correct meaning, called essential NMF. Where if the signer uses just MFs, the meaning will change to another meaning that the signer did not mean [35].

H. Morphology and Structure of ArSL

The grammar rules of ArSL are not the same as grammar rules of the Arabic language. The differences are in the following points: verb tenses, singular and plural differences, prepositional and adverb rules and gender signs. In terms of the sentence structure, ArSL just uses Subject-Verb-Object (SVO) instead of structure SVO, OVS and VOS [30].

4 ArSL and Design Architecture

New technologies that support communication have a significant impact on human life. For deaf people, the developers and researchers tried to use some new technologies to facilitate deaf life by developing some automated systems that can support them in communication in different aspects of their life with others. In this section, we will illustrate the brief explanation of some techniques used in order to implement the automated system for better communication between the community and amongst themselves.

A. Machine Translation

Machine Translation (MT) is a standardized name for the system that builds based on the computer analysis in order to translate between two natural languages. It is used for both text and speech using Natural Language Processing (NLP) and Artificial Intelligence (AI). For example, translating from the source "English" into the target "Arabic" [36], [37]. Also, MT is used for translating text or speech to avatar or video sign language. There are several approaches in MT that can be used based on what we need to translate, and what is a better translation. One of the approaches is Direct-Based translation without considering any grammar rules. For improving the quality of MT, the Rule-Based approach introduced, which built to analyze the syntactic parsing for both source and target language. Another approach is Corpus-Based, which deals with massive data that contain sentences. Also, there exists Knowledge-Based approaches, which consider understanding both target and source text within linguistic and semantic knowledge. Lastly is the google translation which is developed by Google [37].

B. Arabic Speech Recognition

Speech Recognition (SR) is a computerized system that converts the speech into text or sign. This system is used to communicate between humans and machines. It is also known as automatic speech recognition (ASR). SR by machine is a complicated task because of the differences in dialects, contexts and speech styles. For reducing this complexity in SR, the system can exploit the repetition and speech signal structure of the token as multiple sources of knowledge. The sources of SR are built based on knowledge of phonology, syntax, phonemes, grammar and phrases [38], [39]. In addition, the SR system has multiple classifiers. 1) The utterance of speech, which is composed of separate, connected, continued and spontaneous speech. 2) The speaker model, which contains one of both dependently that designed for a specific speaker or independently for different speakers. 3) The size of vocabulary [40].

In terms of the SR process, there are four stages that SR can be implemented in. Analyze the signal speech and then extract the feature by using different techniques for identifying the vector, like MFCC (Mel-frequency cepstral coefficient). After that, we build a model using different techniques like HMMs with the training dataset. In the last stage, we test the model within matching, taking the dissection and measuring the performance based on the error rate [41], [42] as shown in Fig. 1.

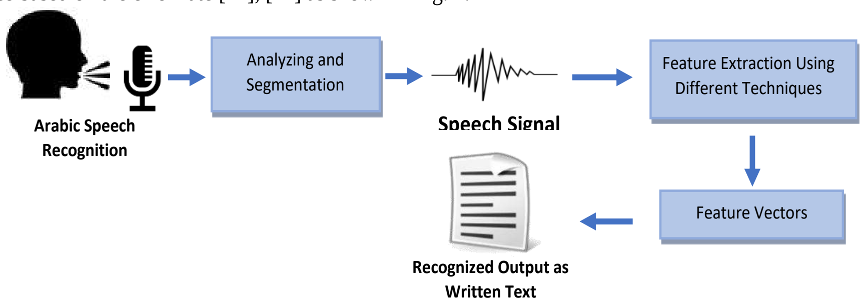


Figure 1: Block diagram to recognize spoken to Arabic written text

C. ArSL Gestures Recognitions

Gesture recognition is defined as the ability of the computer for understanding the gestures and executing the commands based on the performed gestures. The first gesture recognition system was developed in 1993 as a kind of user interface for perceptual computing that helps for capturing human gestures and then transfer it into commands by using a computerized system. It is used with many technologies, particularly in the games field, such as X-box, PlayStation, and Wii Fit. These games use Just Dance and Kinect Sports, which recognizes the hand and some parts of the body [43], [44].

In the sign language field, the gesture recognition system uses the following processes: 1) Recognize the deaf signs. 2) Analyzes the sign. 3) Converts this sign into the meaningful text (word or sentences), voice or expressions that non-deaf can understand. Moreover, there are two main methods for ArSL gesture recognition, which are wearable-based devices and vision-based devices (video or image based). Each of them have their advantages and disadvantages. One of the advantages of wearable-based does not need to look for changes in the background and lighting. The disadvantage is impeding movement. In contrast, the vision-based advantage is easy to move whereas the disadvantage is the effect of changing the background and lighting [45] [46].

Also, each of them has different processes and techniques. In the Vision-based process, one or more than one camera is the main tool that should be available for using this method. On the other hand, wearable-based method is dependent on some types of equipment and computers mainly. In terms of their processes, wearable-based method spells the alphabet by reading the particular information in each finger joint sensor or glove sensor. However, the vision-based has some stages, which are the following:

- The image capturing, using a camera for collecting data (building the corpus) and analyzing the collected images.
- The preprocessing, which starts to prepare the images and identify the information based on the color (segmentation).

- The feature extraction using some techniques like Root Mean Square (RMS) in order to identify the feature vector.
- The classification which classifies based on the feature vector in order to build the model [47] [48]. The process of vision-based gesture recognition is shown in the Fig.2.

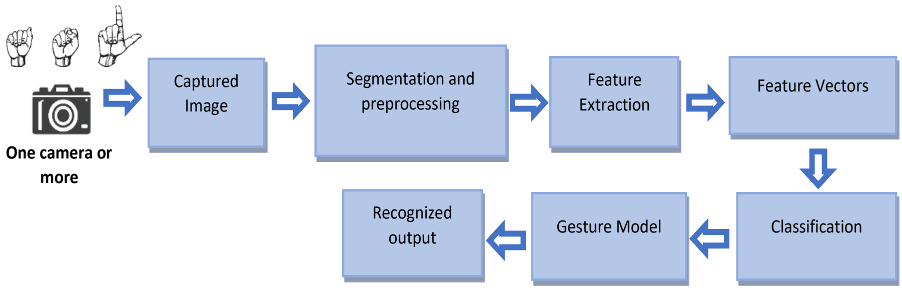


Figure2: Block diagram of the ArSL gesture recognition processing Adapted from [48]

5 Deaf Driver Corpus

In order to create deaf driver corpus, we divided the processes that we will use for this creation into 4 modules. Which are preprocessing, recording, assessment and validation module. These high-level and low-level approaches are represented in Fig.3 and Fig.4.

Figure3: High-level approach of our dataset (corpus) collection

A. Data Collection and Creation

In this section, we explain our data collection, data determination (deaf driver terminologies) and dataset creation and processing in order to create our video corpus through the preprocessing and recording module.

In the preprocessing module, first, we gathered the data based on two level: a word or phrase-level dataset and a sentence-level dataset in order to create a small Arabic dictionary. This dictionary is divided into eight sections (categories). 1) Welcoming "Salam Alaikum [السلام عليكم] and How are you? [كيف حالك؟]". 2) Directions "Left [يسار] and right [يمين]". 3) Place "School [مدرسة] and Deaf Association [جمعية الصم]". 4) Traffic and Transportation "Driving License [رخصة سير] and Traffic light [إشارة ضوئية]". 5) Sentences that are used by deaf drivers when they need to talk with their passengers, for example, "We have arrived [لقد وصلنا]" 6) Sentences that are used by passengers when they need to talk with their deaf drivers, for example, "I do not have cash to pay the amount [لا أملك كاش لدفع المبلغ]". 7) General Words " No and Yes [لا و نعم] and In and On [في و على]". 8) Amount "Dollar [دولار], Riyal [ريال] and 2 Riyals until 100 Riyals [2 ريال الى 100 ريال]".

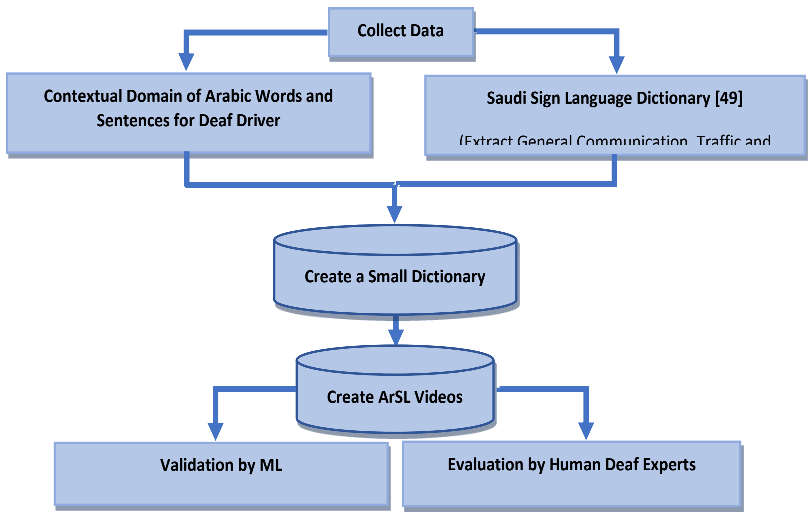


Figure 4: Low-level approach of our dataset (corpus) collection

The total words and sentences are 215. Some of these words and sentences are of the general communication, collected from 2018 edition of the Saudi Sign Language Dictionary [49] and some of them are collected from the contextual domain of the normal conversation between the taxi driver and their passengers. The flowing Table 1 shows part of our created Arabic dictionary.

TABLE IX
AN EXAMPLE OF OUR CREATED DICTIONARY

Section 1: Welcome	**القسم 1: الترحيب**
Word / Phrase / Sentences	**الجملة \ الكلمة \ المصطلح**
1. Salam Aleikum (peace be upon you)	1. السلام عليكم
2. How are you?	2. كيف حالك؟
Section 2: Directions	**القسم 2: الاتجاهات**
1. Left	1. يسار
2. Right	2. يمين
Section 3: Place	**القسم 3: الأماكن**
1. School	1. مدرسة
2. Deaf Association	2. جمعية الصم
Section 4: Traffic and Transport	**القسم 4: حركة المرور والنقل**
1. Driving License	1. رخصة سير
Section 5: Driver	**القسم 5: السائق**
1. We have arrived	1. لقد وصلنا
Section 6: Passenger	**القسم 6: الراكب**
1. I do not have cash to pay the amount	1. لا أملك كاش لدفع المبلغ

In the recording module, we did our videos at a rate of 29.97 FPS using one camera for our ArSL corpus (video capturing) with one of the expert's signer in ArSL. For recording this corpus, we took approximately 20 minutes continuously, where the total of our corpus is 215 words including sentences and signs. The expert signer tried to use one hand to be suitable with the deaf driver context unless if the sign required it to be necessary to use both hands. Then, we did video segmentation by video editing VEGAS (segment every single sign for word or sentences of our dictionary to one video where the total videos are 215). For supporting our future work, we added an Arabic audio and labeled each video with Arabic text that refers to the same ArSL that we recorded. Fig.5 represents one of our captured corpus.

Figure 5: One Captured Video from our ArSL Video Corpus

B. Dataset Evaluation

These collected and created data will be used for evaluation and validation of our corpus. In this section, we explained the evaluation module and the validation will be made in the future work.

In the assessment module, we evaluated each generated corpus' video based on our created dictionary using the human expert evaluation technique. The number of expert participants in ArSL was four and two of them were deaf. We used the quantitative approach (questionnaire) and we divided it into two sections. The first section was a demographic questionnaire (gender - age- education level - if he/she is deaf or not). The second section was a video evaluation based on the related word (phrase) or sentence. Each video attached to each related word or sentence. We asked the participants to evaluate the 215 sign videos if the video for each word or sentence is correct or not. If it is not correct, which means that the video is not related to that particular word or sentence. We asked the four participants to choose one of the correction types that they were supposed to do for each video (adding - replacing - deleting). The way of evaluation is explained in Table 2 using some videos evaluated by one of the experts.

TABLE X
AN EXAMPLE OF THE WAY OF EVALUATION WITH ONE EXPERT

Section One: Welcom					**القسم الاول: الترحيب**
Correction if the translation is wrong: The video of sign needs.			**Video evaluation based on the related word (phrase) or sentences.** √ أو ×	**Word \ phrase \ sentences**	**الكلمة \ المصطلح \ الجملة**
Deleting	**Replacing**	**Adding**			
			√	1- Salam aleikum	1-السلام عليكم
			√	2- Waleikum salam	2-عليكم السلام
	√		×	3- Good morning	3-صباح الخير

The deaf experts' evaluations results based on Word Error Rate (WER) for each category (section) like welcome, directions and place is shown in Fig.6. It means that for each category the result of the evaluation of these videos was wrong. Also, they must be replaced with the correct videos.

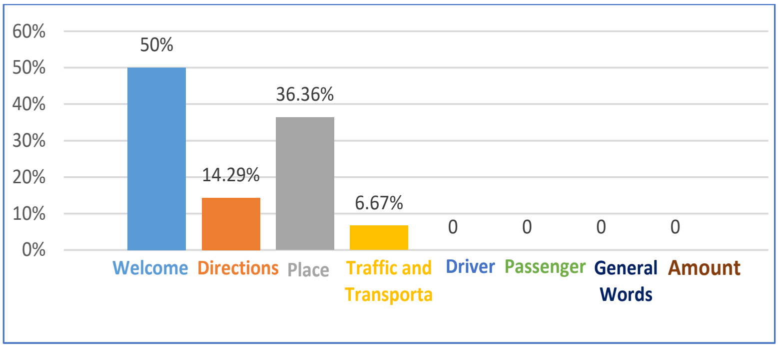


Figure 6: The Deaf Experts' Evaluations Results Based on (WER) for Each Category

As we can see in graph 6, first, the welcoming category has a high percentage of WER which is 50%. After that, the place category has approximately 36%. Thus, we need to reduce the WER in our video corpus that is related to these two categories (sections) for enhancing the communication between deaf drivers and passengers and also in order to describe the correct place for passenger's destination.

The total WER of our video corpus was 10.23% as shown in Table 3. For solving the wrong ArSL videos corpus, we re-captured these videos again based on the corrected signs that the evaluators explained for us.

TABLE III
TOTAL WORD CORRECT AND ERROR RATE FOR OUR CREATED VIDEO CORPUS

	Correct	Wrong
Signs' Videos Corpus	89.8%	10.2%

In the validation module, in the future work, we are going to implement ML technique using python as a programming language. For doing that, we recorded other videos with different signers to divide our data into training and testing datasets in order to measure the accuracy and error rate.

6 CONCLUSIONS AND FEATURE WORK

We have reviewed, through this research, previous studies that have been completed in the field of ArSL and speech recognition and also the translation system that converts ArSL into text and speech or vice versa. We also reviewed used methods that seek to achieve better performance while reducing the error rate in translation. We clarified the difficulties faced by translating ArSL from grammatical, semantic, and syntax. How they affect the accuracy of the translation and recognition. Finally, we described the ArSL dictionary for deaf drivers and we explained the data collection processes to construct our corpus videos. It was recorded by using one camera and then verified with 4 participants, experts in ArSL, of whom two were deaf.

The created ArSL corpus offers possibilities for testing various feature extraction methods and recognition techniques. The dataset extension and validation using Machine Learning (ML) will be implemented in future work. In addition to that, this corpus is going to be used to design our proposed system that facilitates communication with the deaf driver.

ACKNOWLEDGMENT

This work was supported by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah. The authors, therefore, gratefully acknowledge the DSR technical and financial support.

REFERENCES

[1] L. H. Forestal, "A study of deaf leaders' attitudes towards sign language interpreters and interpreting," New York University, 2001.

[2] B. Broecker, "Speaking the Language of Sign: The Art and Science of Signing," *American Annals of the Deaf*, vol. 131, no. 3, pp. 199-200, 1986.

[3] M. Al-Binali and S. Samareen, "Grammar of the Unified Qatari Arabic Sign Language," *Dar Al-Sharq, Doha, Qatar*, In Arabic, 2009.

[4] A. Al-Nafjan, B. Al-Arifi, and A. Al-Wabil, “Design and Development of an Educational Arabic Sign Language Mobile Application: Collective Impact with Tawasol,” in *Universal Access in Human-Computer Interaction. Access to Interaction*, 2015, pp. 319–326, doi: 10.1007/978-3-319-20681-3_30.

[5] Team Mind Rockets, "Mind Rockets Inc, Assistive Technologies for the deaf", Mindrocketsinc.com, 2017. [Online]. Available: http://mindrocketsinc.com. [Accessed: 08-Nov-2018].

[6] T. Aujeszky and M. Eid, "A gesture recogintion architecture for Arabic sign language communication system," *Multimedia Tools and Applications,* vol. 75, no. 14, pp. 8493-8511, 2016.

[7] M. A. Mohandes, "Recognition of two-handed Arabic signs using the CyberGlove," *Arabian Journal for Science and Engineering,* vol. 38, no. 3, pp. 669-677, 2013.

[8] M. F. Tolba, A. Samir, and M. Aboul-Ela, "Arabic sign language continuous sentences recognition using PCNN and graph matching," *Neural Computing and Applications,* vol. 23, no. 3-4, pp. 999-1010, 2013.

[9] N. B. Ibrahim, M. M. Selim, and H. H. Zayed, "An automatic arabic sign language recognition system (ArSLRS)," *Journal of King Saud University-Computer and Information Sciences,* vol. 30, no. 4, pp. 470-477, 2018.

[10] N. Tubaiz, T. Shanableh, and K. Assaleh, "Glove-based continuous Arabic sign language recognition in user-dependent mode," *IEEE Transactions on Human-Machine Systems,* vol. 45, no. 4, pp. 526-533, 2015.

[11] A. Eqab and T. Shanableh, "Android mobile app for real-time bilateral Arabic sign language translation using leap motion controller," in *2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, 2017, pp. 1-5: IEEE.

[12] E. E. Hemayed and A. S. Hassanien, "Edge-based recognizer for Arabic sign language alphabet (ArS2V-Arabic sign to voice)," in *2010 International Computer Engineering Conference (ICENCO)*, 2010, pp. 121-127: IEEE.

[13] M. Al-Rousan, K. Assaleh, and A. Tala’a, "Video-based signer-independent Arabic sign language recognition using hidden Markov models," *Applied Soft Computing,* vol. 9, no. 3, pp. 990-999, 2009.

[14] F. Guesmi, T. Bouchrika, O. Jemai, M. Zaied, and C. B. Amar, "Arabic sign language recognition system based on wavelet networks," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016, pp. 003561-003566: IEEE.

[15] F. Al Ameiri, M. J. Zemerly, and M. Al Marzouqi, "Mobile Arabic sign language," in *2011 International Conference for Internet Technology and Secured Transactions*, 2011, pp. 363-367: IEEE.

[16] D. Dahmani and S. Larabi, "User-independent system for sign language finger spelling recognition," *Journal of Visual Communication and Image Representation,* vol. 25, no. 5, pp. 1240-1250, 2014.

[17] M. Mohandes and M. Deriche, "Arabic sign language recognition by decisions fusion using Dempster-Shafer theory of evidence," in *2013 Computing, Communications and IT Applications Conference (ComComAp)*, 2013, pp. 90-94: IEEE

[18] F. N. H. Al-Nuaimy, "Design and implementation of deaf and mute people interaction system," in *2017 International Conference on Engineering and Technology (ICET)*, 2017, pp. 1-6: IEEE.

[19] H. Luqman and S. A. Mahmoud, "Transform-based Arabic sign language recognition," *Procedia Computer Science,* vol. 117, pp. 2-9, 2017.

[20] A. Tharwat, T. Gaber, A. E. Hassanien, M. K. Shahin, and B. Refaat, "Sift-based arabic sign language recognition system," in *Afro-european conference for industrial advancement,* 2015, pp. 359-370: Springer.

[21] N. A. Sarhan, Y. El-Sonbaty, and S. M. Youssef, "HMM-based arabic sign language recognition using kinect," in *2015 Tenth International Conference on Digital Information Management (ICDIM)*, 2015, pp. 169-174: IEEE.

[22] A. SamirElons, M. Abull-ela, and M. F. Tolba, "Pulse-coupled neural network feature generation model for Arabic sign language recognition," *IET Image Processing,* vol. 7, no. 9, pp. 829-836, 2013.

[23] M. Elpeltagy, M. Abdelwahab, M. E. Hussein, A. Shoukry, A. Shoala, and M. Galal, "Multi-modality-based Arabic sign language recognition," *IET Computer Vision,* vol. 12, no. 7, pp. 1031-1039, 2018.

[24] A. A. Ahmed and S. Aly, "Appearance-based arabic sign language recognition using hidden markov models," in *2014 International Conference on Engineering and Technology (ICET)*, 2014, pp. 1-6: IEEE.

[25] M. ElBadawy, A. Elons, H. A. Shedeed, and M. Tolba, "Arabic sign language recognition with 3d convolutional neural networks," in *2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*, 2017, pp. 66-71: IEEE.

[26] M. Mohandes, S. Aliyu, and M. Deriche, "Arabic sign language recognition using the leap motion controller," in *2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE)*, 2014, pp. 960-965: IEEE.

[27] D. Dahmani and S. Larabi, "User-independent system for sign language finger spelling recognition," *Journal of Visual Communication and Image Representation,* vol. 25, no. 5, pp. 1240-1250, 2014.

[28] N. Aouiti, M. Jemni, and S. Semreen, "Arab gloss and implementation for Arabic Sign Language," in *2017 6th International Conference on Information and Communication Technology and Accessibility (ICTA)*, 2017, pp. 1-6: IEEE.

[29] H. M. Al-Barhamtoshy, N. E. Abuzinadah, A. 3, T. F. 4, A. A. 5 and, and A. A. Allinjawi, "Development Of An Intelligent Arabic Text Translation Model For Deaf Students Using State Of The Art Information Technology," *Biosci. Biotechnol. Res. Commun.*, vol. 12, no. 2, pp. 338–345, Jun. 2019, doi: 10.21786/bbrc/12.2/17.

[30] H. Luqman and S. A. Mahmoud, "Automatic translation of Arabic text-to-Arabic sign language," *Universal Access in the Information Society,* vol. 18, no. 4, pp. 939-951, 2019.

[31] N. Aouiti, "Towards an automatic translation from Arabic text to sign language," in *Fourth International Conference on Information and Communication Technology and Accessibility (ICTA)*, 2013, pp. 1-4: IEEE.

[32] O. H. Al-Barahamtoshy and H. M. Al-Barhamtoshy, "Arabic Text-to-Sign (ArTTS) Model from Automatic SR System," *Procedia Computer Science,* vol. 117, pp. 304-311, 2017.

[33] A. M. Almasoud and H. S. Al-Khalifa, "Semsignwriting: A proposed semantic system for Arabic text-to-signwriting translation," *Journal of Software Engineering and Applications,* vol. 5, no. 08, p. 604, 2012.

[34] S. Ojala, "Towards an integrative information society: Studies on individuality in speech and sign," 2011.

[35] T. Johnston and A. Schembri, *Australian Sign Language (Auslan): An introduction to sign language linguistics*. Cambridge University Press, 2007.

[36] V. K. Verma, S. Srivastava, and N. Kumar, "A comprehensive review on automation of Indian sign language," in *2015 International Conference on Advances in Computer Engineering and Applications*, 2015, pp. 138-142: IEEE.

[37] I. Pradhan, S. P. Mishra, and A. K. Nayak, "A Collation of Machine Translation Approaches with Exemplified Comparison of Google and Bing Translators," Singapore, 2020, pp. 854-860: Springer Singapore.

[38] Y. Chow *et al.*, "BYBLOS: The BBN continuous speech recognition system," in *ICASSP'87. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1987, vol. 12, pp. 89-92: IEEE.

[39] A. Katyal, A. Kaur, and J. Gill, "Automatic Speech Recognition: A Review," *International Journal of Engineering and Advanced Technology (IJEAT) ISSN*, pp. 2249-8958, 2014.

[40] V. Radha and C. Vimala, "A review on speech recognition challenges and approaches," doaj. org, vol. 2, no. 1, pp. 1-7, 2012.

[41] S. K. Saksamudre and R. Deshmukh, "Isolated Word Recognition System for Hindi Language," *International Journal of Computer Science andEngineering,* vol. 3, no. 7, pp. 110-114, 2015.

[42] M. Yankayiş, T. Ensari, and N. Aydin, "Performance Evaluation of Feature Extraction and Modeling Methods for Speaker Recognition."

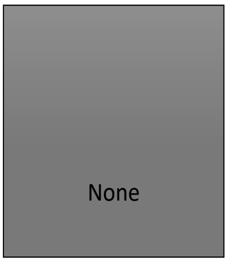
[43] S. Schechter, "What is gesture recognition? Gesture recognition defined," Marxent, 24-Mar-2014. [Online]. Available: https://www.marxentlabs.com/what-is-gesture-recognition-defined/. [Accessed: 16-Jan-2020].

[44] T. Darrell and A. Pentland, "Space-time gestures," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1993, pp. 335-340: IEEE.

[45] P. Paudyal, J. Lee, A. Banerjee, and S. K. Gupta, "Dyfav: Dynamic feature selection and voting for real-time recognition of fingerspelled alphabet using wearables," in *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, 2017, pp. 457-467: ACM.

[46] R. Ambar, C. K. Fai, M. H. A. Wahab, M. M. A. Jamil, and A. A. Ma'radzi, "Development of a Wearable Device for Sign Language Recognition," in *Journal of Physics: Conference Series*, 2018, vol. 1019, no. 1, p. 012017: IOP Publishing.
[47] M. R. Mahmood and A. M. Abdulazeez, "Different Model for Hand Gesture Recognition with a Novel Line Feature Extraction," in *2019 International Conference on Advanced Science and Engineering (ICOASE)*, 2019, pp. 52-57: IEEE.
[48] K. G. Derpanis, "A review of vision-based hand gestures," *Department of Computer Science York University*, 2004.
[49] Saudi Association For Hearing Impairment, Saudi Sign Language Dectionary, Riyadh,2018.

BIOGRAPHY



Samah A. Abbas 1, received the B.A. degree in Economics and Administration, and M.Sc. degrees in Computer Science, Information System from King Abdulaziz University, Saudi Arabia. Currently, she is a Teacher Assistant of Management Information System with the Faculty of Economic and Administration, King Abdulaziz University, Jeddah, Saudi Arabia. She is an Associate member in the international conference USAR- IC 2020 as a leader of the session in ICSM 20 (Business field) in Athena. She is interested in big data, IoT, data science, machine learning and artificial intelligence.

Hassanin M. Al-Barhamtoshy , Hassanin M. Al-Barhamtoshy received the B.S. degree in electronic and communication engineering from Cairo University, and the M.S. degree in systems and computers engineering from the Azhar university, Cairo. He received the Ph.D. degree in systems and computers engineering from Azhar University, Cairo. He was a Professor in the Department of Systems and Computer Engineering at Azhar University. He is currently Professor in the Department of Computer Science and Information Technology at Faculty of Computing and Information Technology, KAU University (2003-now). His interests: language engineering, software engineering, special needs technologies, machine translation, dictionary and lexicon building, speech processing, image processing, image similarity, Sentiment Analysis, and Arabic OCR.

Fahad M. Al-Otaibi , received the B.A. and M.Sc. degrees in computer science from King Abdulaziz University, Saudi Arabia, and the Ph.D. degree in computer science and information systems from Montfort University, Leicester, U.K., in 2014. He is currently an Associate Professor with the Faculty of Computer and Information Technology, King Abdulaziz University. He is interested in smart learning, machine learning, and artificial intelligence.

إنشاء وتطبيق مجموعة لغة الإشارة العربية للسائقين الصم

*سماح أنور عباس، **الحسنين محمد البرهمتوشى، ***فهد مزيد العتيبي

[1] قسم نظم المعلومات الإدارية، كلية الاقتصاد والإدارة، جامعة الملك عبد العزيز، جدة، المملكة العربية السعودية
[2]قسم تقنية المعلومات، كلية الحاسبات وتقنية المعلومات، جامعة الملك عبد العزيز، جدة، المملكة العربية السعودية
[3] قسم نظم المعلومات، كلية الحاسبات وتقنية المعلومات، جامعة الملك عبد العزيز، جدة، المملكة العربية السعودية

[1]sabbas@kau.edu.sa
[2]hassanin@kau.edu.sa
[3]fmmalotaibi@kau.edu.sa

الخلاصة

لغات الإشارة هي لغة عامة يستخدمها الأشخاص الصم حول العالم للتواصل مع الآخرين. ومع ذلك، لا يعرف الأشخاص العاديون عادة لغة الإشارة ولا يتعين عليهم تعلم لغتهم للتواصل معهم في الحياة اليومية. لدعم الصم وتسهيل عملهم، تفتح العديد من التقنيات إمكانياتها للتغلب على هذه العوائق، خاصة من خلال معالجة اللغة الطبيعية (NLP) وفهم النصوص والترجمة الآلية ومحاكاة لغة الإشارة. في هذه الورقة، نركز على المشكلة التي واجهت مجتمع الصم في المملكة العربية السعودية كعضو مهم في المجتمع يحتاج إلى دعم في التواصل مع الآخرين، لا سيما في مجال العمل كسائق. حيث يحتاجون إلى نظام يسهل عملية التواصل مع الركاب باستخدام البرمجة اللغوية العصبية التي تساعد على ترجمة لغة الإشارة العربية (ArSL) إلى صوت والعكس. في هذه الورقة، نناقش الخلفية اللغوية ولغة الإشارة الآلية. بالإضافة إلى ذلك، نقدم نظرة مختصرة على الدراسات التي استخدمت أساليب مختلفة لتحديد لغة الإشارة لترجمتها إلى صوت والعكس صحيح. علاوة على ذلك، فإننا نوضح مجموعتنا، وتحديد البيانات (مصطلحات السائق الأصم) وإنشاء مجموعة البيانات ومعالجتها من أجل تطبيق نظامنا المقترح في المستقبل. لذلك، سيتم تقديم تقييم مجموعة البيانات ومحاكاتها.

الكلمات الرئيسية

لغة الإشارة العربية، التعرف على الكلام، التعرف على لغة الإشارة، معالجة اللغة الطبيعية، سائقي الصم في المملكة العربية السعودية.

Corpus based NLP

A Critical Review of Language Resources and Tools for Arabic Sentiment Analysis

Miramar Etman[1], Sameh Alansary[2]

Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt

[1]miramar.etman@alexu.edu.eg

[2]sameh.alansary@bibalex.org

***Abstract*— Sentiment Analysis field is growing expeditiously, nowadays researchers from a wide range of disciplines are devoting their time and efforts to extract and detect opinions, feelings, views and attitudes towards a specific topic, products, brands and services. With the booming number of people over the Internet where they spend much of their time expressing their opinions and reactions on social platforms, as a result researchers are heading for collecting data from social media platforms. Arabic sentiment analysis has been gaining much attention recently. However, there is insufficiently in the available language resources and tools when comparing Arabic language to English language. This paper will try to shed the light on the sentiment analysis resources and tools developed for Arabic language, so beginners to the Arabic sentiment analysis field could use this paper as a guide.**

***Keywords:* Arabic Sentiment, Lexicons, Datasets, Corpora, Social Media, Dialectal Arabic, Modern Standard Arabic, Sentiment Tools.**

1 INTRODUCTION

Sentiment analysis recently has been the focus of many researchers from a wide range of backgrounds and it is now one of the most important tasks in natural language processing that everyone is seeking to work on. Sentiment analysis or opinion mining are two terms that researchers used interchangeably. However, some researchers declared the differences between them as [1], [2].

Opinion Mining extracts and analyzes people's opinion about an entity while Sentiment Analysis identifies the sentiment expressed in a text then analyzes it. Therefore, the target of SA is to find opinions, identify the sentiments they express, and then classify their polarity [1].

Unfortunately, Arabic language resources and tools are very few compared to English where there is a plenty of resources available free to use. This motivated English sentiment researches to work on sentiment analysis and the outcome results of their work were very satisfied and the progress in the field is growing rapidly and efficiently.

On the other hand, Arabic sentiment researchers devote their time to build resources and tools to help them to encounter the lack existed, but most of the resources made are not available for free and still need further improvements. This paper goal is critically reviewing the available lexicons, data sets, corpora and tools made for Arabic language.

The paper is organized as follows; Section 2 gives background information about Arabic language and challenges in sentiment analysis. Section 3 gives overview about Arabic sentiment analysis architecture, section 4 presents the available Arabic lexicons, Section 5 discusses the available Arabic datasets and corpora, section 6 gives an overview on the used tools, and finally section 7 will conclude the paper.

2 BACKGROUND ABOUT ARABIC LANGUAGE

Arabic language is one of the most used languages in the world; it is the official language of 27 countries and is spoken by more than 422 million people in the Arab world [3].

Arabic language has three different varieties; Classical Arabic (CA) which is the language of Quran (The Holy Book in Islam), Modern Standard Arabic (MSA) which is used in formal contexts (e.g. newspapers, education, television) and Dialectal Arabic (DA) which refers to the colloquial and informal Arabic used in daily communication and mostly seen on the social media platforms.

The number of Arabs on social media is growing rapidly and the amount of opinions, public views and behaviors Arabs express daily is very huge on social platforms and made many researchers head to social media data. The number of Arab Facebook users as of beginning of May 2014 is 81,302,064 up from 54,552,875 in May 2013 [4] and total number of Arab active Twitter users reached 5,797,500 users as of March 2014 [5].

Arabic language made sentiment analysis a very challenging task in terms of its morphological complexity which can be summed up in inflection, agglutination and derivation nature of Arabic, different Arabic dialects present a big challenge, even in MSA [6], Franco-Arabic or Arabic chat Alphabet which Arabs use to write both MSA and DA on social platforms have been one of the most challenging issues in Arabic sentiment Analysis, [7] worked on Franco-Arabic for Algerian Dialect and [8] worked on Franco-Arabic and DA.

3 ARABIC SENTIMENT ANALYSIS ARCHITECTURE

There are four main phases that constitute the cycle of sentiment analysis: preprocessing, feature extraction, feature selection and sentiment classification. Those four phases are similar across the languages; however the implementation of each phase will differ according to the language system. Figure 1 shows the Arabic sentiment architecture.

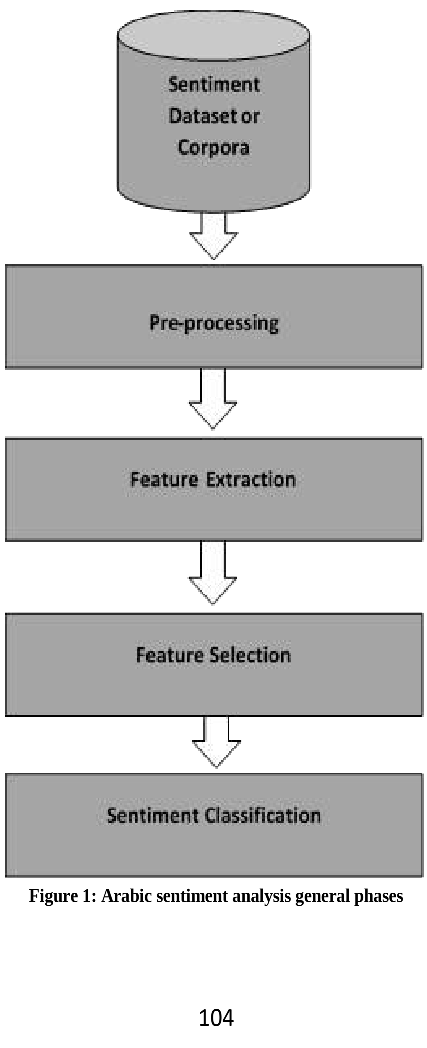


Figure 1: Arabic sentiment analysis general phases

A. Preprocessing

Data preprocessing is one of the most important steps in natural language processing tasks in general and sentiment analysis in particular. Arabic is a morphologically rich language that requires special care during the preprocessing stage [9].

Preprocessing phase includes de-noising (e.g., removing white spaces, HTML tags, special characters, emoticons and URLs), extending abbreviations, tokenization (splitting the text into words), stemming (reducing inflected or derivational words to their word stem), lemmatization (reducing the inflectional forms from each word to a common base or root), normalization (e.g. remove duplicate characters), stop-words removal and negation transformation.

B. Feature Extraction and Selection

The outputs of pre-processing are the extracted text features [10]. Feature extraction methods can be divided into two main approaches lexicon-based approaches and statistical-based approaches.

Lexicon-based approaches usually use a set of seed words or familiar hashtags when dealing with social media platforms and then expanding this seed set using synonyms, antonyms and other resources.

Statistical methods are fully automated methods, which work by extracting linguistic rules from domain-oriented corpus to detect candidate sentiment terms and structures [11].

Those selected text features mainly full under one of four aspects, the first aspect is extracting the terms presence and their frequency (e.g. words and n-grams), the second aspect is concerned with extracting the important parts of speech (e.g. adjectives), the third aspect deals with extracting lists of opinion words and phrases and the fourth aspect deals with negations.

Feature selection is used for filtering irrelevant or redundant features from all your extracted features. Generally, feature selection falls under two main approaches, filter approach and wrapper approach. Filter methods rank the features according to certain metric and select the top-ranked features. Wrapper methods, on the contrary, select the best subset of features by generation and evaluation of different subsets with a classifier [10].

A significance difference between feature extraction and selection is that feature extraction creates brand new features while feature selection keeps a subset of the original features.

C. Sentiment Classification

Sentiment classification deals with classifying a piece of text into positive, negative and neutral. Broadly sentiment classification techniques are divided into three main categories; lexicon-based approach, machine learning approach and hybrid approach.

Machine learning approach relies on machine algorithms to solve sentiment analysis task, machine learning algorithms fall under three main categories; supervised learning, semi-supervised learning and unsupervised learning. Supervised learning methods include four classifiers; decision trees classifiers, linear classifiers (e.g. Support Vector Machines (SVM) and Neural Network), rule-based classifiers, and probabilistic classifiers (e.g. Naïve Bayes, Bayesian and Maximum Entropy).

Lexicon-based Approaches are either Corpus-based approach or Dictionary-based approach. Corpus based approach fall under two methods; statistical and semantic.

Concerning the Arabic language sentiment classification, in [10] survey they declared that there is a dominance of supervised learning over other techniques (semi-supervised, unsupervised, and hybrid techniques) and widely used methods appear to be based on Support Vector Machines (SVM), Naive Bayes (NB), and K-Nearest Neighbors (KNN). Figure 2 shows sentiment classification techniques.

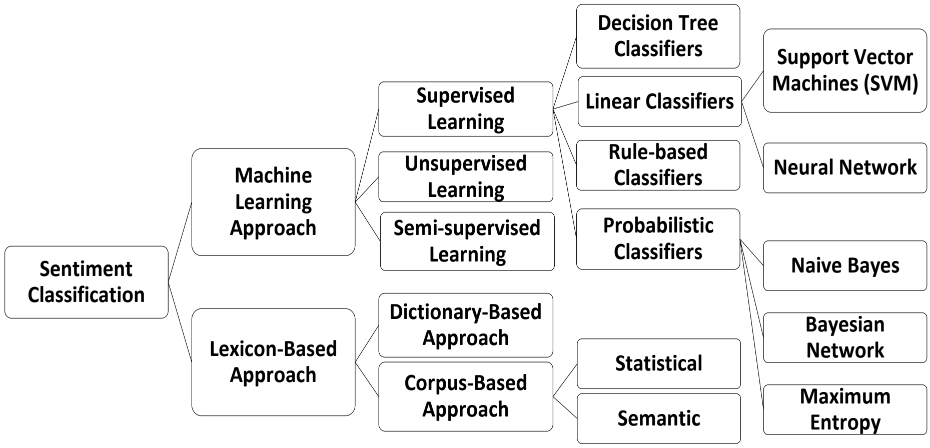


Figure 2: Sentiment classification approaches

4 ARABIC SENTIMENT LEXICONS

The Available Arabic Sentiment lexicons are either automatically or manually created, nowadays there is a trend towards the automatically created lexicons as they save both time and effort. Sentiment lexicons are lists of positive and negative words, optionally with a score indicating the degree of polarity [12].

Lexicons can be based on data collected from social media or reviews or other resources by using machine learning techniques or it can be a translated lexicon from another language (researchers create translated lexicons as a reflection of the limited Arabic resources).

Mohammed et al. [12] presented three sentiment lexicons for Arabic social media that were automatically created based on the idea that hashtag words and emoticons could act as sentiment labels for tweets; the Arabic Emoticon Lexicon was generated by collecting nearly one million tweets that had sad and happy emoticons, the Arabic Hashtag Lexicon was based on the NRC Canada System seed words which was translated into Arabic using Google Translate, the data was pulled from Twitter API and a positive seed hashtag was considered a positive label (pos) and a negative seed hashtag was considered a negative label (neg) [12], the Arabic hashtag lexicon (Dialectal) was man created using seed words taken from a previously created sentiment lexicon by [13]. And work is the best example for showing how researches could benefit from each other work when the resource is made for free use.

They also translated six English sentiment lexicons into Arabic (four of them were manually created and two were automatically created) and they used Google translation to translate the words in the English lexicons, however some words Google were unable to translate. One of the best finding of this work was the observation made about the accuracy of the automatically generated Arabic lexicons which provide accuracy similar to the manual ones around (63%) and the Arabic hashtag lexicon (Dialectal) obtained marked accuracy results (65%). All of the lexicons they presented are available for free use and their work is very valuable in Arabic sentiment analysis.

Badaro et al. [14] created the first publicity available large scale MSA lexicon and it was based on the idea of reusing existed resources to overcome the lack of Arabic resources. They used English WordNet (EWN), Arabic Word-Net (AWN), English SentiWordNet (ESWN) and Standard Arabic Morphological Analyze (SAMA). The creation was based on two approaches; Firstly, Arabic WordNet-based Approach in which a mapping between AWN to ESWN and then mapping between AWN and SAMA and named the resulting lexicon ArSenL-AWN, Secondly the English Gloss-based Approach in which they associated the SAMA lemma entries with English glosses and they named the resulting lexicon as ArSenL-Eng. The ArSenL lexicon was a result of combining the two approaches together.

Eskander & Rambow [15] Presented MSA large scale lexicon, the idea behind SLSA is linking the glosses of AraMorph to the synset terms in SentiWordNet. They also took in consideration while preparing the resources cleaning up the AraMorph especially in cases where part of speech tags was not optimal.

El-Beltagy [16] constructed a lexicon that was a turning point in the field of sentiment analysis; NileULex is an Arabic sentiment lexicon that composes both MSA and Egyptian Colloquial Arabic. Nearly 45% of the lexicon is colloquial and the rest 55% is MSA, NileULex is a phrase and a word level sentiment lexicon, Beside that the lexicon included the most common English transliteration terms and some of the most frequent misspelling words as both are heavily common on social media.

Youssef & El-Beltagy [17] Created MoArLex an Arabic lexicon for use in social media applications and they have taken a new approach in creating their lexicon, rather than creating the lexicon from scratch they extended another lexicon (NileULex [16]) as seed lexicon which showed improvement of the accuracy in sentiment analysis across various datasets. The idea of the expanded lexicon is based on finding the terms that are semantically similar to those in the original seed lexicon that can be found in some accurate manner, and then those terms will most likely be excellent candidates for addition to the lexicon. To the author's best knowledge this work is from the pioneers in expanding lexicon using automatically approaches.

From the previously mentioned work on lexicons it's apparently clear that there is a trend towards building lexicons from social media platforms based on dialectal Arabic data [12], [16], [17], [14], however some lexicons still created for MSA [14], [15]. Table I lists all of the lexicons reviewed with addition of another lexicons which were not reviewed here. Some of the lexicons in Table I are available for use (e.g. [12]).

TABLE XI
Arabic Sentiment Lexicons

Article	**Lexicon**	**Size**	**Construction Approach**	**Source of Data**	**Year**
[12]	Arabic Emoticon Lexicon	43,304 **Positive:** 22,962 **Negative:** 20,342	**Automatic** (Using distant supervision Techniques)	Twitter API	2016
	Arabic Hashtag Lexicon	21,964 **Positive:** 13,118 **Negative:** 8,846	**Automatic** (Using distant supervision Techniques)	Twitter API	
	Arabic Hashtag Lexicon (Dialectal)	20,128 **Positive:** 11,941 **Negative:** 8,179	**Automatic** (Using distant supervision Techniques)	Twitter API	
	English Lexicons Translated into Arabic • NRC Emoticon Lexicon • NRC Hashtag Lexicon		**Automatic**	English Lexicons	

	English Lexicons Translated into Arabic • AFINN • Bing Liu's Lexicon • NRC Emotion Lexicon • MPQA Subjectivity Lexicon	2,476 **Positive:** 878 **Negative:** 1,598	**Manually**	English Lexicons	
		6,789 **Positive:** 2,006 **Negative:** 4,783			
		8,199 **Positive:** 2,718 **Negative:** 4,911 **Neutral:** 570			
		14,182 **Positive:** 2,317 **Negative:** 3,338 **Neutral:** 8,527			
[14]	Arabic sentiment lexicon (Ar-SenL)	**33,995**	**Automatic**	• English SentiWordnet (ESWN) • Arabic Word-Net (AWN) • English SentiWordNet (ESWN)	2014
[15]	SLSA (A Sentiment Lexicon for Standard Arabic)	**34,821**	**Automatic**	• English SentiWordnet (ESWN	2015
[16]	NileULex	**5953** **Compound Positive Phrases:** 416 **Compound Negative Phrases:** 563 **Single Negative Terms:** 369 **Single Positive Terms:** 1281	**Manually**	- Egyptian dialect dataset (NU_EG_Twitter_corpus) • Datasets is one that was collected at a research center in Saudi Arabia	2016
[17]	MoArLex	**36,775**	**Automatic** - Lexicon Expansion	• NileULex Lexicon	2017
[18]	ArSEL (Arabic Sentiment and Emotion Lexicon)	**32,196**	**Automatic**	- DepecheMood - English WordNet - ArSenL	2018
[19]	HILATSA		**semi- automatic** - Combines both lexicon	- ASTD - Mini Arabic Sentiment Tweets Dataset (MASTD)	2019

			based and machine learning approaches	- ArSAS - Arabic Gold Standard Twitter Data for Sentiment Analysis - Syrian Tweets Corpus - Twitter dataset for Arabic Sentiment Analysis (ArTwitter)	
[20]	SentiRDI	18,164 **Positive:** 3,156 **Negative:** 4,169 **Neutral:** 10,839	**Automatic** - If word is not covered by semantic database, the polarity is set **Manually**	- Arabic Semantic Dtabase	2014

5 ARABIC SENTIMENT CORPORA AND DATASETS

There are a very limited available corpora and datasets that are suitable for sentiment analysis tasks. Most of the Arabic sentiment analysis researchers built their own resources as the amount and quality of the freely available are insufficient. Not all Arabic corpora available for text categorization could be used in sentiment analysis for example the Quranic corpus although it is morphologically and syntactically annotated it is not sentimentally tagged and does not provide opinions.

Nabil et al. [21] introduced an Arabic Sentiment Tweets Dataset (ASTD) that was obtained from twitter and its size about 10,000 tweets that were collected in two stages; the first stage by using SocialBakers to determine the most active Egyptian twitter account and in the second stage they used the top trending hashtags in Egypt. They annotated the data using Amazon Mechanical Turk (AMT) Service through an API called Boto and they used four-way sentiment classifications (Subjective Positive, Subjective Negative, Subjective Mixed and Objective) which is very challenging and most of the researchers focus on three-way sentiment classification so their corpus is an addition.

Itani et al. [22] developed a corpus that is based on informal Arabic or dialectal Arabic, the corpus was gathered from Facebook social media platform and they classified the posts as negative, positive, neutral, dual or spam. To the author's best knowledge there is no other work considered spam posts into account. They developed two corpora with the collected posts; the news corpus (NC) from Al Arabiyya News Facebook page and the arts corpus (AC) from the Voice Facebook Page.

Abdellaoui & Zrigui [23] used a distant supervision algorithm to automatically collect and label TEAD (Dataset for Arabic Sentiment Analysis); the data was collected from twitter using the top twenty most used emojis on Twitter according to emojitracker. They classified the tweets as positive, negative and neutral and replaced dialect words with their respective synonyms in MSA using dialect lexicons.
Saleh [24] This corpus represents one of the first sentiment analysis datasets for Arabic language; it was collected manually from movie reviews collected from different web pages and classified reviews into two categories only (positive and negative).

Abdul-Mageed & Diab [25] Created AWATIF a multi-genre MSA corpus, the corpus data relied on three different resources; Wikipedia Talk Pages, Penn Arabic Treebank (PATB) and other web forums. Corpus annotation was divided into three parts; one part was annotated using crowdsourcing on Amazon Mechanical Turk (AMT), second part was annotated by students that received instructions and last part was also annotated by students but it was simpler instructions. One advantage about AWATIF corpus is the fact that its collected from many different domains, but unfortunately this corpus is not available for free use so no other researcher could benefit from.

Aly, & Atiya [26] Built large scale corpus collected from book reviews, they used 1 to 5 score sentiment classification system. And they collected the data from www.goodreaders,com from the first 2143 books in the list of Best Arabic Book. They considered positive reviews those which have ratings 4 or 5, and negative reviews those which ratings 1 or 2. Reviews which have rating 3 are considered neutral and were not included in the polarity classification. To the author's best knowledge most of the works included neutral data in polarity classification.

Elsahar, & El-Beltagy [27] collected multi-genre Arabic reviews corpus from a wide range of domains and their work consider one of the most important works in creating multi–genre corpora. Table II shows more information about the reviewed corpora.

TABLE XIII
SHOWS ARABIC SENTIMENT CORPORA AND DATASETS

Article	Corpus/Dataset	Size	Source of Data	Arabic variety
[21]	Arabic Sentiment Tweets Dataset (ASTD)	10,000 Tweets - 793 Positive - 1684 Negative - 6691 Neutral - 832 Mixed	Twitter	Dialectal Arabic
[22]	The News Corpus (NC)	1000 Posts - 230 Negative - 236 Positive - 161 Dual - 193 Spam - 180 Neutral	Facebook	Dialectal Arabic
	The Arts corpus (AC)	1000 Posts - 224 Negative - 233 Positive - 151 Dual - 197 Spam - 195 Neutral	Facebook	Dialectal Arabic
[23]	TEAD	6 Million Tweets - 3,122,615 Positive - 2,115,325 Negative - 378,003 Neutral	Twitter	Modern Standard Arabic
[24]	OCA (Opinion Corpus for Arabic)	500 Reviews - 250 Positive - 250 Negative	Movie Reviews	Dialectal Arabic
[25]	AWATIF	2,855	Wikipedia TalkPages/Forums	MSA/ Dialectal Arabic
[26]	LABR (Large Scale Arabic Book Reviews)	63,257 (1 to 5 Scale Rate) - 1 - 2 Positive - 3 Neutral - 4 - 5 Negative	GoodReads.com	MSA/ Dialectal Arabic
[27]	Hotel Reviews (HTL)	15,527	TripAdvisor.com	MSA/ Dialectal Arabic
	Restaurant Reviews (RES)	10,970	Qaym.com	MSA/ Dialectal Arabic
	Product Reviews (PROD)	4,272	Souq.com	MSA/ Dialectal Arabic

	Movie Review (MOV)	1,524	Elcinemas.com	MSA/ Dialectal Arabic

6 Arabic Sentiment Analysis Tools

There are few tools available which researchers in Arabic sentiment use, usually those tools help in preprocessing phase or data collection. This section will present the tools used or created for Arabic.

Researchers in [21] used in the first stage of data collection SocialBakers to find out the most active twitter account Egyptian Twitter accounts. They also used Amazon Mechanical Turk (AMT) service for data annotation through Boto API.

In [28] researchers built a tweet collector tool (TCT) by using Java language, the goal behind this tool is helping researches in sentiment analysis. The tool consists of two main phases; the first phase enables the user to search for any specific hashtag or word and retrieve all the related data regardless its size, the second phase consist of extracting sentiment analysis features from dataset.

Researchers in [29] presented a web-based tool for Arabic sentiment analysis using a combination of five parameters as the time of the tweets, preprocessing techniques (e.g. stemming and retweets), n-grams features, lexicon-based methods, and machine-learning methods. The tool was created using R Language by the help of several packages as RWeka, shiny and Twitter, and to the author's best knowledge this is the first web-based tool that target Arabic.

In [30] the researcher presented SentiArabic a lexicon-based sentiment analyzer for Standard Arabic, the analyzer identifies the polarity (positive or negative) of the text. The analyser is based on an extended and modified version of the SLSA a large-scale Sentiment Lexicon for

Standard Arabic, training and evaluation of SentiArabic analyser was based on a new corpus they created for modern standard Arabic and tagged each sentence for polarity.

7 Conclusion

It can be noticed from the previously mentioned work that there is a number of challenges researchers' face when working on Arabic Sentiment Analysis resources and tools but the amount of resources developing is increasing rapidly. Nowadays there is trend towards creating Dialectal Arabic sentiment lexicons which is also reflected in creating dialectal corpora and datasets.

The number of studies in the last five years is concerned with collecting data from social media platforms as the amount of time Arabs spend on social media and the huge content they post about their opinions is growing rapidly.

REFERENCES

[1] Hassan Yousef, Ahmed & Medhat, Walaa & Mohamed, Hoda. (2014). Sentiment Analysis Algorithms and Applications: A Survey. Ain Shams Engineering Journal. 5. 10.1016/j.asej.2014.04.011.

[2] Tsytsarau, M., Palpanas, T. Survey on mining subjective data on the web. Data Min Knowl Disc 24, 478–514 (2012). https://doi.org/10.1007/s10618-011-0238-6.

[3] Boudad, Naaima & Faizi, Rdouan & Rachid, Oulad haj thami & Chiheb, Raddouane. (2017). Sentiment analysis in Arabic: A review of the literature. Ain Shams Engineering Journal. 9. 10.1016/j.asej.2017.04.007.
[4] Facebook in the Arab Region, Arab Social Media Report: http://www.arabsocialmediareport.com/Facebook/LineChart.aspx?&PriMenuID=18&CatID=24&mnu=Cat , (accessed 6 January 2020).
[5] Twitter in the Arab Region, Arab Social Media Report: http://www.arabsocialmediareport.com/Twitter/LineChart.aspx?&PriMenuID=18&CatID=25&mnu=Cat , (accessed 6 January 2020).
[6] Hamdi, Ali & Shaban, Khaled & Zainal, Anazida. (2016). A Review on Challenging Issues in Arabic Sentiment Analysis. Journal of Computer Science. 12. 471-481. 10.3844/jcssp.2016.471.481.
[7] Chader, Asma & Dihia, Lanasri & Hamdad, Leila & Belkheir, Mohamed & Hennoune, Wassim. (2019). Sentiment Analysis for Arabizi: Application to Algerian Dialect. 475-482. 10.5220/0008353904750482.
[8] Duwairi, Rehab & Marji, Raed & Sha'ban, Narmeen & Rushaidat, Sally. (2014). Sentiment Analysis in Arabic tweets. 2014 5th International Conference on Information and Communication Systems, ICICS 2014. 1-6. 10.1109/IACS.2014.6841964.
[9] Duwairi, Rehab. (2014). A Study of the Effects of Preprocessing Strategies on Sentiment Analysis for Arabic Text. Journal of Information Science. 40. 501-513. 10.1177/0165551514534143.
[10] Assiri, Adel & Emam, Ahmed & Aldossari, Hmood. (2015). Arabic Sentiment Analysis: A Survey. International Journal of Advanced Computer Science and Applications. 6. 10.14569/IJACSA.2015.061211.
[11] Tedmori, Sara & Awajan, Arafat. (2019). Sentiment Analysis Main Tasks and Applications: A Survey. Journal of Information Processing Systems. 15. 500-519. 10.3745/JIPS.04.0120.
[12] Mohammad, Saif & Salameh, Mohammad & Kiritchenko, Svetlana. (2016). Sentiment Lexicons for Arabic Social Media.
[13] Refaee, E., & Rieser, V. (2014). An Arabic Twitter Corpus for Subjectivity and Sentiment Analysis. In N. Calzolari (Ed.), *Proceedings of the 9th International Conference on Language Resources and Evaluation* (pp. 2268-2273). European Language Resources Association.
[14] Badaro, Gilbert & Baly, Ramy & Hajj, Hazem & Habash, Nizar & El-Hajj, Wassim. (2014). A Large Scale Arabic Sentiment Lexicon for Arabic Opinion Mining. 10.3115/v1/W14-3623.
[15] Eskander, Ramy & Rambow, Owen. (2015). SLSA: A Sentiment Lexicon for Standard Arabic. 2545-2550. 10.18653/v1/D15-1304.
[16] El-Beltagy, Samhaa. (2016). NileULex: A Phrase and Word Level Sentiment Lexicon for Egyptian and Modern Standard Arabic.
[17] Youssef, Mohab & El-Beltagy, Samhaa. (2018). MoArLex: An Arabic Sentiment Lexicon Built Through Automatic Lexicon Expansion. Procedia Computer Science. 142. 94-103. 10.1016/j.procs.2018.10.464.
[18] Badaro, Gilbert & Jundi, Hussein & Hajj, Hazem & El-Hajj, Wassim & Habash, Nizar. (2018). ArSEL: A Large Scale Arabic Sentiment and Emotion Lexicon.
[19] Elshakankery, Kariman & Farouk, Mona. (2019). HILATSA: A hybrid Incremental learning approach for Arabic tweets sentiment analysis. Egyptian Informatics Journal. 20. 10.1016/j.eij.2019.03.002.
[20] Mobarz, Hanaa & Rashown, Mohsen & Farag, Ibrahim. (2014). Using Automated Lexical Resources In Arabic Sentence Subjectivity. International Journal of Artificial Intelligence & Applications. 5. 01-14. 10.5121/ijaia.2014.5601.
[21] Nabil, Mahmoud & Aly, Mohamed & Atiya, Amir. (2015). ASTD: Arabic Sentiment Tweets Dataset. 2515-2519. 10.18653/v1/D15-1299.
[22] Itani, Maher & Roast, C.R. & Al-Khayatt, Samir. (2017). Developing Resources For Sentiment Analysis Of Informal Arabic Text In Social Media. Procedia Computer Science. 117. 129-136. 10.1016/j.procs.2017.10.101.
[23] Abdellaoui, Houssem & Zrigui, Mounir. (2018). Using Tweets and Emojis to Build TEAD: an Arabic Dataset for Sentiment Analysis. Computacion y Sistemas. 22. 10.13053/CyS-22-3-3031.
[24] Saleh, Mohammed & Martín-Valdivia, Maria & López, L. & Perea-Ortega, José. (2011). OCA: Opinion corpus for Arabic. JASIST. 62. 2045-2054. 10.1002/asi.21598.

[25] Abdul-Mageed, Muhammad & Diab, Mona. (2012). AWATIF: A Multi-Genre Corpus for Modern Standard Arabic Subjectivity and Sentiment Analysis. Proceedings of LREC.
[26] Aly, Mohamed & Atiya, Amir. (2013). LABR: A Large Scale Arabic Book Reviews Dataset. 10.13140/2.1.3960.5761.
[27] Elsahar, Hady & El-Beltagy, Samhaa. (2015). Building Large Arabic Multi-domain Resources for Sentiment Analysis. Lecture Notes in Computer Science. 9042. 23-34. 10.1007/978-3-319-18117-2_2.
[28] Abdallah, Emad & Abo-Suaileek, Sarah. (2019). Feature-based Sentiment Analysis for Slang Arabic Text.
[29] El-Masri, Mazen & Altrabsheh, Nabeela & Ahmed, Hanady & Ramsay, Allan. (2017). A web-based tool for Arabic sentiment analysis.
[30] Eskander, Ramy. "SentiArabic: A Sentiment Analyzer for Standard Arabic." LREC (2018).

I. ACKNOWLEDGMENT

First and foremost, I would like to thank God for giving me the strength, knowledge, ability and opportunity to undertake this research. I also would like to thank my family and all my professors especially Dr. Sameh Alansary for encouraging me to take this research.

BIOGRAPHIES

Miramar Etman: *Master Degree student at Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt*

She graduated from Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, (2018), graduated from Information Technology Institute (ITI), Ministry of communications and Training, Track: Open Source Applications Development, Intake 39. Her research interests include natural language and speech processing and Programming languages.

Dr. Sameh Alansary: *Director of Arabic Computational Linguistic Center at Bibliotheca Alexandrina, Alexandria, Egypt.*

He is professor of computational linguistics in the Department of Phonetics and Linguistics and the head of Phonetics and Linguistics Department, Faculty of Arts, Alexandria University. He obtained his MA in Building Arabic Lexical Databases in 1996, and his PhD from Nijmegen University, the Netherlands in building a formal grammar for parsing Arabic structures in 2002. His main areas of interest are concerned with corpus work, morphological analysis and generation, and building formal grammars. He is also the head of Arabic Computational Linguistics Center in Bibliotheca Alexandrina; supervising and managing the Universal Networking Language project in Library of Alexandria since 1-6- 2005 till now. Dr. Alansary is the co-founder of the Arabic Language Technology Center (ALTEC), an NGO aims at providing Arabic Language resources and building a road map for Arabic Language Technology in Egypt and in the Middle East. He has many scientific works in Arabic Natural Language Processing published in international conferences and periodicals, and a member in many scientific organizations: (1) Egyptian Society of Language Engineering, Cairo, (2) Arabic Linguistic Society - USA, (3) Association of Computational Linguistics - USA – Europe, (4) Universal Networking Language foundation, United Nations, Geneva, Switzerland.

مراجعة نقدية لموارد وأدوات اللغة العربية الخاصة بتحليل الآراء والمشاعر

Miramar Etman[1], Sameh Alansary[2]

Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt

[1]miramar.etman@alexu.edu.eg

[2]sameh.alansary@bibalex.org

ملخص **— ينمو مجال تحليل الآراء والمشاعر بشكل سريع في الوقت الحالي، حيث يكرس الباحثون من التخصصات المختلفة وقتهم وجهودهم لاستخراج واكتشاف الآراء والمشاعر للأشخاص اتجاه موضوع معين أو منتج أو علامة تجارية أو خدمات. ومع تزايد عدد الأشخاص على الإنترنت حيث يقضي معظم الأشخاص وقتهم في التعبير عن آرائهم وردود أفعالهم على منصات التواصل الاجتماعي، وهذا جعل الباحثون يتجهون لجمع البيانات من منصات التواصل الاجتماعي. اكتسب تحليل الآراء والمشاعر في اللغة العربية الكثير من الاهتمام مؤخرًا. ومع ذلك ، لا توجد موارد وأدوات للغة متاحة بشكل كبير عند مقارنتها باللغة الإنجليزية. ستحاول هذه الورقة إلقاء الضوء على مصادر وأدوات تحليل الآراء والمشاعر المطورة للغة العربية، لذلك يمكن للمبتدئين في مجال تحليل الآراء والمشاعر في العربية استخدام هذه الورقة كدليل.**

الكلمات الدالة: **المشاعر العربية ، المعاجم ، مجموعات البيانات ، المدونات اللغوية ، وسائل التواصل الاجتماعي ،العامية العربية ، العربية الفصحى ، أدوات تحليل المشاعر.**

A Pilot Study of Biber's Model for Language Variation Detection: A Language Engineering Approach

Maram Elsaadany

Pharos University, Alexandria, Egypt

maram.elsaadany@pua.edu.eg

Sameh Alansary

Bibliotheca Alexandrina, Alexandria, Egypt
Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt

sameh.alansary@bibalex.org

***Abstract* — this paper is primarily a translation analysis of Tramp's speech and a letter sent to President Trump regarding family separation on "The Leadership Conference on Civil and Human Rights". Google translate is the engine used to translate these two texts into Arabic. The data selected for this analysis is 1000 words in each script. Biber's model (1988) used 67 features to prove that writing is more complicated than speech. The study enforces Biber's claim that writing is more complicated than speech. The findings assure Biber's claim as there were lots of problems in the translation of the speech text into Arabic in comparison to the translated written texts. The Results clarify the fact that google translate has to be adapted and equipped with a new grammar for speech that is different than the one used for writing to achieve the best outcome for both translations in the same language. This paper is a pioneer in its application as there is no research paper adapted such claim in the field of translation.**

Key words: translation, computational, linguistic variations, speech, writing

1 INTRODUCTION

Computational linguistics (CL) combines resources from linguistics and computer science to discover how human language works. Computational linguistics is a vital field in the information age. According to [1],computational linguists create tools for important practical tasks such as machine translation, speech recognition, speech synthesis, information extraction from text, grammar checking, text mining and more. [2]has stressed the idea that contrastive Analysis (CA) is a method that is connected to Contrastive Linguistics, which is considered a branch of linguistics that focuses on illustrating the differences and similarities among two or more languages at different linguistic levels as semantics, syntax, and phonology.

According to [3] earlier programs have been criticized by the lack of a dictionary; to identify linguistic features, they relied on small lists of words that were built into the program structure itself. These lists included prepositions, conjuncts, pronominal forms, auxiliary forms. Since these word lists were relatively restricted, the grammatical category of many words in texts could not be accurately identified, and therefore these programs could not identify all of the occurrences of some linguistic features. The programs have been designed to avoid skewing the frequency counts of features in one genre or another so that the relative frequencies were accurate. The main disadvantage of this earlier approach was that certain linguistic features could not be counted at all. For example, there was no way to compute a simple frequency count for the total nouns in a text, because nouns could not be identified. For these reasons, the second set of programs has been taking place.

The second stage of program development took place during the years (1985-1986). The approach used in this stage is different from that of the first stage. As a result, a general tagging program to identify the grammatical category of each word in a text was developed. The aim is to develop a program that was general enough to be used for tagging both written and spoken texts. For example, the program could not depend on upper case letters or sentence punctuation. This goal is achieved by using a large-scale dictionary together with a number of context-dependent disambiguating algorithms. The main problem that had to be solved is that many of the common words in English are ambiguous as to their grammatical category. Words like "absent" can be either adjectives or verbs; words like "acid" can be either nouns or adjectives. All past and present participial forms can function as noun (gerund), adjective, or verb. A simple word like that can function as a demonstrative, demonstrative pronoun, relative pronoun, complementizer, or adverbial subordinator.

[4] has developed algorithms to disambiguate occurrences of certain words, depending on their surrounding contexts. For example, a participial form preceded by an article, demonstrative, quantifier, numeral, adjective, or possessive pronoun is functioning as a noun or adjective. That is to say, it is not functioning as a verb in this context; given this preceding context, if the form is followed by a noun or adjective then it will be tagged as an adjective; if it is followed by a verb or preposition, then it will be tagged as a noun. Tagged texts enable automatic identification of a broad range of linguistic features that are major for differentiating between genres in English. The tagged texts are subsequently used as input to other programs that count the frequencies of certain tagged items (e.g. nouns, adjectives, adverbs) and compute the frequencies of particular syntactic constructions (e.g. relativization on subject versus non-subject position).There has also been a debate concerning the need for a linguistic comparison of speech and writing. Historically, academics have regarded writing as the true form of language, while speech has been considered to be unstable and not worthy of study. By the early twentieth century, linguists regarded speech as primary and writing as a secondary form of language derived from speech; thus only speech was considered worth serious linguistic analysis. In fact, the historical view that written, literary language is true language continues as the dominant perception to the present time. it might well be the case that neither speech nor writing is primary; that they are rather different systems, both deserving careful analysis. This is in fact the view advocated by some researchers studying communicative competence. In this paper, the researcher is comparing the translation of google translate program in two different forms of the same language, speech and writing. In other words, are the two translations accurate? Is one of them better than the other and why?

2 RESEARCH QUESTION

1. From a language engineering point of view, can a computer program which is capable of translating written texts translate speech texts with same accuracy?

2. Is the translated text in the written form less erroneous than the text translated in the speech form?

3 BIBER'S VARIATIONS

[4] model is the main model upon which this study is based. The initial step is to collect the English and Arabic texts that are used as the data used in this study. Next, text normalization is crucial for any comparison of frequency counts across texts, because text length can vary widely. Translation studies, have only evolved during the last decades [5]. Scientific research in this area is a very recent phenomenon, as stressed by [6]. The call for research in translation is overwhelming as "a whole range of issues seemed to be waiting for examination, and inquiry is overdue". [7] Grammatical competence is concerned with the linguistic structure of 'grammatical' utterances; communicative competence is concerned with the form and use of all language - both speech and writing. Within this framework, neither speech nor writing needs to be considered primary to the exclusion of the other. Rather, both require analysis, and the linguistic comparison of the two modes becomes an important question.

4 THE SAMPLE

The sample used for this pilot study consists of two texts .The first one is the script that represents Donld Trump's speech. It was a telephone conversation with President Zelenskyy of Ukraine. The second text is extracted from a letter written to Donald Tramp regarding the family separation. The machine translation used is Google Translate that translates both texts into Arabic. In order to normalize texts, in this study, the frequency counts of all linguistic features are normalized to a text length of 1000 words so we have to delete some words to normalize the texts to have the same length.

5 Analysis

Translation is a complicated task, during which the meaning of the source-language text should be conveyed to the target-language readers. In other words, translation can be defined as encoding the meaning and form in the target language by means of the decoded meaning and form of the source language. [8] listed eight strategies, which have been used by professional translators, to cope with the problematic issues while doing a translation task as translation by a more general word, translation by a more neutral/ less expressive word, translation by cultural substitution, translation using a loan word, translation by paraphrase using a related or unrelated word, translation by omission or by illustration. Cohesion is the network of lexical, grammatical and other relations which provides links to various parts of the text. These relations organize a text and to some extent create it. One example of this is reference to other words and expressions in the surrounding sentences and paragraphs. Google translation tends to misuse references I in the speech texts. Most of the time pronouns are not with clear references. This is really clear with the pronoun (It) as there is no clear reference to whom it refers.

For example:

Much more than the European countries are doing and they should be helping you more than they are

Google Translation

أكثر بكثير مما تفعله الدول الأوروبية وينبغي أن يساعدك أكثر مما هو عليه

Here, google translation misused the pronoun as the correct translation should be

أكثر بكثير مما تفعله الدول الأوروبية وينبغي أن تساعدك أكثر مما هو عليه

E.g.

They are not working as much as they should work for Ukraine

Google Translation

انهم لا يعملون بقدر ما ينبغي أن تعمل من أجل أوكرانيا

Correct translation

انهم لا يعملون بقدر ما ينبغي أن يعملوا من أجل أوكرانيا

The problem of non-equivalence was apparent in google translation as the source language concept is not lexicalized in the target language. The source language word may express a concept which is known in the target culture but simply not lexicalized. In other words [8] stated that there is no specific term in the target language for this word that is not allocated in the target language word to express it. For example, the word (run more) this means to exceed or to go over. However, google misuse this by translating it into the literal meaning of the word which is totally incorrect .This means that the target language lacks this term.

E.g.

When I was speaking to Angela Merkel she talks Ukraine, but she doesn't do anything.

Google translation:

عندما كنت أتحدث إلى أنجيلا ميركل تتحدث إلى أوكرانيا ، لكنها لا تفعل أي شيء.

This is a literal translation that implies a vague and unclear message

Correct translation

على سبيل المثال عند التحدث إلى أنجيلا ميركل تتحدث عن اوكرانيا دون فعل شى من أجلها

E.g.

I told them that they are not doing quite as much as they need to be doing on the issues with the sanctions

Google translation

إنهم لا يقومون بالقدر الذي يجب عليهم فعله بشأن القضاياالمتعلقة بالجزاءات

Correct translation

إنهم لا يقومون بالقدر الذي يجب عليهم فعله بشأن القضايا المتعلقة بالجزاءا

E.g.

It turns out that even though logically, the European Union should be our biggest partner but technically the United States is a much bigger partner than the European Union

Google Translate:

أنه على الرغم من المنطق ، يجب أن يكون الاتحاد الأوروبي أكبر شريك لنا ولكن الولايات المتحدة من الناحية الفنية شريك أكبر بكثير من الاتحاد الأوروبي

Correct translation

علي الرغم من ان الاتحاد الاوروبي هوشريكنا الأكبر لكن فعليا الولايات المتحده هى اقوى واهم شريك لنا خاصا عند فرض العقوبات على الاتحاد الأوروبي

E.g.

We are almost ready to buy more Javelins from the United States for defense purposes.

Google translate

على وجه التحديد نحن على وشك. على استعداد لشراء المزيد من الرمح من الولايات المتحدة لأغراض الدفاع.

Correct translation

ونحن على استعداد لشراء السلاح اللازم من الولايات المتحدة لأغراض دفاعيه

E.g.

They are not working as much as they should work for Ukraine.

Google translate

انهم لا يعملون بقدر ما ينبغي أن تعمل من أجل أوكرانيا.

Correct translation

لقد قمت بالتحدث إلى انجلينا ماركل بشان الجهود الضئيله المبذولة فى قضايا العقوبات

Presupposed meaning arises from the co-occurrence restrictions on what other words or expressions that we expect to see before or after a particular lexical Item. This was clear in google 's translation we spend a lot of effort and a lot of time as collocational restrictions these are semantically arbitrary restrictions which don't follow the prepositional meaning of a word.

E.g.

Also, I think I should run more often so you can call me

Google translate

أعتقد أنني يجب أن أجري أكثر من مرة حتى تتمكن من الاتصال بي

Correct translation

واعتقد أن على بالاستمرار و مواكبه هذا التقدم لكى يستمر تواصلنا و علاقتنا ببعضنا البعض

Definitely, the implied meaning from this sentence is not "running as in a race" and that is why google misused the word as it should be المحاوله في الاستمرار

[8] referred to problems of non-equivalence above the word level. She suggested lots of strategies to solve such problem as translation by paraphrase. This is also related to google's ignorance of the fixed expressions of the target language.

For example:

We all watched from the United States and you did a terrific job. The way you came from behind

Google Translation:

شاهدنا جميعًا من الولايات المتحدة وقمت بعمل رائع. الطريقة التي أتيت بها من الخلف

Correct translation:

تهانينا على النصر العظيم الذى شهدناه جميعا من الولايات المتحده لقد قمت بعمل اكتر من رائع لقد اجتنزت الفرصه و انتصرت ببراعه

Here the words connote a different meaning rather than the expressed one. The literal translation of the words together does not deliver the right message as collocations and idiomatic expressions are really problematic in their translations. This according to Baker is referred to as translation beyond the word level. Register is a variety of language that a language user considers suitable to a specific situation. Its variation occurs from variations in the field of discourse as the linguistic choices will vary whether the speaker is taking part in a political speech or something else. Google misinterpreted this formality as in translating main words such as (congratulations) and (win big).

The problem of nonequivalence was apparent in google translation in (crowd stick) as this level of difficulty posed can vary tremendously depending on the nature of non-equivalence. Different kinds of non-equivalence require different strategies that the translator should be aware of. However, some strategies are difficult to handle.

Grammar is a set of rules which determines the way in which units such as words and phrases can be combined in a language .Google's translation totally neglected the English structure in the way it translated the sentences into Arabic .A language can of course express any kind of information its speakers need to express, but without using the same grammatical structure

For example

They are not working as much as they should work

Word by word translation is another fatal mistake that google has committed this way of translating word by word reflects the lack of terms and items in the target language which makes the final translation very poor and totally incorrect.

The tension between accuracy and naturalness is always apparent in google translation as Mona Baker stated that it is not easy for a translator to produce a collocation which is typical to the target language. This ideal cannot be achieved with the source collocation. For example, the word earlier is translated in a totally wrong way.

6 FEATURES OF SPEECH

1. Integration: It refers to the way in which a large amount of information is packed into relatively few words in typical writing, because the writer operates under few time constraints and can therefore construct a carefully packaged text. In contrast, typical speech cannot be highly integrated because it is produced and comprehended on-line. Features that are used to integrate information into a text include attributive adjectives, prepositional phrase series, phrasal coordination, and careful word choice.

2. Fragmentation: It refers to the linguistic characteristics of texts produced under severe time constraints, the case for typical speech. Under these conditions, information cannot be carefully incorporated into the text, and the resulting structure is much looser, or fragmented. Linguistic features associated with a fragmented text include clauses strung together with simple conjunctions (e.g., and) or with no connectives at all.

3. Involvement: This refers to those linguistic features which reflect the fact that speaker and listener typically interact with one another, while writer and reader typically do not. Due to this interaction, speakers often make direct reference to the listener (by use of second person pronouns, questions, imperatives, etc,), and they are typically concerned with the expression of their own thoughts and feelings (e.g., marked by use of first person pronouns, affective forms such as emphatics and amplifiers, and cognitive verbs such as think and feel). As a result of this concern, speech often has a distinctly non-informational and imprecise character (marked by hedges, pronoun it, and other forms of reduced or generalized content). These features can be considered together as the characteristics of involved text. In contrast, detachment refers to the characteristics of typical writing which result from the fact that writer and reader usually do not interact (e.g., marked by agentless passives and nominalizations).

7 FEATURES OF WRITING

1. [9] stated that writing is more structurally complex and elaborate than speech, indicated by features such as longer sentences or T- units and a greater use of subordination.
2. As for [10] more explicit than speech, in that it has complete idea units with all assumptions and logical relations encoded in the text.
3. [11] explained that it is more decontextualized, or autonomous, than speech, so that it is less dependent on shared situation or background knowledge.
4. [9] also illustrated that less personally involved than speech and more detached and abstract than speech.
5. [12] showed that it is characterized by a higher concentration of new information than speech (Brown and Yule 1983); and more deliberately organized and planned than speech. [12] noted that 'in writing we have time to mold a succession of ideas into a more complex, coherent, integrated whole', whereas speech, because it is produced on-line, is more fragmented.

8 CONCLUSIONS

There is a huge distinction between speaking and writing in their characteristics. There is also a difference in the channel. There may be many sub-channels available in speaking but only the lexical-syntactic sub-channel available in writing. Also, the opportunity for interaction with the text varies. In one hand, there are no real-time constraints in writing. On the other hand, severe real-time constraints appear in speech. Even these two differences are not absolute. Features such as underlining, bold-face, and certain punctuation marks can be used to represent prosodic or paralinguistic sub-channels in writing. Tape-recorded speech bypasses some of the real-time constraints of speech, more so in comprehension than in production. The recommendations of this study is to build a new grammar to be used in translating speech which is totally different than the grammar used in translating written texts of the same language to avoid the problems that exist due to the differences between the two forms of the same language.

REERENCES

[1] Mitkov, R. (ed.).(2015). The Oxford handbook of computational linguistics. 1st ed. Oxford University Press: Oxford

[2] Fisiak, J. (1981). Some introductory notes concerning contrastive linguistics. In J. Fisiak (Ed.), Contrastive linguistics and the Language Teacher (pp. 1-11). Oxford: Pergamon.

[3] Towell, R. & Hawkins, R. (1994). Approaches to second language acquisition. Clevedon: Multilingual Matters.

[4] Biber, D. (1988). Variation across speech and writing, Cambridge: Cambridge University Press.

[5] Broeck, R. (1986). Contrastive discourse analysis as a tool for the interpretation of shifts in translated texts. In House, J. & Blum Kualka, S. (Eds.) Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies, Tubingen: Gunter Narr P. 37-47.

[6] Gile, D. (1994) Beyond testing towards a theory of educational assessment, London: Falmer Press.

[7] Simon, S. (1996). Gender in Translation: Cultural Identity and Politics of Translation. London and New York: Routledge.

[8] Baker, M. (1992). In other words: A course book on translation. London: Routledge.

[9] CHAFE, Wallace. 1985. Linguistic Differences Produced by Differences Between Speaking and Writing. In OLSON, D., ORRANCE, N. & HILDYARD, A. Literacy, Language and Learning. Cambridge: Cambridge University Press, 1985. pp. 105-123.

[10] Coulmas, Florian. 1996. The Blackwell encyclopedia of writing systems. Oxford: Blackwell.

[11] Dewaele, J.-M. (2000). Gender, social and situational variables in the choice of speech style in native Dutch. Paper presented at the Sociolinguistics Symposium 2000, Bristol

[12] Coffin, Caroline et al. (2003). Teaching Academic Writing: A Toolkit for higher education. Routledge (London).

BIOGRAPHY

She has attained her PH-D degree in Translation studies from the Institute of Applied Linguistics and Translation (2018) , Faculty of Arts, Alexandria University. She has also attained her Masters degree in Applied Linguistics in (2014). Her main areas of interest are applied linguistics, computational linguistics and computational studies in the field of translation. Elsaadany is holding the position of a lecturer at Pharous University.

She is a former English instructor in the Arab Academy for Science and Technology. Furthermore, she is a speaking examiner for IELTS and OET exams.

Dr. Sameh Alansary: *Director of Arabic Computational Linguistic Center at Bibliotheca Alexandrina, Alexandria, Egypt.*

He is professor of computational linguistics in the Department of Phonetics and Linguistics and the head of Phonetics and Linguistics Department, Faculty of Arts, Alexandria University. He obtained his MA in Building Arabic Lexical Databases in 1996, and his PhD from Nijmegen University, the Netherlands in building a formal grammar for parsing Arabic structures in 2002. His main areas of interest are concerned with corpus work, morphological analysis and generation, and building formal grammars. He is also the head of Arabic Computational Linguistics Center in Bibliotheca Alexandrina. He is supervising and managing the Universal Networking Language project in Library of Alexandria since 1-6-2005 till now.

Dr. Alansary is the co-founder of the Arabic Language Technology Center (ALTEC), an NGO aims at providing Arabic Language resources and building a road map for Arabic Language Technology in Egypt and in the Middle East. He has many scientific works in Arabic Natural Language Processing published in international conferences and periodicals, and a member in many scientific organizations: (1) Egyptian Society of Language Engineering, Cairo, (2) Arabic Linguistic Society - USA, (3) Association of Computational Linguistics - USA – Europe, (4) Universal Networking Language foundation, United Nations, Geneva, Switzerland.

للكشف عن التباين اللغوي: منهج هندسة اللغة Biber دراسة تجريبية لنموذج

مرام السعدني

جامعه فاروس- الاسكندريه- مصر

maram.elsaadany@pua.edu.eg

سامح الانصاري

مصر -مكتبه الاسكندريه -الاسكندريه

قسم الصوتيات- كلية الاداب- جامعة الاسكندرية

sameh.alansary@bibalex.org

ملخص - هذه الورقة هي في المقام الأول تحليل ترجمة لخطاب ترامب ورسالة مرسلة إلى الرئيس ترامب بشأن انفصال الأسرة عن "مؤتمر القيادة حول الحقوق المدنية والإنسانية". ترجمة جوجل هو المحرك المستخدم لترجمة هذين النصين إلى العربية. البيانات المحددة لهذا التحليل هي 1000 كلمة في ميزة لإثبات أن الكتابة أكثر تعقيدًا من الكلام. تفرض الدراسة إدعاء بيبر بأن الكتابة أكثر تعقيدًا من الكلام. تؤكد النتائج كل برنامج نصي. استخدم نموذج ادعاءات لأن هناك الكثير من المشاكل في ترجمة نص الكلام إلى العربية مقارنة بالنصوص المكتوبة المترجمة. توضح النتائج حقيقة أن يجب أن يتم تكييفها وتجهيزها بقواعد جديدة للتعبير تختلف عن تلك المستخدمة في الكتابة لتحقيق أفضل نتيجة لكلا الترجمتين في نفس اللغة. تعتبر هذه الورقة رائدة في تطبيقها حيث لا توجد ورقة بحثية تم تكييفها لهذه المطالبة في مجال الترجمة.

الكلمات المفتاحية: الترجمة ، الاختلافات الحسابية ، اللغوية ، الكلام ، الكتابة

Towards Building a Semantic Role Labeling System for Modern Standard Arabic: A Rule-based Approach

Amena Helmy Abd El-khaleq Muhammad Dheif[1], Sameh Saad Abu-Elmagd Alansary[2]

Department of Phonetics and Linguistics, Faculty of Arts, Alexandria University.

amena.helmy@alexu.edu.eg

Sameh.Alansary@bibalex.org

Abstract— ***Semantic role labeling (SRL), the computational identification and labeling of arguments in text, has become a leading task in computational linguistics today, as it is a step towards natural language understanding (NLU). Many linguistic theories had discussed the nature of semantic roles, on the other hand, language engineers headed directly to build automatic statistical-based and rule-based SRL systems that would predict the correct semantic roles of a predicate argument structure. Many systems, corpora, and lexical resources have been developed to tackle the task of semantic role labeling in many languages, but Arabic lacks such attention. The study's main goal is to build a rule-based semantic role labeler for Modern Standard Arabic (MSA), which is the first rule-based SRL system to be developed on MSA, that works on the verbal predicates of the top frequent 40 Arabic VerbNet (AVN) classes. The study defined the input data to be gold-standard fully syntactic trees drawn from the Arabic Penn Treebank (ATB) part1 and part3. Forty knowledge bases were constructed to store the syntactic and semantic features that realize the semantic roles of each AVN class predicates. Features of each AVN class were extracted from the training data of the frequent predicate, but it was used to represent the whole class. The developed system was tested on two types of testing data, the first type is drawn from the sentences of the same frequent verb the knowledge base is built upon and the second type is drawn from the sentences of the other verbs that belong to the same AVN class. The current system achieves a final F1 score of 92.9% over the first type of testing data, F1 score of 86.0% over the second type of testing data and a 91.0% F1 score on both.***

Keywords*: Semantic role labeling, VerbNet, Predicate, Core semantic role, Adjunct semantic role.*

1 INTRODUCTION

Semantic parsing is one of the main topics of NLU that has undergone intense study after syntactic parsing was found to be insufficient for providing the layer of meaning representation needed for proper understanding of natural language. Semantic parsing has mainly two types, the first type is deep semantic parsing and it is defined as "Mapping a natural language sentence to a detailed representation of its complete meaning in a fully formal language that has a rich ontology of types, properties, and relations, and supports automated reasoning" [1]. The Abstract meaning representation (AMR) project [2] is one of the main projects that seeks to provide a fully semantic parsed corpora for English. The second type of semantic parsing is Semantic Role Labeling (SRL) or shallow semantic parsing which is the core topic of the current study. SRL is considered as a shallow semantic parsing because it represents only a subtype of all the semantic relations, which is the predicate-argument semantic relations. SRL task can be defined as the automatic analysis of an input text into the number of propositions that compose it, where a proposition consists of a predicate and its set of arguments, then map the predicate arguments to their semantic relations [3] to properly answer the question of who did what to whom and perhaps when where and why[4]. Figure (1) shows a shallow semantic representation of sentence (1) where it consists of only one proposition and maps each argument to its correct semantic relation.

1) المحققُ قبضَ على المتهمِ في مسرحِ الجريمةِ مساءً .

- al-muḥaqiqu qabaḍa ʻalá al-mutahami fī masraḥi al-jarīmati masāʼan.[4]

[4] Arabic transliteration is provided according to the ALA-LC Romanization Scheme which provides transliteration for Non-Roman languages and is approved by the Library of Congress and the American Library Association [5].

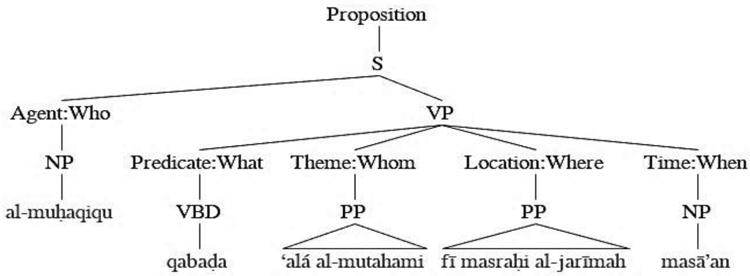


Figure 1: A shallow semantic representation of sentence (1)

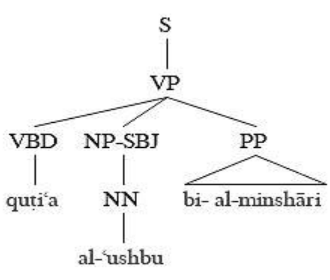
A predicate is the main unit in any proposition, Arab grammarians call predicates the “musnad ” of the proposition. Ref. [6] defines the predicate as “any word (or sequence of words) which (in a given single sense) can function as the predicator of a sentence”. In turn, the predicator is defined as “the word (sometimes a group of words) which does not belong to any of the referring expressions and which, makes the most specific contribution to the meaning of the sentence. Intuitively speaking, the predicator describes the state or process in which the referring expressions are involved.”[6].

The second main unit that makes up a proposition is the predicate arguments. Each argument plays a certain semantic role assigned by the predicate. Predicates assign their arguments two kinds of semantic roles namely core semantic roles and adjunct (or peripheral) semantic roles. On one hand, an argument would be classified as playing a core role if it is essential and central to the event or state being described by the predicate, on the other hand, an argument would be seen as peripheral or adjunct if it provides extra information about the event or the sate being described by the predicate [7].

As mentioned above, SRL is considered important as it provides a layer of shallow semantic representation that depicts the predicate-argument semantic relations which is a major step towards NLU. But what makes SRL more important is that SRL systems abstract and unifies the different argument structure realizations of a predicate which is called “diathesis alternations” [8] into a unified semantic representation [9]. This contribution takes SRL a step ahead of syntactic parsing, for example, in both sentences (2) and (3), the argument “al-‘ushb (the grass)” plays the patient semantic role, though it is realized as a subject in sentence (2) and as an object in sentence (3). The reason behind the different realizations of the same role is the voice changing of the verb in the two sentences. Another difference can be seen with the argument “al-minshār” where it’s realized as an indirect object in sentence (2) and as a subject in sentence (3) while it plays the instrument semantic role in both sentences. The reason behind the different realizations of “al-minshār” argument is the instrument alternation of predicate “qaṭa‘” [10]. Figure (4) shows the shallow semantic representation both sentences (2) and (3) map to.

(2 قُطِعَ العشبُ بالمنشارِ. (quṭi‘a al-‘ushbu bi-al-minshāri)

(3 المنشارُ قطعَ العشبَ.(al-minshāru qaṭa‘a al-‘ushba.)



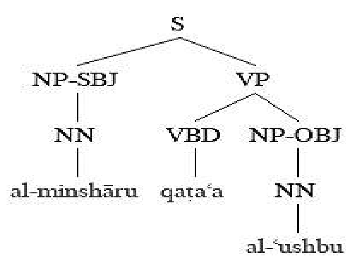


Figure 2: syntactic tree of sentence (2). **Figure 3: syntactic tree of sentence (3).**

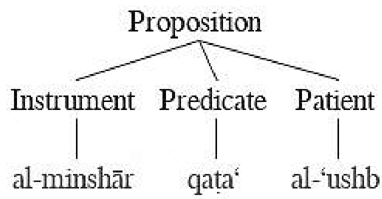


Figure 4: The shallow semantic layer both sentence (2) and sentence (3) map to.

Of course, Many higher-level natural language processing tasks like Information extraction, question answering, automatic summarization, and machine translation rely on the layer of semantic representation provided by an SRL system and consider it as an intermediate stage that need to be performed before pursuing to its main task since it is easier to work on a layer of semantic representation than working on a raw text [11].

2 LITERATURE REVIEW

The notion of semantic roles had been studied thoroughly and extensively in many linguistic theories. It has a numerous different terms across different theories, as it is termed deep semantic cases [12] Theta roles by Chomsky and his followers [13], proto-roles [14], thematic relations by [15], [16] and semantic macro roles [17]. The term "semantic roles" is the term used in the current study.

Unfortunately, these extensive studies were not found to be that fruitful since none of these theories provided a standard universal set of semantic roles that could be applied to any language [9]. The popular proposed lists range from a large set of situation-specific roles such as the Killer and the Victim roles used in the semantic frames theory [18] to a small set of general roles, such as the Agent, Theme, Location, and Goal roles used in [12] to an overgeneralized set of two proto-roles like the Proto-Agent and Proto-Patient roles used in[14].These linguistic theories also fail in providing a unified definition for the main roles like the "Agent" role that they all share. Another problem is the different rolesets designed for each predicate and the different criteria used to differentiate between core and adjunct roles of a specific predicate.

While linguists were trying to settle down their arguments around the notion of semantic roles, computer scientists started to develop and build systems that could map a sentence to its underlying semantic representation, without trying to uncover thoroughly the linguistic side of the phenomenon. Their motivation was not to answer the linguistic questions of what semantic roles are or what are the linguistic principles that govern it, but rather to try to develop systems that could with some linguistic and world knowledge provide a semantic interpretation of a text [11].

Early developed SRL systems were domain-specific like the Air Traveler Information System (ATIS) which was developed to enable users to book their flights automatically [19]. The system understands the users' queries by mapping their queries constituents to a pre-arranged set of semantic roles related to the air travel semantic frame only like destination and flight date roles [20]. SRL systems had undergone a breakthrough after the advent of large corpora annotated with semantic roles like Propbank [21] and FrameNet [7], and because of the availability of other lexical resources like the VerbNet [22] which is a verb lexicon that consists of a number of classes where verbs that behave syntactically alike are assigned the same class, and because of the availability of elaborate machine learning algorithms. This mixture motivated researchers to develop domain-independent statistical SRL systems that could train over a vast amount of annotated data to extract linguistic features that realize semantic roles and then tag unseen other data. The CoNLL, the Conference on Natural Language Learning, devoted three of their yearly meetings for the SRL task [3],[4], [23].

Most of the developed SRL systems tackled English texts till Diab [24] proposed a statistical SRL system that was trained over Modern Standard Arabic (MSA) data. The release of the Arabic version of Propbank [25] made developing such a system feasible. This system was built following the methodology advocated for other SRL systems without considering the special morphological, syntactic, and semantic characteristics of the MSA. The system yielded an F1 score of 81.34%. Ref. [26] proposed another SRL system that considered the special linguistic characteristics of MSA which yielded an F1 score of 82.17% which is to be considered as an improvement in the state of art of Arabic statistical SRL. Although the majority of the SRL systems follow the statistical-based approach, this approach suffers from many drawbacks. To begin with, availability of a large amount of annotated corpora is a need to build a statistical SRL system. There is a direct correlation between the amount of annotated data and the accuracy of the system. But the major flaw of the statistical SRL system is that the developed systems extracted features over-fits the data it trains on, thus when trying to test the developed system on other genres, the system shows a significant performance deterioration.

Rule-based SRL systems are really few. The only well-known rule-based system was developed by [27]. Semantic and syntactic rules were extracted from the FrameNet annotations and then added to a knowledge-base. The system achieved 74.5% accuracy. Though the rule-based approach is not used widely to build SRL systems since extracting and structuring such rules is very exhausting, its bright side comes from its reliance on extracting economic general relevant features that would achieve the task in hand effectively, unlike statistical SRL systems where plenty of features are extracted.

3 METHODOLOGY

This section describes in detail the design and implementation of the MSA rule-based SRL system. It starts by giving a brief description of the resources and tools needed by the system. Then it describes in detail the procedures followed to design the training and testing data, construct the system knowledge bases, and build the rule-based SRL system.

I. Data Resources

This section provides a brief description of the data resources used to build the system and the reason behind choosing these resources.

1) *Arabic Penn Treebank (ATB):* The (ATB) project [28] is a million words treebank, built to support the development of natural language processing research on the written MSA and was accomplished by a team from the Linguistic Data Consortium (LDC) and the University of Pennsylvania. The ATB is a newswire corpus that is comprised of three parts, ATB1 which represents 166K words of 734 stories drawn from the Agence France Presse corpus, ATB2 which represents 168K words of 501 stories drawn from the Ummah Arabic News Text corpus and ATB3 which represents 350K words of 600 stories drawn from Annahar News Agency corpus. The current study selected its dataset from both ATB1 and ATB3, as these two parts are the only available parts for the researcher. The ATB corpus is provided with both morphological and syntactic annotations. The drawback ATB suffers from is the lack of word sense disambiguation (WSD) for the token's lemmas.

2) Arabic VerbNet (AVN): the AVN [8] is a lexical resource that provides a classification of the Arabic verbs according to Levin's verbs classification structure [10] where verbs that behave syntactically alike are assigned to the same class, as they share the same semantic meaning according to Levin's hypothesis. The current study uses the AVN to assign the verbal predicates drawn from the ATB to their corresponding AVN classes. Of course, polysemous and homonymous verbs would be found to belong to multiple AVN classes which is the second advantage AVN provides, as it shows the different senses of these verbs by showing the different AVN classes they belong to.

3) SemLink: Since the current study needs to define a roleset for each verbal predicate, the three resources namely the FrameNet, the VerbNet and the Propbank were the researcher's destination as each of them specify a roleset for each predicate. Unfortunately, these three resources follow different criteria while defining a predicate roleset which yields different rolesets for the same predicate. SemLink is an ongoing project which maps these three resources rolesets of each predicate, which is what the current study needs [29].

4) Arabic And English WordNet: WordNet [30], [31] in general is a large electronic thesaurus-like relational database that groups synonymous nouns, verbs, adjectives, and adverbs that represent a unique lexicalized concept together into sets called synonym-sets. The synonym-sets are interlinked together with a number of lexical relations such as antonymy, hyponymy, meronymy, and entailment, which results in a semantic network that can be navigated. The Arabic WordNet (AWN) database current version AWN.2 covers around 11,269 synonym-sets [32] and the English WordNet Current version EWN.3 covers around 117,597 synonym-sets [30].

J. Tools

This section provides a brief description of the tools used to build the SRL system and the reason behind choosing these tools.

5) *Python*: Python [33] is an interpreted powerful high-level, object-oriented programming language created by Guido van Rossum. It is not really a tool, but it represents the programming language used to build the system. The reasons behind choosing Python as the language to work with are, the availability of many APIs written in the Python language, and the availability of the natural language toolkit (NLTK) which is the main reason behind using Python. Python 3.6 is the version used for building the study's SRL system.

6) Extensible Markup Language (XML): XML is a markup language designed to store and transport data. Again, XML is not a tool, but it represents the language chosen to build the knowledge bases.

7) *Natural language toolkit (NLTK):* NLTK is a famous open source Python library that contains corpora and modules which supports the development of NLP applications. Using NLTK in the current study facilitated working with and traversing through the complicated syntactic tree data structure to extract the needed information and allowed the direct use of the EWN as it is available within its corpora.

8) *Google translate API:* Google Translate API [34] is a free Python library that translates texts from a source language specified by the user to a target language. It can also detect the language of an input text. The reasons behind using the Google translate API are its high speed and the reliability and consistency of its translation. A study made by [35] proved that Google Translate outputs are slightly better than the outputs of Bing, Babylon and Systranet machine translation Systems.

K. Procedure

This section provides in detail the procedure followed to, firstly choose the study's list of semantic roles, secondly select the AVN classes along with their verbal predicates that this study will focus on, thirdly design the core roleset of each AVN class, fourthly design the study's adjunct semantic roles, fifthly design and annotation of data, sixthly design the linguistic features that will be extracted to be stored in the knowledge base of each AVN class, seventhly construct the SRL system and finally design the testing data and how the performance of the system was calculated.

1) Choosing the semantic roles list: Proposing the study's semantic roles list was a troublesome challenge since a standard agreed-upon list of semantic roles does not exist. Many semantic roles lists were proposed by many theories, corpora, and projects, each designed based on different criteria. The current study adopted a different number of criteria that defines the characteristics of a proper list of semantic roles. These criteria are adopted from [36], [37] The criteria state that:

I. The set of semantic roles should be small in size [36].
II. "Semantic roles are neither syntactic nor lexical structures but are semantic categories" [37].
III. Semantic roles list must not contain roles defined for specific verb or classes of verbs" [37].
IV. Semantic roles list must be universal, it could describe the semantic relations within any language.

Regarding the adopted criteria, the study mainly adopted its list of semantic roles from the LIRCIS Project [37] since most of the adopted criteria are drawn from that project besides the project's proposed list was tested for completeness on many languages and yielded successful results. Some other roles were drawn from The UNL project set [38] since the UNL set was meant to be a universal set to suit any language. These roles are not found in the LIRICS project but were found in Arabic. Other roles were drawn from the Propbank set of semantic roles [39] since the Propbank set could be seen universal as it was used to annotate corpora from eight languages one of them is the Arabic language. The adopted semantic roles list consists of 32 roles.

2) Assigning the PATB verb lemmas to their AVN class: The second step considered assigning the PATB verb lemmas to their AVN classes. The verb lemmas extracted comprises 2,355 verb lemmas. The AVN is comprised of 336 verb classes populated with around 7,774 verb entries [8]. The output of this step was 366 files, with each file representing one of the AVN classes, populated with the AVN class verb lemmas along with their frequencies in the ATB.

One major problem encountered while extracting the frequencies of the verb lemmas was the need to disambiguate the polysemous and homonymous verbs senses like "أصاب" (aṣāba) polysemous verb which is assigned to the two AVN classes "jaraḥ" class and "ishtará" and has a frequency of 116 sentences that needs to be disambiguated according to these two meanings to correctly assign the frequency of each sense. A layer of verb sense disambiguation was required to tackle this problem, unfortunately, this layer was not available, thus the researcher carried on the disambiguation phase manually.

3) Designing the core roleset of the selected AVN classes: Designing the core roleset of the selected AVN classes considers the part where the number and types of the AVN classes core semantic roles are specified. The three popular resources namely the FrameNet, the VerbNet and the Propbank designed their core semantic roles of each class or frame depending on different criteria, which yielded different rolesets for the same class across the three resources. Thus, the study needed to define its own criteria upon which the rolesets of each class could be selected. The basic criterion the study adopted states that the obligatory participants (core roles) are implied in the semantics of each verb whether or not they are syntactically instantiated [40]. When these semantically obligatory arguments are not explicitly syntactically expressed, it is termed implicit arguments[40]. FrameNet design of the core frame elements depends on many criteria, one of them is the previous criterion, which is stated as follows "A frame element which, when omitted, receives a definite interpretation, is also core" [7]. For that reason, the current study adopted the FrameNet design of the core roleset which yielded a need to map the study's AVN classes to their corresponding FrameNet frames. Mapping each AVN to their corresponding FrameNet frame was accomplished using the SemLink resource. After this mapping step, another step was needed which concerns mapping the situation-specific frame elements of the chosen frame to their corresponding semantic role tags from the studies proposed list.

4) Designing the study's adjunct roleset: The current study adopted the Propbank criteria, where the same set of adjuncts is used across the chosen classes unless one of the adjuncts is specified as a core role in a class, it is then transferred from the adjunct set to the core set as the case with "ḥaṣala" AVN class where both the "Time" and the "Location" roles are specified as core roles of that class where these two roles are in most case adjunctive roles.

5) *Annotating data with semantic roles:* the researcher manually annotated the trees of each AVN class with their semantic roles, to be used afterwards to extract the relevant linguistic features with which each semantic role is realized, then properly structure and organize these features in the class knowledge base to be used by the system eventually.

6) Design of the Linguistic features to be extracted: One of the current study's goals is to design an economic number of linguistic features that MSA uses to realize the semantic roles of any predicate.

The study designed a number of features related to the predicate which were found to play a significant role in the assignment of the predicate semantic roles to its arguments, these features are:

- Voice: the voice feature was extracted directly from the label of the predicate in the syntactic tree, which has the value of either active or passive voice.
- The syntactic frame: the syntactic frame represents the sister nodes of the verbal predicate in the syntactic tree and the sisters of the verb phrase (VP) that contains that predicate. This feature collects all the predicate arguments that represent the candidates that would realize the predicate semantic roles.
- The predicate sense: the predicate sense is inferred from the AVN class name the predicate belongs to.

- The roleset of the predicate: both the core and adjunct roles of the predicate are retrieved from the XML lexicon that is built for each class, which will be discussed later.

Another list of features was extracted, which is related to each argument which are:

- Phrase type: phrase type feature represents the syntactic category (NP, PP, etc.) of the argument. It is drawn from the label of the argument in the syntactic tree.
- Function tag: This feature indicates the grammatical function of the NPs whether it is subject or object, and the other function tags the ATB provides are discarded as the TMP, LOC. etc.
- Status: the status feature shows if the argument is dropped, topicalized, sister to the predicate or sister to the VP.
- Distance: the distance feature indicates how far the argument is from its predicate. This feature is calculated only for the arguments that are sisters to the predicate.
- Headword: The headword feature was firstly introduced in [20] and It is used to show the selectional restrictions of the lexical items that might fill a semantic role. The headword extraction process follows the headword rules defined by Collins [41].
- The headword lemma: the lemma of the headword is extracted.
- The POS tag of the headword: the POS tag of the headword is extracted from the label of the headword in the syntactic tree. The reason behind extracting the head lemma POS is the fact that it replaces to some extent the name entity feature, as proper nouns lemmas which map mostly to a person, organization, or location name entities.
- Semantic features of the headword: this feature considers extracting the semantic features of the headword lemma. To achieve this task, the same procedure followed by [20] of getting semantic features by traversing the WordNet hypernyms hierarchy of the headword lemma was followed. The list of hypernyms of that lemma is considered as the semantic features of the headword lemma. AWN version 2 was used. AWN was found to be suffering from poor coverage since many of the searched lemmas weren't found which affected the results of the system dramatically. To overcome this problem, lemmas that are not found in the AWN were translated using the Google translator API and then their hypernyms were drawn from the EWN version 3 which is available among the NLTK corpora.
- Content word: this feature was introduced by [42], where it selects a content word of the argument. Content word feature was extracted for PP arguments only where the right most child is selected. Content word feature was introduced to give better information when the headword feature is less informative, like the case with the PP arguments.
- Lemma of the content word.
- POS tag of the content word.
- Semantic features of the content word.

7) Knowledge base construction: The knowledge base construction phase considers the part where the extracted features are properly structured and stored in a knowledge-base to be used later by the system. For each AVN class, a knowledge base was created with the XML language. Each AVN class knowledge base contains the AVN class name, its mapped FrameNet frame, and the Arabic Propbank (APB) frame-file name of the frequent verb. The knowledge base is then divided into two parts, the core semantic roles part and the adjunct semantic roles part. For each role either core or adjunct, the role tag name is provided according to the study's proposed list of semantic roles, along with its APB correspondent tag. The features of each role are then split according to the predicate voice since it was found that the features of the same role change according to the predicate voice.

For each role and within each voice, the different phrase types that were found to realize that role in the annotated data were extracted and stored in the knowledge base along with their frequencies. Each phrase type is then constrained with other features which are the function tags that cooccur with the phrase type if are, the different statuses of the same phrase type and the different number of distances the same phrase type may occur away from its predicate. The semantic selectional restrictions of the headword lemmas and the content word lemmas of that phrase type is also stored but according to an order of priorities which goes as follows: if the list of the semantic features of the headword lemma is retrieved successfully, then the most proper semantic feature that can be seen as the selectional restriction that made the headword fit for the role it realizes is extracted and stored, on the other hand if the system fails to retrieve such list, the POS tag of the headword is stored but only if the POS tag again plays the role of the selectional restriction, otherwise the headword lemma is stored. The content words of the phrase type are stored in the lexicon following the same procedure of storing the headword. Each phrase type of each role has a minimum score attribute that represents the minimum number of features an argument needs to be marked with, in order to be classified with that role. Also, each phrase type has a score attribute that represents the ideal number of features an argument needs to be marked with, in order to be classified with that role.

8) Building the Rule-based SRL system: Building a rule-based semantic role labeling system for MSA is the ultimate goal of the current study. The input data that shall be fed to the system must be a fully syntactic parsed sentence according to the ATB parsing convention. The strategy of choosing to work with syntactic hand-annotated data was justified by the researcher need to avoid the system performance errors that may result from feeding erratic input data to it.

The first step the system performs considers detecting the verbal predicate in the syntactic tree and then extracting its/their lemma(s). The lemma extracted was then searched and if it was found to be one of the 372 verbs the current study focuses on, the predicate features explained earlier were extracted from the syntactic tree.

The second step, then extracts from the fed syntactic tree the syntactic nodes that are highly probable to be assigned a semantic role. The current study followed the same hypothesis of [43] where it was claimed that the candidates that highly represent the predicate semantic roles are located in the predicate immediate clause which represents the clause that contains the predicate. Thus, the study focused on retrieving the syntactic arguments that are either sisters to the predicate or sisters to the VP that contains the predicate. Retrieving the candidate arguments of the extracted predicate was not a challenge-free task, two major challenges were encountered. The first challenge encountered pertains to retrieving empty arguments, and the second challenge pertains to retrieving arguments that have an explicit pronoun “dhamīr ẓāhir” headword. These two problems needed to be solved with anaphora resolution layer the system lacked, thus, the problem was left unhandled.

The last step is the critical step in the whole system, as it considers the part where the system predicts and assigns the correct roles to the retrieved arguments. This step starts with retrieving the AVN class knowledge-base, the verb belongs to. The system firstly extracts the core and adjunct roles features from the knowledge base according to the predicate voice and then calculates the probability of each retrieved syntactic argument to be assigned any of the class core or adjunct roles. This calculation is done firstly by comparing the phrase type of each syntactic argument to the phrase types of each role (PTs), and if the argument PT was found among any role PTs the probability score of that argument to be assigned that role becomes one, and consequently the other argument’s extracted features get compared to that role PT constriction rules which includes:

- Searching the argument headword lemma among that role headword lemmas, if found the argument probability score is incremented by one, if not found, the argument headword lemma’s semantic features are searched among that role selectional semantic features and if one of the argument features was found within the role features, the argument probability score gets incremented by one, if not found the headword lemma POS tag is searched among that role POS tags and if found the probability score gets incremented by one, otherwise the probability score is left unchanged.
- Searching the argument content word if it is different from the headword follows the same strategy explained above.
- Searching the argument function tag FT among that role FTs, if found the probability score gets incremented by one, otherwise the probability score is left unchanged.
- Searching the argument status among that role statues, if found the probability score gets incremented by one, otherwise the probability score is left unchanged.
- Searching the argument distance among that role distances, if found the probability score is incremented by one, otherwise the probability score is left unchanged.

After the calculation of the argument probability score for each role, the argument is added as a potential realization of any role if its final score equals or exceeds the minimum score specified for that role.

After getting the potential arguments of each role, a filtration process takes place where each role keeps only the argument with the highest score and the other potential arguments of that role are erased even if they have the minimum role score.

After the filtration process is done, another filtration process takes place where the scores of the arguments that are assigned multiple roles are compared to assign that argument the role at which it gets the highest score.

In some cases, it was found that the same argument may get the optimum score of two different roles. At this condition, the system assigns that argument the more frequent role of the two role candidates.

In some other cases, some arguments are left without being assigned any semantic role as their features suffice none of the roles specified in the knowledge base, at this time the argument gets assigned with a "None" role. The system specifies for each sentence its syntactically unexpressed obligatory semantic roles. The full system flow diagram is shown in figure (5).

9) *Testing phase: T*he testing phase considers first the strategy followed to design the testing data, and second, the measurement calculated to evaluate the SRL built system.

Designing the testing data followed two different paths which resulted in two types of testing data. The first type represents unseen testing sentences drawn from the same frequent verb, to find out the system accuracy when implemented over unseen data which is called testing data type 1. The second type represents the testing sentences drawn from the other verbs that belong to the same class to find out if the features drawn from the annotated data of the most frequent verb would apply correctly to the other class predicates which is called testing data type 2. The researcher claims that the system would perform accurately when assigning core roles of any AVN class with both types of testing data, and it would perform moderately when assigning adjunct roles. Figure (6), shows the system output for one of the testing sentences.

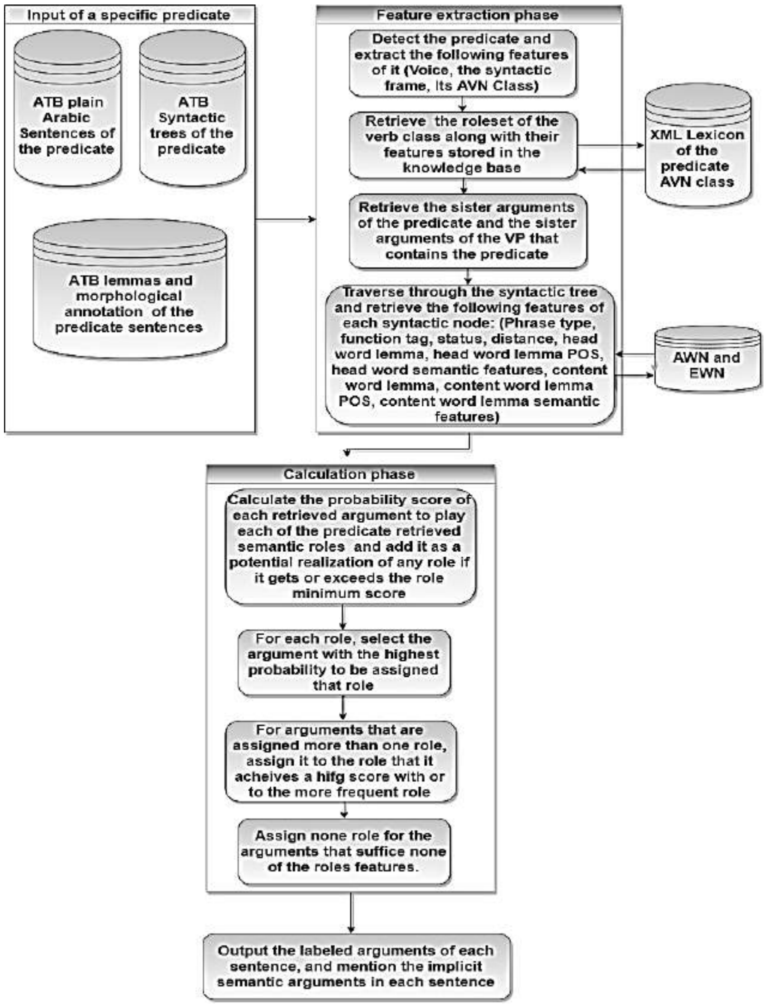


Figure 5: The full system flow diagram

وقد قتل اكثر من 90 شخصا منذ بداية تشرين الاول/اكتوبر في مجازر واعتداءات نسبت الى الجماعات الاسلامية المسلحة المعادية لسياسة الوفاق الوطني التي ينتهجها الرئيس عبد العزيز بوتفليقة وفق حصيلة اعدت استنادا الى الصحف .

- Predicate Voice: Passive
- Type: Core
  Patient: اكثر من 90 شخصا
  Optimal score: 5 arg_score: 5
- Type: Adjunct
  Time: منذ بداية تشرين الاول/اكتوبر
  Optimal score: 4 arg_score: 4
- Type: Adjunct
  Setting: في مجازر واعتداءات نسبت الى الجماعات الاسلامية المسلحة المعادية لسياسة الوفاق الوطني التي ينتهجها الرئيس عبد العزيز بوتفليقة
  Optimal score: 4 arg_score: 4
- Type: Adjunct
  Adverbial: وفق حصيلة اعدت استنادا الى الصحف.
  Optimalscore:4 arg_score:4
- **Not_instintiated: Agent, Cause, Instrument, Means

Figure 6: The system output for an input sentence.

The system is evaluated with respect to precision, recall and the F1 (F-measure) of the predicted arguments. To calculate these measurements, the study must first define when an argument is correctly labeled. The study adopted the definition of a correct assigned argument from [44] which states, for an argument to be recognized as labeled correctly, the words spanning the argument as well as its semantic role tag must be correct. When one of these two criteria drops the assigned argument is considered wrong. The second two criteria are adopted from [45]. These two criteria state that:

1. "If an argument is a reference (R-arg) to some other argument arg, then this referenced argument must exist in the sentence.
2. If there is a C-arg argument, then there has to be an arg argument; in addition, the C-arg argument must occur after arg." [45].

The current system allows the duplicate roles since it was found that Arabic allows such repetition of roles, especially with the adjunct roles.

The measurements calculated are the Precision, recall and F1 each calculated for the core roles, adjunct roles, and all roles. Precision is defined as "the proportion of arguments predicted by a system which are correct" [3]. It is calculated with the following formula: Precision = roles correctly assigned / roles assigned relevant. Recall is defined as "the proportion of correct arguments which are predicted by a system [3]. It is calculated with the following formula: Recall = roles correctly assigned / total of roles. Finally, the F1 measure "computes the harmonic mean of precision and recall and is the final measure to compare the performance of systems" [3], It is calculated with the following formula: F1 = (2 * Precision * Recall) / (Precision + Recall).

4 RESULTS

This section shows the results of the built rule-based SRL system over the manually annotated data and testing data. Table (I) shows the number of sentences of the training and testing data.

Table I

THE NUMBER OF SENTENCES OF THE TRAINING AND TESTING DATA.

Manually annotated data sentences of all classes	Testing data of type 1 sentences	Testing data of type 2 Sentences	Total of Testing data sentences	Total
2786	1439	821	2260	5046

The manually annotated data represents 55.2 % of the whole sentences fed to the system, the other 44.8% of sentences represent the testing data fed to the system which is comprised of 63.7% sentences that represent testing data type 1and 36.3 % of the sentences represent testing data type 2. Running the system over the manually annotated data was done firstly to find out if the linguistic features extracted from the manually annotated data was structured properly and secondly to find out if the calculation procedure adopted by the system to predict the propositions' semantic roles is accurate. Table (II) depicts the precision (P), recall (R) and F1 of the system results over the manually annotated data, testing data of type 1 and type 2.

TABLE II

PRECISION (P), RECALL (R) AND F1 OF THE SYSTEM RESULTS OVER THE TRAINING DATA, TESTING DATA OF THE SAME FREQUENT VERB AND OF THE OTHER AVN CLASS VERBS TESTING DATA.

Data	Core P	Core R	Core F_1	Adjunct P	Adjunct R	Adjunct F_1	All roles P	All roles R	All roles F_1
Manually annotated data	97.6	96.4	96.9	94.5	92.1	93.3	96	95.2	95.6
Testing data type 1	97.1	96.7	96.9	91.1	82	86.3	93.1	92.8	92.9
Testing data type 2	94.5	90.3	92.4	82	69	75	87	85.1	86

Figure (7) shows the overall system performance over the three datasets according to the core, adjuncts, and all roles.

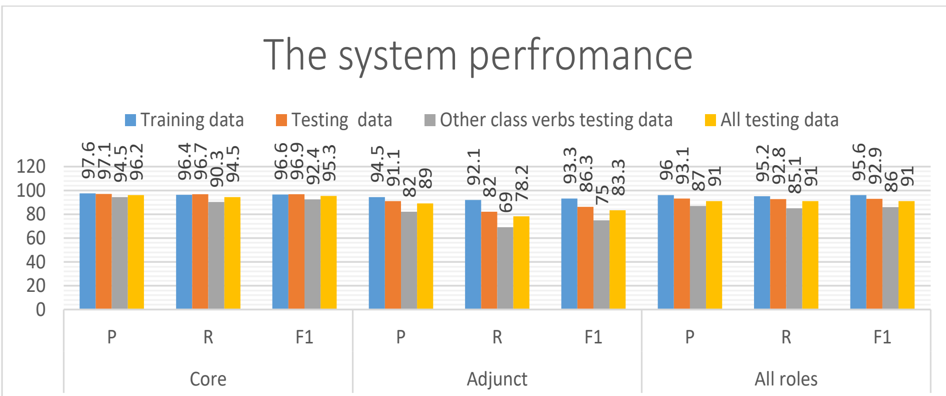


Figure 6: the overall system performance over the three datasets according to the core, adjunct and all roles.

5 DISCUSSION

The results achieved by the system shall be discussed in this section to give a linguistic insight of the results that goes beyond just the numbers.

L. Core results Vs. adjunct results.

The system achieved nearly consistent F1 scores while predicting the core roles across the manually annotated data, the first type and the second type of testing data as the core F1 scores of the three data sets are 96.4%, 96.7%, and 90.3% respectively. On the other hand, the system showed a significant drop while predicting the adjunct roles across the three sets of data as the adjunct F1 scores of the three data sets are 93.3%, 86.3%, and 75% respectively. The core roles which are described in [14] as more important than the adjunct ones in human life since who did what to whom is more important than where or when it happened. For that reason, core roles show less variable behavior when realized to be easily understood. On the other hand, the adjunct roles, gets its high variability not only because of its less importance in a sentence if compared to the core role, but also because they occur with most of the predicates which suggest that the better procedure that needs to be followed while working with adjunct roles is to extract their feature cumulatively form the different AVN classes and the final lexicon of the whole adjunct roles shall be used by all the AVN classes.

M. Wrong trees

The researcher found out that for achieving high accuracy with syntactic parsing, one needs to have a sufficient knowledge of the predicates semantic roles especially the core ones since the reasons behind many of the erratic trees found in the ATB are explained by the lack of availability of such knowledge for the annotators, which resulted in many confusions while parsing the trees.

N. Metaphoric roles

Some of the encountered arguments were metaphorically expressed. For example, the “aṣ-ṣaḥīfah” argument was used numerously to express the “Agent” of the “qāla” predicate. Using that argument refers actually to the journalist who wrote or said a message via that journal. The dilemma here was how should the researcher annotate the semantic role of such argument, based on its literal meaning or on the meaning it stands for. The researcher chose not to work on the argument literal meaning and to assign the semantic role of these constituents based on the intended meaning.

O. The completeness of the study’s proposed list of roles

The currents study proposed a set of semantic roles for Arabic predicates with the goal that this list covers all the semantic relations the Arabic predicates assign. The defined list proved its success since it was easy to map the FrameNet specific-frame elements to their correspondent general semantic roles the study proposed. One problem was only encountered while mapping the “Topic” FrameNet role found in most communication-related frames, which represents the topic about which a speaker is talking. Since the utterance communicated by the speaker was mapped to the “Theme1” semantic role, the “Topic” FrameNet role was assigned to “Theme2” semantic role, as each Theme represents a different semantic relation.

P. Pipeline nature of SRL system

The proposed SRL system follows a pipeline of tasks that need to be accomplished correctly first for the system to yield a good performance. The main prerequisites any SRL system is built upon is the availability of a rich morphological and syntactic analysis of the input text, the availability of word sense and anaphora coreference assignment and the availability of different types of lexicons that maps the Arabic predicates to their rolesets, provided the ontological categories of the Arabic lemmas with high coverage. The pipeline nature of the developed system is a major drawback of the current system since any missing layer of these layers would turn the SRL system work incomplete and the erratic output of any of these layers would affect the performance of the currents system.

6 CONCLUSION AND FUTURE WORK

Rule-based SRL systems are not frequent in NLU domain. The Current study is considered as an attempt to build a rule-based semantic role labeler for Modern Standard Arabic (MSA). The uses the general roles list of Agent, Theme unlike most of the Arabic SRL systems which only the Propbank tags. The developed system achieved a final F1 score of a 91.0% F1 score on the testing data.

The researcher looks up to build a robust system that accepts an input syntactic trees parsed automatically, rather than working on gold-standard syntactic trees. The researcher also aims at covering the 337 AVN classes verbs in the new system. Not only the verbal predicates shall be the point of research, but also other kinds of predicates like the nominal, and adjectival predicates. The researcher needs to build a system with higher accuracy, and this system could be integrated with other systems.

II. REFERENCES

[1] R. J. Mooney, "Semantic Parsing: Past, Present, and Future," Austin: University of Texas, 2018.
[2] L. Banarescu, C. Bonial, S. Cai, M. Georgescu, K. Griffitt, U. Hermjakob and N. Schneider, "Abstract Meaning Representation for Sembanking, "in *Proc. of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pp.178-186, Sofia, Bulgaria, 2013.
[3] X. Carreras and L. M`arquez, "Introduction to the CoNLL-2005 Shared Task: Semantic Role Labeling," in *Proc. of the Ninth Conference on Computational Natural Language Learning*, pp.152-164, Ann Arbor, Michigan, 2005.
[4] L. M`arquez, X. Carreras, K. C. Litkowski and S. Stevenson, "Semantic Role Labeling. An Introduction to a special issue," *Journal of Computational Linguistics*, vol.34, no.2, pp.145-159, 2008.
[5] *ALA-LC romanization tables: transliteration schemes for non-Roman scripts*. Washington: Cataloging Distribution Service, Library of Congress, 1997
[6] J. Hurford, B. Heasley and M. Smith, *Semantics: A Coursebook*, Cambridge University Press, 2007.
[7] J. Ruppenhofer, M. Ellsworth,M. R. Petruck, C. R. Johnson and J. Scheffczyk, "FrameNet II: Extended Theory and Practice," Berkeley, California: International Computer Science Institute, 2016
[8] J. Mousser, "A Large Coverage Verb Lexicon for Arabic," Doctoral dissertation, University of Konstanz, 2013.
[9] M. Palmer, D. Gildea and N. Xue, *Semantic Role Labeling*, Morgan and Claypool Publishers, 2010.
[10] B. Levin, *English verb classes and alternations : a preliminary investigation*, University of Chicago, 1993.
[11] S. S. Pradhan, "Robust semantic role labeling," Doctoral dissertation, University of Colorado, 2006.
[12] C. J. Fillmore, "The case for case," *In collection. Bach, & H. R., Universals in Linguistic Theory*, pp. 1-88, 1968.
[13] L. Haegeman, *Introduction to government and binding theory*, 2nd ed, Oxford : Blackwell, 1999.
[14] D. Dowty, "Thematic Proto-Roles and Argument Selection," *journal of Language*, Vol. 67, No. 3, pp.547-619, 1991.
[15] J. Gruber, "Studies in Lexical Relations," Doctoral dissertation, Massachusetts Institute of Technology, 1965.
[16] R. Jackendoff, "The Status of Thematic Relations in Linguistic Theory," *journal of Linguistic Inquiry*, vol. 18, No. 3, pp. 369-411. 1987
[17] R.Valin, "Semantic macroroles in role and reference grammar," 2004.
[18] C. J. Fillmore and C. Baker, "A Frames Approach to Semantic Analysis," *In B. Heine, and H. Narrog, The Oxford Handbook of Linguistic Analysis*, 2012.
[19] S. Miller, "A Fully Statistical Approach to Natural Language Interfaces," in *Proc. of the 34th Annual Meeting on Association for Computational Linguistics*, pp.55-61, 1996.
[20] D. Gildea and D. Jurafsky, "Automatic Labeling of Semantic Roles," *journal of Computational Linguistics*, vol.28, no.3, pp. 245–288, 2002.
[21] M. Palmer, D. Gildea and P. Kingsbury, "The Proposition Bank: An Annotated Corpus of Semantic Roles," *journal of Computational Linguistics*, vol.31, no.1, pp. 71-106, 2005.
[22] K. K. Schuler, "Verbnet: A Broad-coverage, Comprehensive Verb Lexicon," Doctoral dissertation, University of Pennsylvania, 2005.
[23] M. Surdeanu, R. Johansson, A. Meyers, L. M`arquez and J. Nivre, "The CoNLL-2008 shared task on joint parsing of syntactic and semantic dependencies," in P*roc. of the Twelfth Conference on Computational Natural Language Learning*, pp.159-17, Manchester, England, 2008.
[24] M. Diab, A. Moschitti, and D. Pighin, CUNIT: "A Semantic Role Labeling System for Modern Standard Arabic," *in Proc. of the Fourth International Workshop on Semantic Evaluations*, (SemEval-2007), pp.133-136, Prague, Czech Republic, 2007.
[25] M. Palmer, O. BabkoMalaya, A. Bies, M. Diab, M. Maamouri, A. Mansouri, and W. Zaghouani, "A Pilot Arabic Propbank," *in Proc.of the Sixth International Conference on Language Resources and Evaluation*, pp.28-30, Marrakech, Morocco, 2008.
[26] M. Diab, A. Moschitti and D. Pighin, "Semantic Role Labeling Systems for Arabic using Kernel Methods," *in Proc. of ACL-08: HLT*, pp. 798-806, Columbus, Ohio, 2008.
[27] L. Shi and R. Mihalcea, "An Algorithm for Open Text Semantic Parsing," *in Proc. of the 3rd Workshop on Robust Methods in Analysis of Natural Language Data*, pp. 59-67, Stroudsburg, PA, USA, 2004.
[28] M. Maamouri, A. Bies, T. Buckwalter and W. Mekki," The penn Arabic treebank: Building a large-scale annotated Arabic corpus," *NEMLAR Conference on Arabic Language Resources and Tools*, 2004.

[29] M. Plamer, "SemLink-Linking PropBank, VerbNet, FrameNet," in *Proc. of the Generative Lexicon ConferenceGenLex-09*, Pisa, Italy, 2009.
[30] Princeton University (2010), About WordNet,Available from: https://wordnet.princeton.edu/, (accessed 28 August 2020).
[31] G. A. Miller, "WordNet: A Lexical Database for English," *journal of Commun. ACM*, vol. 38, no.11, pp.39-41, 1995.
[32] Y. Regragui, L. Abouenour, F. Krieche, K. Bouzoubaa and P. Rosso, "Applications, Arabic WordNet: New Content and New applications," in *Proc. of Global Wordnet Conference*, Bucharest, Romania, 2016.
[33] Learn Python Programming Web Site, https://www.programiz.com/python-programming, (accessed 28 August 2020).
[34] Googletrans Web Site, https://pypi.org/project/googletrans/,(accessed 28 August 2020).
[35] H. Al-Sarhan, A. Darabsh, I. Al-Husban and R. AlShalabi, "Evaluating Machine Translations from Arabic Into English and Vice Versa,"*in Proc. of ICETMS*, vol.8, Dubai, UAE, 2017.
[36] W. Croft, *Syntactic Categories and Grammatical Relations: The cognitive organization of information*, University of Chicago Press, 1991
[37] A. Schiffrin and H. Bunt(2007), "LIRICS Deliverable.D4.3. Documented compilation of semantic data categories". Available from http://lirics.loria.fr, (accessed 28 August 2020).
[38] R.Martins, "Le Petit Prince in UNL," in *Proc. of the Eighth International Conference on Language Resources and Evaluation,* (LREC'12), pp. 3201–3204, Istanbul, Turkey, 2012.
[39] A. Mansouri, M. Foster, J. D. Hwang, O. Babko-Malaya and M. Palmer, *ARABIC PropBank ANNOTATION GUIDELINES*, 2013.
[40] R. Jackendoff and P.W. Culicover, *Simpler Syntax*, Oxford University Press, 2005.
[41] M.J. Collins, "Head-driven Statistical Models for Natural Language Parsing", Doctoral Dissertation , University of Pennsylvania,1999
[42] M. Surdeanu, S. M. Harabagiu, J. Williams, and P. Aarseth, "Using Predicate-Argument Structures for Information Extraction," *in Proc. of the 41st Annual Meeting of the Association for Computational Linguistics,* vol.1, pp. 8-15, USA, 2003.
[43] J. Lim, Y. HwangY, S. Park and & H. Rim, "Semantic Role Labeling using Maximum Entropy Model," *in Proc. of the Eighth Conference on Computational Natural Language Learning (CoNLL-2004) at HLT-NAACL Boston*, pp. 122–125, Massachusetts, USA, 2004.
[44] X. Carreras, Màrquez, and Lluís, "Introduction to the CoNLL-2004 Shared Task: Semantic Role Labeling," *in Proc. of the Eighth Conference on Computational Natural Language Learning*, (CoNLL), pp. 89–97, Boston, Massachusetts, USA, 2004.
[45] V. Punyakanok, D. Roth, W. Yih and D. Zimak, "Semantic Role Labeling via Integer Linear Programming Inference," *in Proc. of the 20th International Conference on Computational Linguistics*, pp. 1346–es, Geneva, Switzerland, 2004.

III. BIOGRAPHY

Dr. Sameh Alansary: *Director of Arabic Computational Linguistic Center at Bibliotheca Alexandrina, Alexandria, Egypt.*

He is professor of computational linguistics in the Department of Phonetics and Linguistics and the head of Phonetics and Linguistics Department, Faculty of Arts, Alexandria University. He obtained his MA in Building Arabic Lexical Databases in 1996, and his PhD from Nijmegen University, the Netherlands in building a formal grammar for parsing Arabic structures in 2002. His main areas of interest are concerned with corpus work, morphological analysis and generation, and building formal grammars. He is also the head of Arabic Computational Linguistics Center in Bibliotheca Alexandrina. He is supervising and managing the Universal Networking Language project in Library of Alexandria since 1-6-2005 till now. Dr. Alansary is the co-founder of the Arabic Language Technology Center (ALTEC), an NGO aims at providing Arabic Language resources and building a road map for Arabic Language Technology in Egypt and in the Middle East. He has many scientific works in Arabic

Natural Language Processing published in international conferences and periodicals, and a member in many scientific organizations: (1) Egyptian Society of Language Engineering, Cairo, (2) Arabic Linguistic Society - USA, (3) Association of Computational Linguistics - USA – Europe, (4) Universal Networking Language foundation, United Nations, Geneva, Switzerland.

Amna Helmy: *Assistant teacher at the Phonetics and Linguistics department, Faculty of Arts, Alexandria University*

She got a Bachelor of Arts in 2013 (Phonetics and Linguistics department – Faculty of Arts - Alexandria university – Alexandria - Egypt).She has graduated from Information technology institute (ITI) in 2014 intake 34.She worked as a demonstrator at Phonetics and Linguistics department , Faculty of Arts ,Alexandria university form 2013 till 2019 .She has discussed her Master thesis in 2019 about semantic role labeling. She had participated in three published papers.

نحو بناء نظام لتعيين الأدوار الدلالية للغة العربية المعاصرة: اتجاه مبني على قواعد لغوية

Amena Helmy Abd El-khaleq Muhammad Dheif[1], Sameh Saad Abu-Elmagd Alansary[2]

Department of Phonetics and Linguistics, Faculty of Arts, Alexandria University.

amena.helmy@alexu.edu.eg

Sameh.Alansary@bibalex.org

ملخص— تعيين الأدوار الدلالية هو التحليل الأوتوماتيكي للنص المدخل إلي عدد العناصر المكونة له، حيث يتكون الاقتراح من المسند ومجموعة العناصر الخاصة به، ثم يربط متعلقات الفعل بعلاقاتها الدلالية؛ أو بعبارة أخرى فإنه يسعى إلى الإجابة بشكل صحيح على سؤال من فعل ماذا لمن ولماذا ومتى وأين ولماذا.

يتمثل الهدف الرئيسي للدراسة الحالية في بناء معيّن للأدوار الدلالية يقوم على قواعد الفصحى العربية المعاصرة، وهو أول نظام لتعيين الأدوار الدلالية معتمد على قواعد لغوية يتم تطويره على الفصحى العربية المعاصرة، ويشكل الأربعون فعل الأكثر شيوعا من مجموعة شبكة الأفعال العربية (AVN) اهتمام الدراسة حاليا. وتستخدم الدراسة قائمة أدوار دلالية عامة مثل المنفذ والضحية، وتعد الأولى من حيث الاستخدام في تصنيف الأدوار العربية حيث أن الأدوار الوحيدة المستخدمة في الأنظمة السابقة هي تلك المستخدمة في مدونة PropBank. وقد حقق النظام المطور نسبة نجاح على بيانات الاختبار بمتوسط 91.0٪.

الكلمات الرئيسية في ورقة البحث : *تعيين الأدوار الدلالية، شبكة الأفعال، المسند، الأدوار الدلالية الرئيسية، الأدوار الدلالية الثانوية.*

دور القواعد اللغوية في التمييز الآلي بين معاني الحروف

خالد مصطفى أبو شبانة*[1]، أحمد عبد الغني**[2]

*1 قسم اللغة العربية، كلية الآداب، جامعة الإسكندرية، مصر

**2 قسم الصوتيات، كلية الآداب، جامعة الإسكندرية، مصر

*1kh33shabana@gmail.com

**2hmd.abdelghany@gmail.com

***ملخص*__ تتناول هذه الورقة الدور الذي تلعبه القواعد اللغوية للغة العربية في بناء التطبيقات الحاسوبية التي تقوم بمعالجة اللغة الطبيعية (NLP)، مثل تطبيقات الترجمة الآلية والتلخيص الآلي للنصوص، والتصحيح النحوي والإملائي، وإحصاء المفردات، فمن المشكلات التي تواجه تلك التطبيقات وهي تخص جانب الدلالة مشكلة المعاني المختلفة التي يحتملها اللفظ الواحد والتي يصعب على الحاسب التمييز بينها، والبحث يركز على ما يخص حروف المعاني العربية (Particles) وذلك لما تلعبه من دور محوري في الربط بين الجمل، فهي في اللغة بمثابة المفاصل في الجسد، وأيضًا لتعدد المعاني التي تكون لكثير من هذه الحروف مع لزومها هيئة واحدة لأنها مبنية دائمًا، وذلك يزيد الأمر تعقيدًا حيث لا يوجد شيء في هيئتها يتغير بتغير معناها، كما أن الحرف نفسه يستخدم مع الأفعال بمعانٍ مختلفة، فالأمر إذًا يحتاج إلى تحليل أعمق من الشكل، والبحث يحاول رصد القواعد اللغوية والتي قد تكون تركيبية أو سياقية والتي من الممكن أن يتم حوسبتها بحيث تمكن التطبيق من تحديد معنى الحرف في الجملة التي يقوم بمعالجتها، فقد قام الباحث بضبط بعض القواعد اللغوية والتي تساعد على تحديد معاني الحرف في السياق الوارد فيه وقام بصياغة هذه القواعد صياغة حاسوبية، ثم قام بتجربتها عمليًا من خلال تطبيق حاسوبي، يعد بمثابة أداة (SDK) يمكن دمجها في تطبيقات أخرى، وقد استمد الباحث النصوص التي تم تطبيق القواعد عليها من عدة مصادر، أولها: القرآن الكريم، والثاني: إحدى المدونات اللغوية وهي المدونة اللغوية العربية الدولية التابعة لمكتبة الإسكندرية (ICA)، والثالث: المقالات المنشورة على مواقع الإنترنت.**

***الكلمات المفتاحية:* القواعد اللغوية – اللبس الدلالي - حروف المعاني – المعالجة الآلية – المدونات اللغوية – التطبيقات الحاسوبية**

1- مقدمة

المقصود بالحروف هنا حروف المعاني، وهي أحد أقسام الكلام، وسميت بذلك للتمييز بينها وبين حروف المباني وهي حروف الهجاء التي يبنى بها الكلم، وأما حروف المعاني فهي الحروف التي تدل على معانٍ، وتؤدي وظائف في الجملة، والحرف هو أحد أقسام الكلام، وقد عرفه النحاة القدماء بأنه "ما دل على معنى في غيره"[1]، أي أن معناه لا يتضح بنفسه ولكنه يحتاج إلى غيره، وللحروف دور كبير في اللغة وتأليف الكلام، فهي تقوم بوظيفة الربط بين أجزاء الجمل بل وبين الجملة وأختها حتى يتماسك النص وتلتحم أجزاؤه، كما أن وظيفة التعليق التي يقوم بها الحرف تعد من الوظائف التي تبنى عليها كثير من الجمل العربية، والتعليق بالأداة أشهر أنواع التعليق في اللغة العربية الفصحى, فإذا استثنينا جملتي الإثبات والأمر بالصيغة "قام زيد، وزيد قام، وقم", وكذلك بعض جمل الإفصاح, فإننا سنجد كل جملة في اللغة الفصحى على الإطلاق تتكل في تلخيص العلاقة بين أجزائها على الأداة[5][7]، وإننا في كثير من الأحيان لا نستطيع فهم النص ولا معرفة المراد منه من غير تحديد وظيفة الحرف ومعرفة معناه بدقة، من أجل ذلك وقع اختيار الباحث على هذا الموضوع وهو ضبط القواعد التي تعين على فك لبس معاني الحروف من أجل المعالجة الآلية للنصوص العربية.

2- مفهوم الحرف

اختلف مفهوم الحرف عند اللغويين القدماء والمحدثين، وذلك يرجع إلى اختلافهم في أقسام الكلام (POS)، فأقسام الكلام عند القدماء ثلاثة: اسم وفعل وحرف[8]، فالحرف هو القسم الثالث عندهم وهو ما دل على معنى في غيره،

---

[5] اللغة العربية معناها ومبناها، تمام حسان، 123.

والحرف عندهم مبني لا محل له من الأعراب، ومنه ما يعمل ومنه غير العامل ولكن لا يعمل فيه غيره، أما عند المحدثين فقد حاول بعضهم إعادة النظر في أقسام الكلام ووضع تقسيمًا جديدًا مثل تمام حسان الذي جعل اقسام الكلام سبعة، وهي: (الاسم والصفة والفعل والضمير والخالفة والظرف والأداة)[6][7]، وقد اختلفت نظرته للحرف فأدخل فيه بعض الأسماء والأفعال التي تؤدي وظيفة الحرف النحوية وهي الربط والتعليق، وأطلق عليه اسم الأداة وعرفها بأنها: "مبنى تقسيمي يؤدي معنى التعليق، والعلاقة التي تعبر عنها الأداة إنما تكون بالضرورة بين الأجزاء المختلفة من الجملة"[7][7]، وتنقسم الأداة عند تمام حسان إلى قسمين: الأداة الأصلية: وهي الحروف ذات المعاني كحروف الجر والنسخ والعطف، وهذا ينطبق على الحرف عند القدماء، والأداة المحولة: وهي ما كان يؤدي وظيفة الحرف من الأسماء والأفعال والظروف والضمائر[7]، والباحث يميل إلى هذا التقسيم؛ لأن الكلمات عندما تؤدي وظيفة واحدة يجب وضعها تحت قسم واحد ولا داعٍ لتشتيتها بين الأقسام.

3- القواعد اللغوية ودورها في تحديد معاني الحروف في السياقات المختلفة

إن القواعد اللغوية لها دور في تحديد المعنى السياقي للحرف، ولا يشترط أن تكون القواعد نحوية، بل قد ترجع هذه القواعد إلى ما يصاحب الحرف من كلمات ذات خصائص معينة، أو تنتمي إلى قسم معين من أقسام الكلم، فعند تحقق ذلك نستطيع أن نحدد معنى الحرف في هذا السياق أو نقوي احتمالًا معينًا ونستبعد احتمالًا آخر، فعلى سبيل المثال إذا جاء بعد "ما" "إلا" تكون "ما" في هذا السياق هي النافية، ويكون هذا الأسلوب للحصر وليس للاستثناء، وأيضًا إذا جاء بعد "إن" فعل ناسخ ثم تلا ذلك اللام التي للتوكيد فإن ذلك يقوي احتمال كون "إن" في هذا السياق المخففة من الثقيلة والتي تستخدم لتوكيد الجملة الفعلية. والباحث يحاول استخراج مجموعة من القواعد التي تساعد على تحديد معنى الحرف في سياقه، ولا شك إن هذه الخطوة من المعالجة يجب أن تسبقها خطوات أخر، فيجب أولًا تقطيع النص (Tokenization)وذلك بفصل الكلمات عما يسبقها أو يلحق بها من السوابق واللواحق، ثم وسم النص (Tagging) والمقصود به تحديد قسم الكلام (POS) الذي تنتمي إليه كل كلمة في النص، فهذه الخطوات يجب أن تتم قبل تطبيق القواعد؛ لأن هذه القواعد تعتمد على نوع الكلمات المصاحبة للحرف، فيجب أن يمر النص أولًا على أحد التطبيقات التي تقوم بتقطيع النص وتحديد أقسام الكلام ثم بعد ذلك يتم تطبيق القواعد من أجل فك اللبس عن معاني الحروف ، وقد تمكن الباحث من ضبط جملة من القواعد وتجربتها، وهذه القواعد لا تحصر جميع الحروف بكل ما تحتمله من معاني إنما نستطيع من خلالها تحديد بعض المعاني لبعض الحروف، والمجال مازال مفتوحًا لضبط قواعد أخر.

4- قواعد تمييز معاني الهمزة

أ- *ضبط همزة الاستفهام*

قد تختلط همزة الاستفهام بهمزة النداء نظرًا لاتحاد الصورة "أ"، فهل نستطيع التمييز بينهما عن طريق القواعد؟ هناك مواضع انفردت بها همزة الاستفهام يمكن أن تضبطها القاعدة، فإذا وجدنا الهمزة في أحد تلك المواضع فنستطيع الجزم بأنها همزة الاستفهام ونقوم باستبعاد احتمال أن تكون همزة النداء، وهذه المواضع هي:

- إذا جاء بعد الهمزة فعل، فهي همزة الاستفهام لأن همزة النداء يجب أن يتلوها اسم، فإنك حين تقول: "أذهب أحمد إلى العمل؟" تكون الهمزة هنا للاستفهام بلا شك، ولا يمكن أن تأتي همزة النداء أبدًا في هذا التركيب؛ لأنك لا تنادي الفعل. ويمكن ضبط هذه القاعدة حاسوبيًا على النحو التالي:

"أ" + فعل = استفهام

- إذا جاء بعد الهمزة حرف عطف (الواو- الفاء – "ثم") أو حرف نفي ("لا"- "ما"- "لم") أو كلاهما معًا ("أولم"- "أوما"- "أفلا") فإنها أيضًا همزة الاستفهام، فقد قرر النحاة أن همزة الاستفهام وحدها لها الصدارة وأنها تتقدم حتى على حروف العطف وحروف النفي، كما في {أَلَمْ يَجِدْكَ يَتِيمًا فَآوَى}[8]، وفي {أَثُمَّ إِذَا مَا وَقَعَ آمَنتُم بِهِ}[9]، وفي {أَفَلَا تَعْقِلُونَ}[10]. قال المرادي في أثناء كلامه عن الهمزة: "وهي أصل أدوات الاستفهام، ولأصالتها استأثرت بأمور، منها، تمام التصدير بتقديمها على الفاء والواو و"ثم"، وكان الأصل في ذلك تقديم حرف العطف على الهمزة، لأنها

[6] اللغة العربية معناها ومبناها، تمام حسان، 122- 124.
[7] المرجع السابق، 123.
[8] الضحى، 6.
[9] يونس،51.
[10] البقرة، 44.

من الجملة المعطوفة. لكن راعوا أصالة الهمزة، في استحقاق التصدير، فقدموها "[11][1]، وهذا يمكن ضبطه حاسوبيًا كما يلي:

"أ" + (نفي / عطف) = استفهام

- إذا جاء بعد الهمزة حرف جر ("في"- "على"- "من") أو ظرف ("عند"- "مع") مثل {أَلَكُمُ الذَّكَرُ وَلَهُ الأُنثَى}[12]، {أَعِندَهُ عِلْمُ الْغَيْبِ فَهُوَ يَرَىٰ}[13]، "أفيكم محمد؟"، "أمنكم العقلاء؟"، ويكون الضبط الحاسوبي في هذه الحالة على النحو التالي:

"أ"+ (حرف جر/ ظرف) = استفهام

- وجود "أم" في الجملة بعد الهمزة يدل على أنها همزة الاستفهام، كما تقول: "أحضر عليٌّ أم سعيد؟". ولا تجيء "أم" مع همزة النداء، ويمكن صياغة ذلك كما يلي:

"أ" + "أم" = استفهام.

ب- ضبط همزة التسوية

قد تخرج الهمزة عن الاستفهام إلى معانٍ أخر تتحدد من خلال السياق، من هذه المعاني التسوية، وتعني أنه سيان عند المتكلم وقوع الفعل من عدمه، كما تقول: "سواء عليَّ أحضر محمد أم لم يحضر"، وكما في قوله تعالى: {سَوَاءٌ عَلَيْنَا أَجَزِعْنَا أَمْ صَبَرْنَا مَا لَنَا مِن مَّحِيصٍ}[14]، فكي تكون الهمزة للتسوية لابد أن يتقدمها في الجملة كلمة "سواء" وأن يتأخر عنها في الجملة "أم"، وذهب بعض النحاة إلى أنه قد يتقدمها كلمة تدل على هذا المعنى مثل ("ما أدري"- "ما أبالي") وما في معناها، ولكن رجح ابن هشام أن هذه الكلمات لا تنافي معنى الاستفهام فلا تدل على التسوية[15][10]، ويمكن ضبط هذه القاعدة حاسوبيًا على النحو التالي:

"سواء" + "أ" + "أم" = تسوية

5- قواعد تمييز معاني "إن"

الحرف "إن" له عدة أنواع، فهناك "إنَّ" المشدَّدة الثقيلة، وهناك "إن" المخففة من الثقيلة، وهناك "إن" النافية، و"إن" الشرطية، وهناك "إن" الزائدة، وكلها تأخذ هيئة واحدة ، فهل من سبيل للتمييز بينها عن طريق القواعد؟ صحيح تتميز "إنَّ" الثقيلة بالتشديد والفتح، لكن إذا غاب التشكيل تصبح الصورة واحدة، ولذلك لابد من البحث عن قواعد تمكن من التمييز بين تلك الصور أو على الأقل ضبط بعضها.

أ- تمييز "إنْ" المخففة من الثقيلة

تدخل "إن" (بكسر الهمزة) المخففة من الثقيلة على الجملتين الاسمية والفعلية، ولكنها تأتي مع الفعلية أكثر، ودائمًا يصحبها اللام التي للتوكيد، وحيث وُجِدت "إن" وبعدها اللام المفتوحة فيُحْكَم عليها بأن أصلها التشديد"[16][10]، ومن ذلك قوله تعالى: {وَإِن كَانَتْ لَكَبِيرَةً إِلَّا عَلَى الَّذِينَ هَدَى اللَّهُ}[17]، وقوله: {إِنْ كَادَ لَيُضِلُّنَا عَنْ آلِهَتِنَا لَوْلَا أَنْ صَبَرْنَا عَلَيْهَا}[18]، وكذلك قوله عز وجل :{وَإِنْ يَكَادُ الَّذِينَ كَفَرُوا لَيُزْلِقُونَكَ بِأَبْصَارِهِمْ لَمَّا سَمِعُوا الذِّكْرَ}[19]، وقوله: {وَإِنْ كُنْتَ

---

[11] الجنى الداني، المرادي، 31.
[12] النجم، 21.
[13] النجم، 35
[14] إبراهيم، 21.
[15] مغني اللبيب، ابن هشام، 62.
[16] مغني اللبيب، 37.
[17] البقرة، 143
[18] الفرقان، 42.
[19] القلم، 51.

مِنْ قَبْلِهِ لَمِنَ الْغَافِلِينَ}[20]، ومن ذلك أيضًا: {وَإِنْ وَجَدْنَا أَكْثَرَهُمْ لَفَاسِقِينَ}[21]، وأيضًا قوله: {وَإِنْ كُنَّا لَمُبْتَلِينَ}[22]، فنلاحظ في الشواهد السابقة أنّ "إن" تدخل على الجملة الفعلية، وأنه في الغالب يتبعها فعل ناسخ ماض، كما في ("كانت"، "كنت"، "كنا"، "كاد")، وقد يأتي مضارعًا، كما في "يكاد"، وقد يأتي غير ناسخ، كما في "وجدنا"، وقد رأينا أن اللام التي للتوكيد لم تفارقها أبدًا في كافة النماذج المتقدمة، وهذا يجعلنا نقول: إذا جاء بعد "إن" فعل، وجاءت بعدها اللام فهي "إن" المخففة من الثقيلة، والتي تأتي لتوكيد الجملة الفعلية، وأما القاعدة فتصاغ حاسوبيًا كما يلي:

إن + فعل + لام التوكيد = إن المخففة من الثقيلة المؤكدة للجملة الفعلية

ب- تمييز "إنْ" النافية

وأما "إن" النافية، فتدخل على الأسماء وعلى الأفعال، وتفيد النفي مثل "ما" النافية، وفي كثير من الأحيان يأتي بعدها "إلا" أو "لمّا" ، وهذا لا يعني أنها لا تجيء إلا هكذا، فقد تأتي بدونهما، ولكن هذا نادر، "فقد وردت "إن" النافية في القرآن في عشرة ومئة موضع كلها مقترنة بـ"إلا" أو "لما" عدا سبع آيات[23][9]، ومن شواهد ذلك قوله تعالى: {إِنْ أَنْتَ إِلَّا نَذِيرٌ}[24]، وقوله عز وجل: {وَإِنْ مِنْ أُمَّةٍ إِلَّا خَلَا فِيهَا نَذِيرٌ}[25]، وقوله سبحانه: {إِنْ عَلَيْكَ إِلَّا الْبَلَاغُ}[26]، وقد يأتي بعد "إن" النافية فعل، كما في قوله: {إِنْ أَرَدْنَا إِلَّا الْحُسْنَى}[27]، وكذلك {إِنْ يُرِيدُونَ إِلَّا فِرَارًا}[28]، فنستطيع القول إن "إن" إذا جاء بعدها "إلا" فهي النافية، ونصوغ القاعدة كما يلي:

"إن" + "إلا" = "إن" النافية

ولكن قد رأينا في أحد الشواهد أن "إن" قد جاءت بعدها "إلا" وحكمنا عليها بأنها "إنّ" المخففة من الثقيلة، وهو قوله تعالى: {وَإِن كَانَتْ لَكَبِيرَةً إِلَّا عَلَى الَّذِينَ هَدَى اللَّهُ}[29]، والسبب الذي جعلنا نقول إنها المخففة من الثقيلة هو وجود لام التوكيد بعدها ("لكبيرة")، وهذا يحتم علينا إذا وجدنا "إن" وبعدها فعل أن نبحث أولًا عن اللام فإذا وجدناها فهي المخففة من الثقيلة، وإذا لم نجدها نبحث عن "إلا" فإذا وجدناها فهي النافية، وهذا الأمر يمكن برمجته حاسوبيًا.

وأما عن "لمّا" فقد ورد في التنزيل {وإن كلٌ لمّا جميعٌ لدينا محضرون}[30]، {وإن كل ذلك لمّا متاع الحياة الدنيا}[31]، {إن كل نفس لمّا عليها حافظ}[32]، والضابط الحاسوبي هنا:

"إن" + "كل" + "لمّا" = نفي

ولابد هنا أن تكون "كل" مرفوعة، أما إذا كانت منصوبة فتكون "إن" في هذا التركيب هي الثقيلة أو المخففة منها، وتنوين النصب يمثّل بالألف التي تكون في نهاية "كل"، ويمكن تمثيل ذلك حاسوبيًا كما يلي:

"إن" + "كلا" + "لمّا" = "إنَّ" / "إنْ" المخففة منها

6- قواعد تمييز معاني "إذا"

"إذا" ظرف للزمان المستقبل مضمَّن معنى الشرط[10]، ولها ثلاث أحوال: "إذا" الفجائية، و"إذا" الشرطية، و"إذا" الظرفية، فأما الفجائية فهي التي تدل على المفاجأة، وتدخل على الجملة الاسمية، كما في الآية الكريمة: {فَأَلْقَاهَا فَإِذَا هِيَ حَيَّةٌ تَسْعَى}[33]، وأما الشرطية فتدخل على الجملة الفعلية غالبًا، وهي تربط بين جملتين الأولى جملة الشرط، والثانية جواب الشرط، فيجب أن يأتي بعد الفعل جملة فعلية أو اسمية مقترنة بالفاء، ومن ذلك قوله عز وجل: {وَإِذَا قِيلَ لَهُمْ لَا تُفْسِدُوا فِي الْأَرْضِ قَالُوا إِنَّمَا نَحْنُ مُصْلِحُونَ }[34]، أما الظرفية فتشير إلى زمان وقوع الحدث، ويأتي بعدها فعل

---

[20] يوسف، 3.
[21] الأعراف، 102.
[22] المؤمنون، 30.
[23] معاني النحو: فاضل السمرائي، 35/4.
[24] فاطر، 23.
[25] فاطر، 24.
[26] الشورى، 48.
[27] التوبة، 107.
[28] الأحزاب، 13.
[29] البقرة، 143
[30] يس، 32
[31] الزخرف، 35
[32] الطارق، 4
[33] طه، 20.
[34] البقرة، 11.

دائمًا ويكثر مجيئها بعد القَسَم. وعند تتبع الشواهد في المدونة العربية العالمية (ICA) تبين للباحث أن الاستخدام المعاصر لـ"إذا" الشرطية يأتي أحيانًا على الترتيب المعتاد: "إذا" ثم فعل الشرط ثم جواب الشرط، كما في الأمثلة :

"فإذا رفض المجلس جاء الرئيس بوزارة أخرى"

"فإذا سأل سائل فردا من هو ؟ أجاب إجابة في صيغة جمعية"

وأحيانًا يتقدم الجواب على "إذا" وفعل الشرط كما في الأمثلة التالية:

"عليه أن يتوسع إلى مناطق أخرى إذا نجحت التجربة"

"لم يستطيعوا أن يقرؤوا ، أو يكتبوا شيئا ، إذا لم تكن أمامهم أية حروف سوى حرف واحد"

أما الفجائية فلابد أن يتلوها اسم[10]، وفي الغالب إذا كان هذا الاسم ضميرًا فإنه يكون مقترنًا بباء الجر، كما نرى في تلك الشواهد:

"فإذا بها تكتشف وفاته ، فتضيق بها الدنيا"

"انتصبت واقفة ، لأسرع لنجدته .. فإذا بي أرى رجلا ينبثق من بين البيوت"

وفي بعض الأحيان يليها اسم ظاهر كما في:

"بدأ يرفس الهواء ، ثم يهبط ويرمح مرة ، مرتين .. فإذا المحراث يتفكك ويتطاير قطعا رغم أنه من الخشب القوي"

وفي كل الأمثلة السابقة نلاحظ مجيء فعل مضارع بعد الاسم أو الضمير بعدها ("تكتشف"، "أرى"، "يتفكك")، وقد يأتي بعدها ما يقوم مقام الفعل المضارع كاسم الفاعل أو الصفة المشبهة، كما في قوله تعلى: {حَتَّى إِذَا فَتَحْنَا عَلَيْهِمْ بَابًا ذَا عَذَابٍ شَدِيدٍ إِذَا هُمْ فِيهِ مُبْلِسُونَ}[35]، وكما في قول الشاعر: "قَد غَفَونا وَاِنتَبَهنا .....فَإِذا نَحنُ غَرقى وَإِذا المَوتُ أَمَم"[2]. وأحيانًا يجيء بعد الاسم فعل ماض مسبوق بـ"قد" كما في المثال: "خرجنا لملاقاته على الطريق فإذا به قد أقبل والأنوار تسبقه". وهذا يعني أن أهم ما يميز "إذا" الفجائية دخولها على الجملة الاسمية، وغالبًا ما يكون المبتدأ ضميرًا وفي كثير من الأحيان في الاستعمال المعاصر يتصل بباء الجر، وأما عن الخبر ففي الغالب يكون جملة فعلية فعلها مضارع، وقد يكون الفعل ماضيًا مسبوقًا بـ"قد"، وقد يكون الخبر مفردًا.

وأما "إذا" الظرفية فتتشابه كثيرًا "مع" الشرطية، فيأتي بعدها فعل قد يكون ماضيًا كما في قوله عز وجل: {وَالنَّجْمِ إِذَا هَوَى}[36]، وقد يكون مضارعًا كما في قوله تعالى: {وَاللَّيْلِ إِذَا يَغْشَى}[37]، وقد يأتي بعدها اسم، فقد وقع الخلاف بين النحاة في "إذا" التي يتلوها اسم كما في الآية: {فَإِذَا النُّجُومُ طُمِسَتْ}[38]، والآية: { إِذَا السَّمَاءُ انْشَقَّتْ}[39].

ومن ذلك كله نستطيع أن نقول إنَّ "إذا" حين يتلوها فعل تنحصر بين الشرطية والظرفية، ونستبعد كونها فجائية، فتصبح القاعدة:

"إذا" + فعل = الشرطية/الظرفية

أما إذا تلاها اسم فإننا ننظر إلى ما بعده فإن كان فعلًا ماضيًا غير مسبوق بـ"قد" فهي شرطية أو ظرفية، وإن كان غير ذلك فهي فجائية، ونصوغ القواعد كما يلي:

"إذا" + اسم + فعل ماض ليس مسبوق بـ"قد" = شرطية/ظرفية

"إذا" + اسم + (صفة/فعل مضارع/فعل ماض مسبوق بـ"قد") = فجائية

7- قواعد تمييز معاني "ما"

إنّ "ما" من الكلمات التي تتعدد أنواعها ومعانيها، فهناك "ما" النافية، و"ما" الموصولة، و"ما" الاستفهامية، و"ما" الشرطية، وغير ذلك، وبغض النظر عن كونها اسمًا أو حرفًا فسوف ننظر إليها على أنها تؤدي وظيفة الحرف في الجملة طبقًا لوجهة النظر التي تضع كل الكلمات التي تؤدي معنى الحرف في قسم واحد وتسمّيه الأداة، ويحاول الباحث التمييز بين بعض أنواع "ما" عن طريق القواعد.

أ- ما الاستفهامية

[35] المؤمنون، 77.

[36] النجم، 1.

[37] الليل، 1.

[38] المرسلات، 8.

[39] الانشقاق، 1.

من القواعد الظاهرة التي تميز "ما" الاستفهامية عن غيرها وجوب حذف الألف منها إذا سبقها حرف جار، قال ابن هشام: "وَيجب حذف ألف مَا الاستفهامية إذا جُرّت وإبقاء الفتحة دَلِيلا عَلَيْهَا نَحْو "فيمَ" و"إلام" و"علام""[40][10]، ومن نماذج ذلك قوله تعالى: {فَنَاظِرَةٌ بِمَ يَرْجِعُ الْمُرْسَلُونَ}[41]، وقوله: {لِمَ تَقُولُونَ مَا لَا تَفْعَلُونَ}[42]. لكن قد تشتبه "ما" الاستفهامية إذا لحقت ببعض حروف الجر ببعض الكلمات مثل "لَمْ"، وأيضًا "عَمَّ" قد تشتبه بـ "عَمَ"، و"عَلَامَ" قد تشتبه بـ "عَلَّام"، وهذا ليس داخلًا في نطاق هذا البحث، أما عند اتصالها بحروف الجر الأخرى كـ "في" و"إلى" والباء فيمكن تمييزها بحذف الألف منها.

ب- "ما" المصدرية أو الموصول الحرفي:

تأتي "ما" مصدرية فتدخل على الفعل فتكون هي والفعل في تأويل مصدر، مثل: {فَاتَّقُوا اللَّهَ مَا اسْتَطَعْتُمْ}[43]، والمعنى: "فاتقوا الله استطاعتكم"، وأيضًا مثل قوله عز وجل: {وَدُّوا مَا عَنِتُّمْ}[44]، أي: "ودوا عنتكم"، وهناك حالة واحدة يمكن تمييز "ما" المصدرية فيها آليًا، وهي عندما تقع بعد كاف التشبيه بين فعلين متماثلين[45][10]، كما في قوله تعالى: {وَإِذَا قِيلَ لَهُمْ آمِنُوا كَمَا آمَنَ النَّاسُ قَالُوا أَنُؤْمِنُ كَمَا آمَنَ السُّفَهَاءُ}[46]، والمراد: "إذا قيل لهؤلاء آمنوا مثل إيمان الناس قالوا أنؤمن مثل إيمان السفهاء"، ومنه أيضًا قوله تعالى: {إِنْ تَكُونُوا تَأْلَمُونَ فَإِنَّهُمْ يَأْلَمُونَ كَمَا تَأْلَمُونَ}[47]، أي: "يألمون مثل ألمكم"، ومن أمثلة ذلك أيضًا، قوله تعالى: {إِنَّا أَرْسَلْنَا إِلَيْكُمْ رَسُولًا شَاهِدًا عَلَيْكُمْ كَمَا أَرْسَلْنَا إِلَى فِرْعَوْنَ رَسُولًا}[48]، والفعلان السابق لـ"كما" والتالي لها لا يشترط أن يتطابقا، فقد يجيء أحدهما ماضٍ والآخر مضارع، وقد يجيء أحدهما فعل أمر والآخر ماضٍ كما في "آمنوا كما آمن"، فلابد أن يكون البحث هنا عن مادة الفعل وليس الفعل نفسه، ثم نصوغ القاعدة :

فعل +كاف التشبيه + "ما" + فعل من نفس مادة الأول = "ما" المصدرية

ت- ما النافية:

تعد "ما" النافية من حروف النفي التي تدخل على الأسماء وعلى الأفعال، ويمكن تمييزها آليًا في الأحوال الآتية:

- إذا جاء بعد "ما" "كان" وبعدها جار ومجرور ثم أتبع بمضارع مسبوق بـ"أن" فهي نافية، كما في قوله عز وجل: {مَا يَكُونُ لِي أَنْ أَقُولَ مَا لَيْسَ لِي بِحَقٍّ}[49]، وقوله: {وَمَا يَكُونُ لَنَا أَنْ نَعُودَ فِيهَا إِلَّا أَنْ يَشَاءَ اللَّهُ رَبُّنَا}[50]، وفي قوله تعالى: {ما كَانَ لِلَّهِ أَنْ يَتَّخِذَ مِنْ وَلَدٍ }[51]، وقوله: {مَا كَانَ لَكُمْ أَنْ تُنْبِتُوا شَجَرَهَا}[52]، وكذلك قوله عز وجل: {وَمَا كَانَ لِبَشَرٍ أَنْ يُكَلِّمَهُ اللَّهُ إِلَّا وَحْيًا}[53]، ومن ذلك أيضًا "وضع النقاط فوق حروف ما كان لها أن تكتب بدونه"، وكذلك "تتساوى في لحظات الاختيار الأصعب أحزاب ما كان لها أن توضع في سلة واحدة"، فكل هذه المواضع وغيرها قد جاءت فيها "ما" وبعدها "كان" في الماضي أو المضارع، ثم بعد ذلك "أن" مع الفعل المضارع، وهي نافية في كل ذلك بلا استثناء، فعلى هذا يمكن ضبط القاعدة:

  "ما" + "كان/يكون" + "أن" + فعل مضارع = "ما" النافية

- إذا جاءت "ما" ثم تلاها "إلا" أو "غير" أو "سوى" فهي النافية، وهنا يكون الأسلوب أسلوب حصر، وهناك نماذج كثيرة لذلك منها قوله تعالى: {وَمَا مِنْ إِلَهٍ إِلَّا اللَّهُ}54، وقوله عز وجل: {وَمَا مُحَمَّدٌ إِلَّا رَسُولٌ}55، وقوله سبحانه:

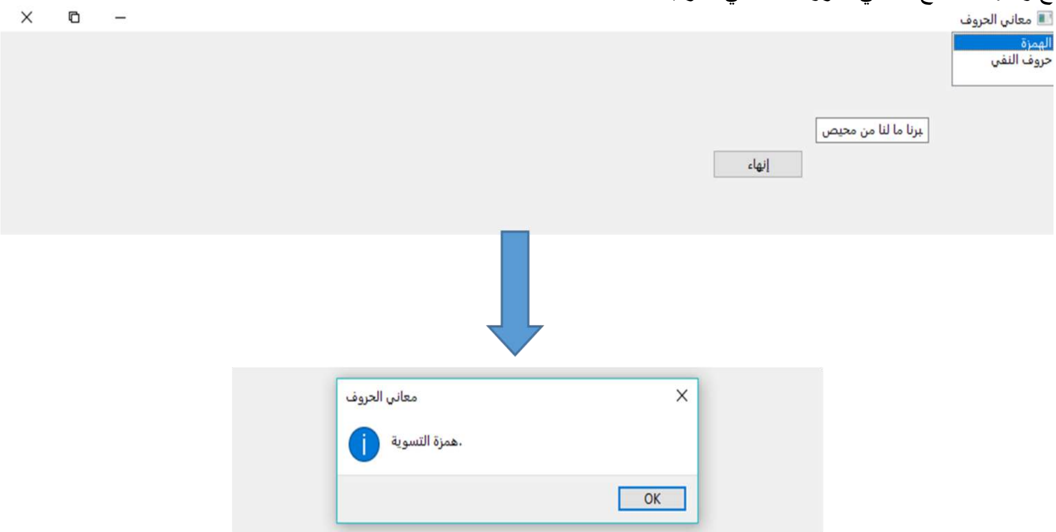
---

[40] مغني اللبيب، 393.
[41] النمل، 35.
[42] الصف، 2.
[43] التغابن، 16.
[44] آل عمران، 118.
[45] انظر مغني اللبيب، 400.
[46] البقرة، 13.
[43] النساء، 104.
[48] المزمل، 15.
[49] المائدة، 116.
[50] الأعراف، 88.
[51] مريم، 35.
[52] النمل، 59.
[53] الشورى، 51.
[54] آل عمران، 62.
[55] آل عمران، 144.

{وَمَا أَرْسَلْنَاكَ إِلَّا رَحْمَةً لِلْعَالَمِينَ}[56]، وكذلك قوله تعالى: {مَا لَكُمْ مِنْ إِلَهٍ غَيْرُهُ}[57]، وقول الشاعر: "نعيب زماننا والعيب فينا...وما لزماننا عيب سوانا". ومن أمثلة ذلك أيضًا "ما كان منه إلا أن يسلم ساقيه للريح"[58]، وكذلك "ليكون مثلا يقتدي به رفاقه الذين ما كان يلقى منهم سوى الخشونة"[59]، فتكون القاعدة كما يلي:

"ما" + "إلا/غير/سوى" = "ما" النافية في أسلوب الحصر

وقد قام الباحث بتصميم تطبيق حاسوبي بغرض اختبار صحة القواعد التي توصل إليها في تحديد معاني الحروف آليًا، بحيث يتم إدخال جملة تحتوي على الحرف المراد تحديد معناه، ثم الضغط على الحرف بحيث يقوم الحاسوب بتطبيق القواعد المتعلقة بهذا الحرف، ثم بعد ذلك تظهر رسالة تحتوي على معنى الحرف الذي توصل إليه الحاسوب بعد تطبيق القواعد، وفي حالة عدم توافر القواعد في النص المدخل فإن التطبيق يظهر رسالة تفيد عدم إمكانية تحديد معنى الحرف، فهذه القواعد لا تحصر جميع معاني الحروف، وإنما هى تضبط جملة من تلك المعاني، والشكل التالي يوضح واجهة المعالج الدلالي لحروف المعاني العربية.



شكل (1): واجهة المعالج الدلالي لحروف المعاني العربية

8- النتائج والتوصيات

قد أوضحت هذه الورقة إمكانية استخراج قواعد لغوية يمكن تغذية التطبيقات الحاسوبية بها لترفع قدرتها على فك اللبس عن معاني الحروف وتحديد دورها الدلالي أو على الأقل ترجيح بعض المعاني المحتملة للحرف في النص الوارد به، وقد تمكن الباحث من استخلاص بعض هذه القواعد وصياغتها صياغة ملائمة للحاسوب، كما قام بتجربة هذه القواعد من خلال تطبيق حاسوبي بسيط للتحقق من صحتها، ويوصي الباحث بضرورة مواصلة البحث في مجال استخلاص القواعد اللغوية التي يمكن أن تدعم تطبيقات المعالجة الآلية للغة العربية مما يساعد على تحسين تلك التطبيقات بشكل يخدم اللغة العربية والناطقين بها ومحبيها.

[56] الأنبياء، 107.
[57] الأعراف، 59،65،73،85.
[58] المدونة اللغوية العربية الدولية، www.bibalex.org/ica/ar
[59] المدونة اللغوية العربية الدولية، www.bibalex.org/ica/ar

المراجع العربية

[1] ابن قاسم المرادي، الجنى الداني في حروف المعاني، تحقيق: فخر الدين قباوة، محمد نديم فاضل، دار الكتب العلمية، بيروت، ط: 1، 1992م.
[2] ديوان حافظ إبراهيم.
[3] ديوان الشافعي
[4] أحمد بن عبد النور المالقي، رصف المباني في شرح حروف المعاني، تحقيق: أحمد الخراط، دار البشير، جدة، ط: 3، 2002م.
[5] ابن يعيش، شرح المفصل، دار الكتب العلمية، بيروت – لبنان، ط: 1، 2001م.
[6] سيبويه، الكتاب، تحقيق: عبد السلام محمد هارون، مكتبة الخانجي، القاهرة، ط: 3، 1988م.
[7] تمام حسان، اللغة العربية معناها ومبناها، عالم الكتب، ط: 5، 2006م.
[8] ابن جني، اللمع في العربية، تحقيق: فائز فارس، دار الكتب الثقافية ـ الكويت.
[9] فاضل صالح السمرائي، معاني النحو، دار الفكر للطباعة والنشر والتوزيع، الأردن، ط: 1، 2000 م.
[10] ابن هشام، مغني اللبيب عن كتب الأعاريب، تحقيق: مازن المبارك، محمد علي حمد الله، دار الفكر، دمشق ط: 6، 1985م.
[11] المبرد، المقتضب، تحقيق: محمد عبد الخالق عظيمة، عالم الكتب، بيروت.
[12] المدونة اللغوية العربية الدولية، www.bibalex.org/ica/ar
[13] مجلة فكر الثقافية، https://www.fikrmag.com/article_details.php?article_id=368

السيرة الذاتية

خالد مصطفى أبو شبانة، باحث دكتوراه قسم اللغة العربية كلية الآداب جامعة الإسكندرية، موضوع رسالة الدكتوراه يدور حول حوسبة حروف المعاني العربية حصلت على الماجستير في تخصص اللغة من جامعة دمنهور سنة 2015، أعمل بوزارة التربية والتعليم بالإسكندرية، كما أعمل مدربًا للحاسب الآلي للمكفوفين وضعاف البصر، وكذلك معلمًا للغة العربية عن بعد.

The Role of Linguistic Rules In automatic disambiguation of Particles Meanings

Khaled Mustafa Abou-shabana

PhD researcher, Arabic language department, faculty of Arts, Alexandria University

kh33shabana@gmail.com

Abstract— This paper deals with the role played by Arabic linguistic rules in designing natural language processing NLP applications Such as machine translation applications, automatic text summarization, grammar and spelling correction, and vocabulary statistics. One of the problems that face these applications, which is related to the semantic side, is the problem of the various meanings that one word can tolerate and that it is difficult for the computer to distinguish between them, and the research focuses on what is related to the Arabic particles. This is because of the central role it plays in linking the sentences. As well as particles always keep its form and don't change when its meaning changes. Also when used with a verb it give the verb more than one meaning So it needs deep analysis in order to catch its meaning. The research attempts to monitor linguistic rules that may be syntactic or contextual and that can be computerized so that the application enables to determine the meaning of the particle in the sentence that it processes, the researcher has set some linguistic rules that help to define the meanings of the particle in the context in which it was formulated and formulated these Grammar Computer Formulation, then tried it in practice through a computer application, which serves as an SDK tool that can be integrated into other applications.

Keywords: **Linguistic Rules -Semantic Ambiguity - Particles - NLP – Linguistic Corpora - Computer Applications**

Speech processing

قياس انفعالات الممثل الصوتية باستخدام تقنيات هندسة اللغة

صديقة عبد المنعم عبد الغنى أحمد لاشين

أستاذ مساعد

*قسم الدراسات المسرحية-كلية الآداب-جامعة الإسكندرية
4عمارات الطابية-السرايا-الإسكندرية
siddeeka_lasheen@yahoo.com

ملخص البحث:

عنوان البحث" *قياس انفعالات الممثل الصوتية باستخدام تقنيات هندسة اللغة* "،هدف البحث إلى تقييم أسس العملية الإبداعية وإخضاعها للبحث العلمى تماشيا مع التكنولوجيا الحديثة. فعن طريق التحليل العلمى للصوت باستخدام البرامج الحديثة يمكننا أن نحكم على مدى تحكم الممثل بصوته فى التعبير عن الانفعالات المختلفة كخطوة أساسية فى الوصول إلى حالة الإبداع. مشكلة الدراسة هى ما مدى تحقق فرضية قياس الانفعالات التى يعبر عنها الممثل عن طريق تحليل الصوت مع عدم المساس بالحالات الإبداعية . لذا وجب تقسيم الدراسة إلى ثلاثة أقسام، القسم الأول وهو بعنوان "فن الآداء التمثيلى وعلاقته بالصوت" وتتناول الباحثة فيه الصوت وعلاقته بفن التمثل، وماهية قواعد الإلقاء التى تستخدم فى مجال الأداء التمثيلى، والقسم الثانى بعنوان: الصوت وعلم هندسة اللغة، وتتناول الباحثة أهم المصطلحات المستخدمة فى مجال هندسة اللغة، أما القسم الثالث فهو الجانب التطبيقى ووقع اختيار الباحثة على نموذج أداء صوتى تمثيلى من مسرحية هاملت تأليف ويليم شكسبير، وأداء عدد من الممثلين المختلفين مكانيا وزمانيا. وتم تحليل الأصوات فيها باستخدام برنامج برات PRAAT . وينتهى البحث بعرض أهم النتائج ومناقشتها ثم التوصيات، وأخيرا قائمة المصادر والمراجع.

كلمات مفتاحية: *Keywords:* الانفعال الصوتى-هندسة اللغة- الأداء التمثيلى-قياس الصوت- .PRAAT-

مقدمة

مما لا شك فيه أن صوت الممثل هو آداة من أدواته الرئيسة للتعبير، فهو عنصر أساسى لا يقل أهمية عن الجسد فى عملية التعبير؛ حيث إنه لابد أن تتضح فيه صفات الشخصية المؤداة، وكامل تفاصيلها من سن وجنس وصفات نفسية واجتماعية وغيرها مما ذكرها المنظرون لفن المسرح، بدءأ من أرسطو حتى النظريات الحديثة، وما بعدها. فالأبعاد الأساسية التى لابد أن يعبر عنها المثل، أولها البعد الطبيعي: ويعنى بها طبيعة الشخصية من الناحية التشريحية والوظيفية. ثم البعد الاجتماعى : ويخص صفات الشخصية من الناحية البيئية والعلاقات الاجتماعية والوضع الاقتصادى، وأخيرا البعد النفسى: ويخص صفات الشخصية المؤداة النفسية"[1]

وتكمن القدرة الإبداعية للمثل فى كيفية توظيف أدواته الرئيسة (صوت- جسد) فى خلق الشخصية المسرحية. ومن هنا ظهرت مدارس تدريب وإعداد الممثل بشكل عام من ناحية، وإعداده لأداء الدور المسرحى من ناحية أخرى وذلك للوصول إلى حالة الإبداع الخلاقة. ففن الأداء التمثيلى ‘‘فن خلّاق، كفن العازف، يستخدم الممثل من خلاله كل مفردات جسمه وصوته لنفس الغرض.. لينقل خيال الكاتب بصدق وبراعة.. إنه إنسان خلاق مبدع من خلال أدواته ولغته الخاصة إذا عرف كيف يستخدمها.’’[2]

ولكن هل تعتبر حالة الإبداع الخلاقة هذه حالة غير قابلة للقياس؟ حالة تحكمها ذات الممثل، أو المؤدى دون الاستناد على أسس علمية لقياسها، ومن ثم الحكم عليها؟

فمن المتعارف عليه أن الأداء التعبيرى لصوت الممثل وجسده فن إبداعى، فكم من الممثلين برعوا كل منهم فى أداء الدور نفسه، حيث يُظهر ذلك كم التدريب الذى تعرض له الممثل، ومخزون الخبرة لديه ومن ثم هذه العملية ذاتية إبداعية. ولكن كيف يمكن إخضاع هذه العملية الإبداعية للقياس؟ وهنا تظهر مشكلة الدراسة، وقد اختارت الباحثة أحد أدوات التعبير للممثل والذى يتمثل فى صوته . وذلك للأسباب الآتية:

1- أن الصوت بشكله المجرد- بعيدا عن الشخصية المؤداه- ، هو عنصر قابل للقياس. خاصة مع الثورة التكنولوجية المعاصرة.
2- هناك بعض القواعد الحاكمة لأداء الممثل الصوتى، تلك القواعد تتمثل فى فن الإلقاء بقواعده الرئيسة من مخارج حروف، وتفخيم وترقيق، ونبر، تنغيم ...إلخ والتى يتم قياسها فى الأداء التمثيلى من خلال (عملية السمع) الأذن. وهى عملية ذاتية أيضا، تخضع لمدى إلمام مدرب الأداء التمثيلى بقواعد اللغة، فضلا عن إلمامه بقواعد فن الإلقاء، ولكن ما هو معيار القياس فى حال عدم إلمام المدرب بتلك القواعد؟ مع اطلاع الباحثة علي بعض برامج تحليل الصوت وجدت أن هذه العناصر أيضا قابلة للقياس من

خلال البرامج الإليكترونية التى تحدد شده الصوت، ونوعه، ودرجته، وكذلك قواعد النبر المتحكمة فيه. هذه البرامج تسهل على المدرب وضع معايير للحكم على مدى إتقان الممثل وتمكنه من اللغة وقواعدها، دون الاعتماد على المعايير الذاتية فى عملية التقييم.

3- المحرك الثالث لهذه الدراسة يتمثل فى اهتمام الباحثة بالدراسات البينية التى تجمع أكثر من مجال مما يفيد ويثرى البحث العلمى.

4- اذا كان عنصر الصوت هو عنصر قابل للقياس والحكم عليه، فإن ذلك يفيد مدربي الأداء التمثيلى للوقوف على أهم المشاكل التى يعانى منها صوت الممثل، ومن ثم إيجاد الطرق العلمية الفعالة لحل تلك المشاكل.

5- التقييم الذاتى هو عملية تعتمد على خبرة المدرب، وحالته النفسية، وتركيزه فى تفاصيل الشخصية محل التقييم.

والتساؤل الذى يطرح نفسه هنا، ويمثل مشكلة الدراسة هو ما مدى تحقق فرضية قياس الانفعالات التى يعبر عنها الممثل عن طريق تحليل الصوت مع عدم المساس بالحالات الإبداعية، وأقصد هنا بأنه -على سبيل المثال- هناك اثنان أو أكثر من الممثلين قد برعوا فى أداء دور ماكبث ، كل بطريقته، وقد نرى أنهم على نفس القوة فى تحقيق حالة الإبداع، هذا التقييم ذاتى، ماذا لو تم إخضاع أصواتهم إلى القواعد العلمية لتحليل أصواتهم ؟ هل نظل على الرأى نفسه؟ أم يتغير؟

ليس الهدف فى هذا البحث انتفاء الدور الإبداعى للممثل، ولكن الهدف هو تقييم أسس العملية الإبداعية وإخضاعها للبحث العلمى تماشيا مع التكنولوجيا الحديثة. فعن طريق التحليل العلمى للصوت باستخدام البرامج الحديثة يمكننا أن نحكم على مدى تحكم الممثل بصوته فى التعبير عن الانفعالات المختلفة كخطوة أساسية فى الوصول إلى حالة الإبداع.

ولمحاولة الإجابة عن هذه التساؤلات، اختارت الباحثة عنوان: "قياس انفعالات الممثل الصوتية باستخدام تقنيات هندسة اللغة" ، لذا وجب تقسيم الدراسة إلى ثلاثة أقسام ، القسم الأول وهو بعنوان "فن الآداء التمثيلى وعلاقته بالصوت" وتتناول الباحثة فيه الصوت وعلاقته بفن التمثيل، وماهية قواعد الإلقاء التى تستخدم فى مجال الأداء التمثيلى، والقسم الثانى بعنوان: الصوت وعلم هندسة اللغة، وتتناول الباحثة أهم المصطلحات المستخدمة فى مجال هندسة اللغة، أما القسم الثالث فهو الجانب التطبيقى، ووقع اختيار الباحثة على نموذج أداء صوتى تمثيلى من مسرحية هاملت تأليف ويليم شكسبير، وأداء عدد من الممثلين المختلفين مكانيا وزمانيا. وسيتم تحليل الأصوات فيها باستخدام برنامج برات وهو تطبيق تحليل الإشارات الصوتية ومعالجتها PRAAT . وينتهى البحث بعرض أهم النتائج ومناقشتها ثم التوصيات، وأخيرا قائمة المصادر والمراجع.

أولا: الصوت وعلاقته بفن الأداء التمثيلي:

يقسم الآداء التمثيلى وفق ثلاث أنواع رئيسة، تتمثل فى الآتى:[3]

1- نوع ملفوظ يعتمد علي اللفظ الصوتي المعبر مثل التمثيل الإذاعي، يدرك بالأذن
2- نوع حركي يعتمد علي الحركة المعبرة، مثل التمثيل الصامت،والرقص يدرك بالعين
3- نوع ملفوظ وحركي، مثل التمثيل علي خشبة المسرح والتلفزيون والسينما.

وفى هذا التصنيف، نجد أن الصوت بشكله المجرد يعد أحد دعائم فن الممثل الرئيسة ، حتى فى الفنون التى تعتمد على الحركة، فهناك بعض الهمهمات أو الأنات التى قد يعبر بها الممثل عن انفعالاته المختلفة، وينسحب الأمر على لحظات الصمت المعبرة ، واستخدام المؤثرات الصوتيه أو الموسيقى التى تهدف إلى نقل الحالات الشعورية المختلفة.

- فن الإلقاء والانفعال:

فن الإلقاء:[4] :(Didacticism) "كلمة مأخوذة من اللاتينية وتعني الكلام، وتستعمل للدلالة على فن اللفظ طريقة الكلام أو طريقة إلقاء الشعر أو النثر. "
أما الإلقاء في المسرح:
"هو فن لفظ النص المسرحي ، ومقوماته مهارة نطق مخارج الحروف وحسن استخدام نبرة الصوت ونغمته وشدته وسرعة الكلام وإيقاعه. وتتفاوت طبيعة الإلقاء في المسرح ما بين إبراز قيمة الصوت كعنصر سمعي في لفظ أقرب إلى الترتيل وهذا هو التنغيم ، وبين إبراز المعنى من خلال إلقاء طبيعي يشبه لهجة الحديث العادي ، وذلك تبعا لاختلاف مدارس التمثيل"

يقول الفنان المصري الراحل عبد الوارث عسر عن فن الإلقاء وعلاقته بالأداء التمثيلى: "إنه فن النطق بالكلام على صورة توضح ألفاظه ومعانيه، وتوضيح اللفظ يعني دراسة الممثل للحروف الأبجدية وصفات كل حرف، ليخرج من الفم سليما معافى، كاملا، واضحا، بلا عيوب لفظية أو نطقية، أما توضيح المعنى فيأتي من خلال دراسة وافية للصوت الإنساني: طبقاته، حدته، وأيضا دراسة موسيقية للأصوات من حيث النغمات التي تتناسب والمعاني (الحوار الرومانسي مثلا يتطلب صوتا ناعما هادئا).. لتأتي مطابقة لدرجة الصوت وتحدث وقعا لطيفا على أذن السامع.. وفي مجال التدريب نطلب من الممثل مثلا أن يؤدي مقطعا من مسرحية هاملت لشكسبير وبالأصوات الأربعة المعروفة: السوبرانو، الألتو، الباص، التينور.. وعلينا كممثلين هنا أن نتقيد بقواعد النطق المهمة وهي: مخارج الحروف والسكتات أو (التمبو) وهي من أبرز العناصر في أداء الجمل المسرحية الطويلة، إذ إنها تمثل الفواصل الزمانية في الحوار المسرحي أو أداء المونولوج" [5]، [6]، [7]

وقد حصر فرحان بلبل فن الإلقاء فى غايات ثلاث:

إيصال المعانى التى يقصدها المتكلم، والتعبير عن المشاعر والعواطف التى يتضمنها النص

وأخيرا، كشف جماليات الأسلوب الأدبى للكلام .[8]

ولكن قواعد النطق السليمة، وكشف جماليات الأسلوب الأدبى ليست معيارا على مدى نجاح الممثل فى التعبير عن الانفعالات المختلفة، فهذه القواعد عنصرا ضروريا للممثل، ولكنها ليست الأساس فى توصيل التعبير، وخير مثال على ذلك بعض المسرحيات أو التمثيليات الإذاعية، التى يبرع الممثلون فى استعراض مهاراتهم الصوتية بعيدا عن نقل الصوت للانفعال، فعلى سبيل المثال مسرحية هاملت الإذاعية [9] فالأداء التمثيلى فيها لا ينقل الأنفعالات المختلفة، تحول الأداء الصوتى فيها وكأن الممثلون يقرؤن النص، ولا يؤدونه انفعاليا بالرغم من تمكنهم من قواعد فن الإلقاء.

وهنا تبرز أهمية الإيقاع وارتباطه بتدريب الصوت ، فالإيقاع (التمبو)Tempo يتكون - في المسرحية- من خلال انفعالات للممثلين، وهو سرعة الأداء أثناء الإلقاء أو الغناء أو الحوار أو الموسيقى، ويعرفه الكسندر دين على انه: "سرعة النمط الإيقاعي ويمكن وصفه بأنه سريع وبطيء أو متوسط السرعة، والتغير في التمبو لا يغير بحال في النمط الإيقاعي والأساسي، والمسرحية التي تسير على نبرة واحدة هي مسرحية مملة، رتيبة، ومع ذلك ينبغي ألا تؤدي التنويهات إلى كسر النبرة الأساسية" [10]

وليس المقصود بالإيقاع فى الأداء التمثيلى سرعة الممثل أو بطئه، ولكن إيقاع الشخصية الذى يتناغم مع الهدف الأعلى للمسرحية المقدمة. لذا اهتمت مدارس التمثيل الحديثة والمعاصرة بتدريبات الصوت ومدى ملاءمته للشخصية المؤداه، فضلا عن الاهتمام بخصائص الشخصية الصوتية. فيحدد الممثلون الميزات المرغوبة في الصوت ويوائمون

أصواتهم بناء على هذه الميزات. ووفق ذلك يتحتم على الممثل أن يدرس متطلبات كل مشهد على حدة.وفق الهدف العام من النص المسرحى المقدم.

ويلجأ الممثل فى تحديده للخصائص الصوتية للشخصية إلى الذاكرة الانفعالية، تلك التى تحدد الانفعالات المراد التعبير عنها بواسطة الصوت والجسد.

وتقوم الذاكرة الانفعالية غير الذاكرة الحسية على إحياء المشاعر والانفعالات السابقة، وبعثها من جديد في مواقف درامية حسب سياقها الذهني والنصي، وإعادة المشاعر كما وقعت في الماضي لإسقاطها على مواقف درامية في الحاضر فوق منصة المسرح مشابهة لتلك التي وقعت في الماضي. وفي هذا، يقول ستانسلافسكي:" إن اكتمال تجاربنا الخلاقة كلها وقوتها يتناسبان مع قوة ذاكرتنا الانفعالية، ودقتها، ومضائها تناسبا طرديا.أما إذا كانت ذاكرتنا ضعيفة، فإن المشاعر التي تثيرها تكون باهتة وهزيلة، ولا قوام لها، وتنعدم قيمتها على المنصة؛ لأنها لا تستطيع التأثير على الجماهير التي تجلس فيما وراء الأضواء الأرضية. وثمة درجات كثيرة لقوة الذاكرة الانفعالية؛ وأن كلا من مؤثراتها ومركباتها يختلف اختلافا كبيرا .[11]

ويعني هذا أن الممثل لكي يؤدي دوره بصدق عليه أن يستذكر تجاربه الشخصية التي تعرض لها في الواقع، ويعيشها مرة أخرى، ويتذكر تفاصيلها معايشة ومعاناة. ومن ثم، يعيد تشغيل الذاكرة الانفعالية فوق الخشبة وفق متطلبات الشخصية المسرحية ويعد ذلك من أهم الآليات الفعالة لربط الدور المسرحي بالحياة، وربطه أيضا بالصدق النابع من الذات الحية. فالممثل هنا هو الوسيط بين الشخصية المؤداه، وحياته الواقعية.

وعلى ذلك فالذاكرة الانفعالية مرتبطة ارتباطا بمخزون الخبرات عند الممثل، وهذا ما يجعل الإبداع متفردا، ولكن الممثل فى تلك العملية الإبداعية، يستخدم عناصر خاضعة للمعايير الموضوعية، وقابلة للقياس، مثل قواعد فن الإلقاء، وكذلك شدة الصوت، ودرجته، وحدته...إلخ فإلى أى مدى نستطيع الحكم على مدى صدق الأداء الصوتى للممثل؟

ثانيا - الصوت وعلم هندسة اللغة:

تتناول الباحثة فى هذا القسم الدراسة بعض المفاهيم الأساسية لعلم هندسة اللغة، والتى تعد بالنسبة لعلماء اللغة ومتخصصى علم الصوتيات أمرا بديهيا، وأسسا متعارفا عليها، ولكن ترى الباحثة ضرورة إدراجها هنا بسبب عدم وجود مراجع فى الأداء التمثيلى تحتوى على تلك الموضوعات.

إن اللغة كائن حى يعيش ويتعايش، يؤثر ويتأثر، ويتطور كذلك، وتنقسم الدراسة الصوتية ومناهجها في العصر الحديث إلى عدة فروع، أهمها:

أولا: علم الأصوات phonetic، وهو علم خاص بالدراسة الصوتية البحتة سواء كانت فيزيائية أو أكوستيكية، من صفات الحروف ومخارجها وتشريح جهاز النطق الإنساني وتطور نطق الحروف عبر الزمن في اللغات المختلفة والدراسة الصوتية المقارنة للأصوات والحروف بين لغتين أو أكثر....إلخ، وكل هذه الدراسات يلزمها مناهج بحثية متنوعة ومختلفة، كالمنهج الوصفي، أو التاريخي، أو المقارن، أو التجريبي...

ثانيا: علم وظائف الأصوات phonology وهو علم يختص بوظيفة الأصوات داخل الكلمة ووظيفتها الدلالية، وهو أحد مستويات اللغة وفرع مكمل للنظام الصرفي ونظام تركيب الجملة في اللغة، ومحور دراسته يدور حول الفونيمphoneme وعُرّف الفونيم: بأنه أصغر وحدة لغوية صوتية مجردة تفرق بين كلمة وأخرى، مثل الفرق بين الفعل (ضَرَبَ) فعل مبني للمعلوم، و(ضُرِبَ) فعل مبني للمجهول، عن طريق الضمة والفتحة كفونيميمن يحددان معنى الفعل الثاني ويفرقانه عن نظيره.[12] و[13]

وتجمع الدراسة الحالية بين علم الأصوات، وعلم وظائفها، وربط هذه الفروع بالأداء التمثيلى.

- الهندسة اللغوية: "هى العلم الذى يبحث فى اللغة البشرية كأداة طيعة لمعالجتها فى الآلة" [14] و [15] ويهدف علم الهندسة اللغوية إلى معالجة اللغة حاسوبيا بغرض الوصول إلى برامج تطبيقية تقوم على معالجة جوانب اللغة كافة: مثل تحليل نحوى، تحليل صرفى، تحليل دلالى، ترجمة آلية، تعليم إليكترونى.... وغيرها. ويستفيد المهندسون من الطاقة الصوتيّة بتحويل الصوت إلى إشارات كهربائيّة، وإدخالها إلى أجهزة الكمبيوتر، ثم تحويلها مرة أخرى إلى طاقة صوتيّة يتم الاستماع إليها، كما يمكن استعمال مجموعة من التقنيات تختصّ بتضخيم الصوت، وتعديله، وإضافة أصوات أخرى إليه، وإعادة إنتاجه مرة أخرى.[16]

- تطبيقات هندسة اللغة فى الحياة الواقعية.

امتد مجال هندسة اللغة، وتحليل الصوت باستخدام البرامج الحديثة إلى عدد من المجالات الحياتية، فهناك بعض الشركات التى تستعين بتك البرامج لاختيار موظفيها وترقيتهم؛ حيث يتعرض الموظف أو المتقدم للوظيفة إلى اختبارات لصوته، وتقوم هذه الاختبارات على تحليل تسجيل صوتي لإجابته مدته 15 دقيقه عن طريق الكمبيوتر. ويتضمن ذلك تحليل نغمة الصوت، واختيار الألفاظ، وتركيب الجملة، للتعرف على السمات الشخصية، مثل الاستعداد للتغيير، والحماسة، وتفهّم مشاعر الآخرين، ويلخص برنامج الكمبيوتر المستخدم في ذلك، وفي جزء من الثانية، سماتك الشخصية، من خلال مخططات، ورسوم بيانية، تكشف عن مدى دماثة خلقك، وطموحك الشخصي لتحقيق مكانة أفضل، وقدرتك على التنظيم، مقارنة بالنموذج المثالي الذي وضعته الشركة مسبقا. [17]

كما استخدمت تكنولوجيا تحليل الصوت فى المجال الطبى، فقد تمكن فريق من الباحثين من تطوير برنامج كومبيوتر يعتمد على الذكاء الإصطناعى (AI) قادر على المساعدة فى تشخيص إضطراب ما بعد الصدمة (PTSD) لدى قدامى المحاربين من خلال تحليل أصواتهم.

وتوصلت الدراسة - التى نشرت فى عدد إبريل من مجلة (الإكتئاب والقلق) - إلى أن أداة الذكاء الاصطناعى يمكن أن تميز بدقة تصل إلى 89% بين أصوات من يعانون اضطراب ما بعد الصدمة أو كانوا من الأصحاء". وقد شارك في الدراسة نحو 53 مشاركا يعانون من اضطراب ما بعد الصدمة، ونحو 78 من قدامى المحاربين من الأصحاء، حيث تم ربط برامج الذكاء الإصطناعى بأنماط ميزات صوتية محددة مع اضطراب ما بعد الصدمة، بما فى ذلك الكلام الأقل وضوحا ونبرة هامدة. وفي الوقت الذى لم تكتشف فيه الدراسة الحالية آليات مرض اضطراب ما بعد الصدمة، إلا أن النظرية هى أن الأحداث الصادمة تغير إشارات المخ التى تعالج الإنفعال ونبرة العضلات، والتى تؤثر على صوت الشخص. [18]

كما تم استخدام برامج تحليل الصوت فى مجال حل الجرائم والقضاء، حيث يتم عرض تلك التسجيلات على خبراء الأصوات ، الذين يقومون بأخذ بصمة صوت المتهمين، ومن ثم يضاهون تلك الأصوات بالأصوات الواردة بالتسجيلات، وثبت من خلال عملية المضاهاة تطابق الأصوات، وعليه يتم إعداد تقرير وافً بتلك التفاصيل، ويقدم لجهات التحقيق التى تستند إليه فى توجيه الاتهامات أو نفيها. [19]

وفى المجال الفنى، فقد استخدمت الأجهزة الحديثة فى تحليل صوت المطرب عبد الحليم حافظ، ولقد كشفت عمليات التحليل والقياس ايضا عن مرحلتين متميزتين في حياة صوت العندليب والحد الفاصل بين تلك المرحلتين هي اغنيه (نار يا حبيبي نار)قبلها لم تكن اثار المرض تبدو في التحليلات ثم بدأت هذه الاثار تتزايد ويقل معها انتاج عبد الحليم الغنائي وهذا يكشف عن المعاناة الانسانيه التي كان يمر بها ليبقى مستمرا في طريق الكفاح الفني. [20]

تعددت التطبيقات الحياتية المرتكزة على مجال التحليل العلمى للصوت، وذلك فى عدة مجالات حيوية ودقيقة ولكن لم تجد الباحثة أثرا لاستخدامه فى مجال المسرح، بصفة عامة، والأداء التمثيلى بشكل خاص. وهذا ما يعنى به موضوع الدراسة. وقبل الوصول إلى تحديد برنامج التحليل الصوتى المستخدم فى العينة المختارة فى البحث، لابد لنا من التعرض إلى تعريف الصوت بالإضافة إلى أهم الخصائص الفيزيائية له.

تعريف الصوت:

الصوت هو موجة ميكانيكية تنتقل خلال وسط ما (غاز، سائلا، صلبا)، حيث يصدر الصوت عن جسم معيّن وينتقل عن طريق مجموعة من التضاغطات والتخلخلات إلى المستقبِل، ويستطيع الإنسان أن يميّز هذه الأصوات من خلال عضو السمع لديه وهو الأذن، ويوجد العديد من العوامل التي تؤثّر في قوّة انتقال الصوت، مثل طبيعة الوسط الذي ينتقل فيه وكثافته. [21] و[22]

الخصائص الفيزيائية للصوت:

تتحدد خصائص الصوت من خلال موجته، ويتحكم بها:

أ- طول الموجة wavelength : وهى المسافة بين أية نقطة من الموجة، ونظيرتها فى الطور الذى يليها.

ب- سعة الموجة Amplitude : وهى شدة إشارة الموجة الصوتية، ويستدل عليها فى المنحنى الموجى بارتفاع الموجة، فكلما علت كلما كان الصوت أعلى.

ت- التردد أو التواتر Frequency : هو عدد الموجات التى تتجاوز نقطة معينة خلال فترة زمنية محددة، ووحدة قياسها Hz (موجة فى الثانية). ويتعلق التردد بسرعة اهتزاز مصدر الصوت، فعند زيادته يزداد تردد الصوت الصادر عنه، وكلما زاد تردد الصوت كلما كان أكثر حدة، والعكس صحيح.

أما فيما يتصل بالصوت اللغوى، فإنه يجرى التركيز على عدة أمور، أهمها المدة، التردد، السعة، والبوانى الصوتية.

أ- المدة (s): تعكس الحجم الزمنى الذى يشغله صوت معين حين نطقه.

ب- التردد: وللصوت اللغوى نوعان من التردد؛ تردد البوانى الصوتية الذى يتعلق بتكوين الجهاز الصوتى، والتردد الأساس(F0) المتعلق بالنبضات الفردية الناتجة عن اهتزاز الوترين الصوتيين خلال وحدة زمنية، وهو يعكس النبرة PITCH

ت- السعة db: ويتم تحديدها بقتامة الأشرطة، فكلما زادت شدة طاقة الصوت المعطى فى وقت وتردد معينين، كلما ازدادت قتامته.

ث- البانية الصوتية (النطاق الرنيني) : هى تركيز الطاقة الأكويستيكية حول تردد معين فى موجة الكلام. [23]

تنطلق هذه الدراسة مما يطلق عليه التحليل الدلالى للصوت،

التحليل الدلالى للصوت:

إن الدلالة الصوتية تتحقق عن طريق أحد طريقين:

1- *طريق الأصوات وصفاتها التي تتألف منها الكلمة:*

وعلى سبيل المثال، الفرق بين كلمتى نضح ونضخ، فالنضح للقليل والنضخ لتدفق السائل بقوة وكثرة ومنه قوله تعالى: "فيهما عينان نضاختان"، ومن هنا اقترنت قوة المعنى والدلالة (تدفق الماء بقوة وكثرة) بصوت الخاء على حين اقترنت الدلالة الخفيفة والمعنى الأضعف (تسرب الماء بضعف) بصوت الحاء، والخاء أقوى في النطق من الحاء.

2- *عن طريق الأداء، Intonation :*

للأداء وظائف دلالية متنوعة في كل لغة بشرية حية، وهناك كثير من الدلالات التي نستمدها من طريقة نطق الكلمة أو الجملة حسب السياق الواردة فيه، وهذه الدلالات ليس لها مقابل على مستوى المفردات والجمل، والذي حمل إلينا هذه الدلالات هو الأداء Intonation،وهو رفع الصوت وخفضه أثناء الكلام للدلالة على المعاني المختلفة للجملة أو للكلمة الواحدة، وطريقة نطق الكلمات والجمل يعطينا كثيرا من الدلالات المتنوعة للكلمة أو الجملة الواحدة.

هذا وقد اهتم علم اللغة الحديث بالتغيرات داخل الكلمات نفسها ، وشكلت موضوع علم الصرفMorphology ، والعلم الذي يختص بدراسة تنظيم الكلمات في نسق معين يشكل موضوع علم النحوSyntax، وكلا العلمين يعدان من أهم موضوعات علم اللغة linguistics، الذي يركز على اللغة نفسها، ومؤخرا تم دمج المستوى الصرفي مع المستوى النحوي لتحليل التراكيب اللغوية. [24] و[25] و[26]

تناولت الباحثة فى هذا المبحث الإطار النظرى التأسيسي للبحث، وسوف تنتقل إلى المبحث الثانى الذى يتناول التجربة العملية تفصيلا.

المبحث الثانى: النموذج التطبيقى:

(قياس الانفعال الصوتى لشخصية هاملت باستخدام برنامج برات)

ونعرض فى هذا المبحث خطوات إجراء التجربة العملية بدءًا من أولى خطوات اختيار العينة، والبرنامج المستخدم فى القياس، مرورا بمراحل التجربة كاملة.

أولا: البرنامج المستخدم فى القياس:

يوجد عدد لا نهائي من البرامج التى يتم من خلالها تحليل الصوت، ويطالعنا العلم بصفة مستمرة على إضافة المزيد من الخصائص على البرامج التى تتعامل مع تحليل الصوت، ومن أشهر هذه البرامج برنامج PRAAT الذى يعمل على تعديل الصوت، وفصله، مع تحرير وتعديل على ملفات الصوت وتحليلها. وهذا التطبيق يستخدم فى مجال الصوتيات والفونولوجيا، ولكنه مستخدم على نطاق واسع فى مجالات أخرى تتعلق باللسانيات، مثل علم النفس، والموسيقى، والأنثروبولوجيا.[27]

وباستشارة المتخصصين فى مجال الصوتيات اقترحوا على الباحثة العمل بالبرنامج نفسه ، حيث وجدوا أنه يحقق هدف دراستى، فضلا عن إنه من البرامج الدقيقة فى عملية القياس.

برات وتعنى بالهولندية "الكلام"، وهو برنامج مجانى لتحليل ومعالجة الموجات الصوتية، كتبه ويشرف عليه بول بورسما، وديفيد وينك Paul Boersma and David Weenink

من معهد علوم الصوتيات،بجامعة أمستردام- هولندا[28]

ثانيا: عينة البحث:

قامت الباحثة باختيار نموذج مسرحى يتمثل فى شخصية هاملت [29]من عدد من العروض المختلفة لممثلين مختلفين .

- تمثل النموذج الأساسي فى أداء الممثل "لورانس أوليفيه Laurence Kerr Olivier" [30] (1907-1989) لمونولوج الكينونة باعتباره نموذجا للقياس. وذلك للأسباب الأتية:

أولا: ليس خافيا على المسرحيين بشكل عام، ودارسي فن التمثيل أن الممثل لورانس أوليفيه هو أفضل الممثلين على الإطلاق، وقيل عنه أنه أفضل ممثل فى العالم الانجليزى فى القرن العشرين".

ثانيا: حصل على جائزة الأوسكار كأفضل ممثل عن دوره فى فيلم هاملت –(1948)، عينة البحث.

ثالثا: كان وراء إحياء أعمال شكسبير التاريخيه، وبادر بتقديمها على شاشات السينما.

- تنوع الاختيار بين المسرح والسينما، حيث إن الهدف هنا هو قياس الانفعال فحسب، وقد تأكدت الباحثة أن كل ما تم تقديمه لم يكنب معالجات أو رؤي درامية مختلفة عن رؤية شكسبير نفسه.

أما النماذج الأخرى التى سيتم تقييم أدائها الصوتى فهى كالتالى:

- أدريان ليستر[31] Adrian Lester هاملت (2016)
- ميل جيبسون[32] Mel Gebson هاملت (1990)
- دافيد تينانت[33] Daived tenant هاملت (2010)
- أندرو سكوت [34] Andrew Scotte هاملت (2018)
- كينيث براناه [35] Kenneth Branagh هاملت (1996)

تم اختيار الجملة الأساسية من مونولوج (الكينونة) -هكذا يطلق المسرحيون عليه-. فهو المونولوج الذى يستهله هاملت بجملة (أكون أو لا أكون...تلك هى المشكلة (To be or not to be..that is the question)، وفى هذه الجملة تتجسد أعلى مستويات الصراع الداخلى للشخصية، ويتبدى القلق والحيرة، كما يظهر كامل ضعف شخصية هاملت. تم تقسيم عينة البحث وفق الآتى:

أ- لورانس أولييفيه: وهو الأساس الذى يتم القياس عليه وسوف يرمز له بحرف R Reference .
بالإضافة إلى خمسة نماذج أخرى أربع منها مسرحية و الخامس سنيمائية. وهم:
1- كينيث براناه وسوف نرمز له بحرف K
2- أندرو سكوت ونرمز له بحرف S
3- أدريان ليستر ونرمز له بحرف L

4- دافيد تانينت ونرمز له بحرف D
5- ميل جيبسون ونرمز له بحرف M

ب - سوف يتم قياس أربعة عناصر رئيسة باستخدام برنامج برات وهى:
أولا: مدة الصوت Duration
ثانيا: الدرجة الصوتية pitch
ثالثا: شدة الصوت intensity
رابعا: النبر intonation
ثانيا: خطوات إجراء القياس:

أولا: مدة الصوت:

تم تقسيم الجملة إلى ثلاث مقاطع رئيسة ، تتخللهم فترتى صمت وذلك وفق طريقة تقسيم الجملة عند لورانس أوليفيه (أساس القياس)

1- To be
2- Silence
3- Or not to be
4- Silence
5- That is the question

ثانيا : الدرجة الصوتية:

وتم قياس الذبذبات الصوتية للنماذج كافة، وقد تم قياس الفرق بين أعلى تردد وأقل تردد لكل نموذج. ، فمن المتعارف عليه أن الذبذبة الصوتية تتوقف على طول وحجم الطيتين الصوتيتين، وبعد الأذن عن مصدر الصوت، وكذلك على جنس المتكلم سواء رجل أم امرأة . وكلما قلت الاهتزازات يعنى ذلك أن الصوت أكثر غلظة. والعكس صحيح. أى أنها تتناسب تناسبا عكسيا مع الصوت.

ويظهر فى الجدول رقم (1) قياس الذبذبات الصوتية لكل النماذج، وكذلك الفرق بين أقل ذبذبة وأعلى ذبذبة:

جدول رقم (1)

قياس الفرق بين أقل ذبذبة وأعلى ذبذبة

SIGNAL	MEAN PITCH	MINIMUM	MAXIMUM
Reference signal	99.96204072865116Hz	75.61388686444576 Hz	277.6546165681976 Hz
K	141.303843207913 Hz	96.37671077085244 Hz	415.5544620362446 Hz
S	136.49406993824252 Hz	91.13895729860967 Hz	490.8729153747852 Hz
L	92.79184993398123 Hz	75.412602429521 Hz	400.3779727431688 Hz
D	114.58316061493304 Hz	91.69784472234956 Hz	146.88185632767784 Hz
M	109.91871532544411 Hz	74.9046198220042 Hz	100.6300021207116 Hz

يتضح من الجدول أن أصوات كل من كينيث براناه K وكان قياسه 141 هرتز، وأندرو سكوت S 136 هرتز، ودايفيد تانينت D 114 هرتز وميل جيبسون M 109 هرتز وهو يعتبر وفق القياس من الأصوات المتوسطة فى الرجال والتى تتراوح بين (109، 163) هرتز عند الرجال ، وتمثلت الأصوات الغليظة فى صوت كل من أدريان ليستر L وكان قياسه 92 هرتز ، ولورانس أوليفيه R 99.9 هرتز ويفسر ذلك بكونه من الطبقات الصوتية الغليظة. وبهذا تراوحت الأصوات بين المتوسطة والغليظة.

ثالثا: قياس شدة الصوت :

ويعنى بها شدة إشارة الموجة الصوتية، فكلما ازدادت القيمة كلما كان الصوت أعلى. أى أن شدة الصوت تتناسب طرديا مع الصوت. ويجمل الجدول الآتى (جدول رقم 2) القياسات الثلاث السابقة (الفترة الزمنية)، ودرجة اهتزاز الصوت، وشدة الصوت:

Reference	To be	Silence	Or not to be	**Silence**	That is the question
Duration	0.6266 sec	2.1576 sec	1.2338 sec	**2.3255**	2.0219
Pitch	108.42067787776764 Hz		97.31674818860985 Hz		96.52632499138801
intensity	57.201977896702516 dB		55.40933660833833 dB		52.63798422597694 dB
K	To be	Silence	Or not to be	**Silence 3**	That is the question
Duration	0.76 sec	1.79 sec	1.67 sec	**2.24 sec**	1.25 sec
Pitch	141.22745914808908 Hz		128.70262570556983 Hz		404.4665324679772 Hz
intensity	62.229461836531314 dB		51.950588301240984 dB		46.801405836345864 dB
S	To be	Silence	Or not to be	**Silence 3**	That is the question
Duration	0.855 sec	4.8 sec	2.25	**3.36 sec**	0.984
Pitch	133.66330449110515 Hz		122.8685620201669 Hz		162.5718191367407 Hz
intensity	58.992758753470234 dB		53.91157454139703 dB		53.093383257231906 dB
L	To be	Silence	Or not to be	**Silence 3**	That is the question
Duration	1.3		1.49	**2.2**	1.2
Pitch	87.13514771064078 Hz		83.71691000877189 Hz		89.28264695435283 Hz

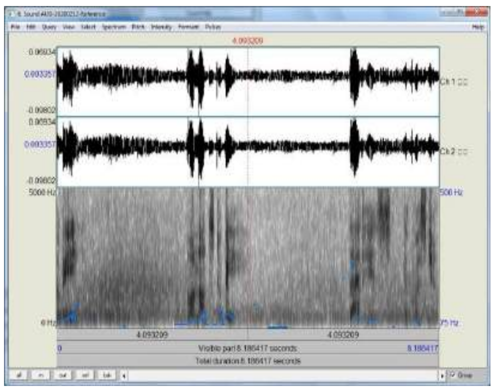
Intensity	62.15333684856982 dB		59.330048871770245 dB		58.574580803824425 dB
D	To be Or not to be			**Silence 2**	That is the question
Duration	1.6 sec			**1.84**	1.05
Pitch	109.91871532544411 Hz				122.73726514654592 Hz
intensity	65.67681964448114 dB				65.8575646610382 dB
M	**To be Or not to be That is the question**				
Duration	2.11 sec				
Pitch	87.06416826156647 Hz				
intensity	54.00598655525319 dB				

جدول رقم (2)

يوضح الفترة الزمنية، ودرجة الصوت وشدته

ومن المتعارف عليه أن شدة الصوت (القوة أو الضعف) يقصد بها مقدار ضخامة الصوت، واندفاعه خارج الفم بغض النظر عن طبقته متوسطة أو غليظة أو حادة. و بذلك فهى تنشأ من مقدار الرنين وقوة دفع الصوت إلى الخارج. وفيما يأتى توضح الأشكال الأتية قياس شدة الصوت لكل من النماذج المختارة.

أولا: النموذج الأساسى: R

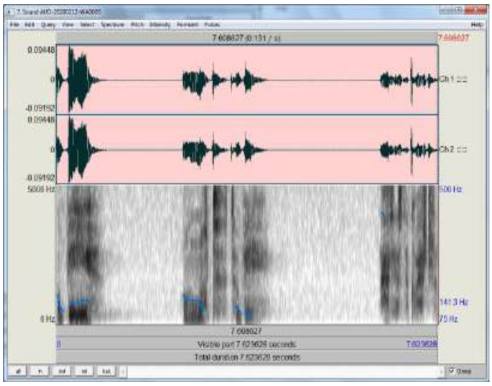


شكل رقم (1)

قياس شدة الصوت عند لورانس أوليفيه

يتضح أنه: فى نموذج القياسR لورانس أوليفيه قد تراوحت شدة الصوت بين 57 ديسيبل فى المقطع الأول، فقد بلغ أعلى درجاته، وانتهى بـ 52 ديسيبل فى المقطع الأخير . وهو ما يتناسب مع طريقة الأداء الصوتى لتساؤل هاملت (To be or not to be that is the question) (to be) هى التى لابد أن تكون الشدة فى هذه الجملة من حيث أنه يريد أن يكون، وقد اتخذ فترة صمت قيمتها 2.1 ثانية لينتقل إلى المقطع الثانى من الجملة، أما كلمة (or not to be) فأتت شدتها أقل،وقد بلغت 55 ديسيبل ولزم له فترة صمت أخرى 2.3 ثانية تمهيدا للانخفاض التدريجى فى المقطع الأخير (that is the question) ، ومن ثم تأتى أقل شدة من سابقتها من الجمل. وتعبر الشدة فى كلمة دون غيرها إحدى وسائل الوصول إلى إبراز المعنى المراد توصيله، كما أنها – وفق قواعد فن الإلقاء- تمكن الممثل من إبراز الحالة الفكرية والنفسية لشخصية المتكلم. كما أن لفترات الصمت أهميتها فى إبراز المعنى المراد توصيله، والتركيز عليه دون غيره بشكل يمنع الالتباس، ويحدد إيقاع الجملة. وبناء على ذلك فلقد جاء نموذج القياس R مستخدما لقواعد الإلقاء كافة (الارتفاع والانخفاض، فتراة الصمت، الإيقاع)

2- كينيث براناه K



شكل رقم(2)

شدة الصوت عند كينيث براناه

فى الشكل السابق، عند k، قد تراوحت شدة الصوت بين 62 ديسيبل فى المقطع الأول، فقد بلغ أعلى درجاته، وانتهى بـ 46.8 ديسيبل فى المقطع الأخير . وقد اتخذ فترة صمت قيمتها 1.7 ثانية لينتقل إلى المقطع الثانى من الجملة، أما كلمة (or not to be) فأتت شدتها أقل،وقد بلغت 51.9 ديسيبل ولزم له فترة صمت أخرى 2.2 ثانية تمهيدا للانخفاض التدريجى فى المقطع الأخير (that is the question) ، ومن ثم تأتى أقل شدة من سابقتها من الجمل.

3- أدريان ليستر L

شكل رقم(3)

شدة الصوت عند أدريان ليستر

فى الشكل السابق، عند L، قد تراوحت شدة الصوت بين 62 ديسيبل فى المقطع الأول، فقد بلغ أعلى درجاته، مثل كينيث براناه تماما وانتهى بـ 58.5 ديسيبل فى المقطع الأخير . وقد اتخذ فترة صمت قيمتها ثانية واحدة لينتقل إلى المقطع الثانى من الجملة، أما كلمة (or not to be) فأتت شدتها أقل،وقد بلغت 59.1 ديسيبل ولزم له فترة صمت أخرى 2.2 ثانية تمهيدا للانخفاض التدريجى فى المقطع الأخير (that is the question) ، ومن ثم تأتى أقل شدة من سابقتها من الجمل.

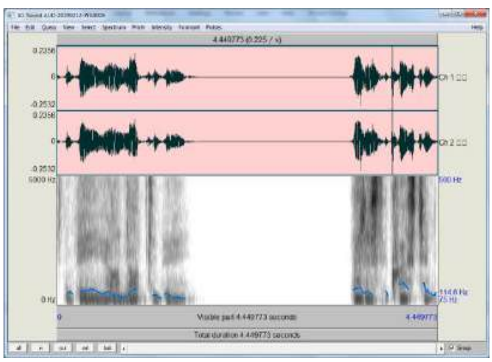
4- أندرو سكوت S

شكل رقم(4)

قياس شدة الصوت عند أندرو سكوت

فى الشكل السابق، عند S، قد تراوحت شدة الصوت بين 58.9 ديسيبل فى المقطع الأول، فقد بلغ أعلى درجاته، . وقد اتخذ فترة صمت قيمتها 4.8 ثانية لينتقل إلى المقطع الثانى من الجملة فى كلمة (or not to be) فأتت شدتها أقل،وقد بلغت 53.9 ديسيبل ولزم له فترة صمت أخرى 3.3 ثانية ، أما شدة المقطع الأخير (that is the question) لم تختلف عن شده المقطع الثانى حيث بلغت 53 ديسيبل. لقد سجلت فترات الصمت أعلى قياسا لها فقد تعدت الفترة الزمنية للجملة المنطوقة، مما قد يصيب المشاهد بالملل، خاصة وأن شدة المقطعين الأخيرين لم يحدث اختلاف بيهما، فلا يوجد تنويع فيما بينهما.

5-دايفيد تانينت D

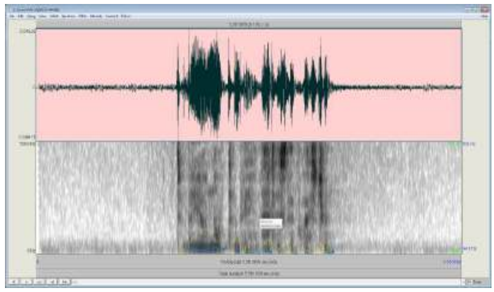


شكل رقم (5)

شدة الصوت عند دايفيد تانينت

قسم تانينت الجملة إلى مقطعين صوتيين فقط الأول (To be or not to be) ثم المقطع الثانى (that is the question) وجاءت شدة الصوت فى كلا المقطعين واحدا بلغ 65 ديسيبل. مما يدل على تساوى شده كل الجملتين، وهو ما يتنافى مع التنويع الصوتى المطلوب للأداء. كما أنه فصل بين المقطعين بفاصل زمنى مدته 1.84 ثانية، إلا أنه –كما ذكرنا- لم ينوع فى طبقة الصوت بين كل من المقطعين.

6- ميل جيبسون M



شكل رقم(6)

شدة الصوت عند ميل جيبسون

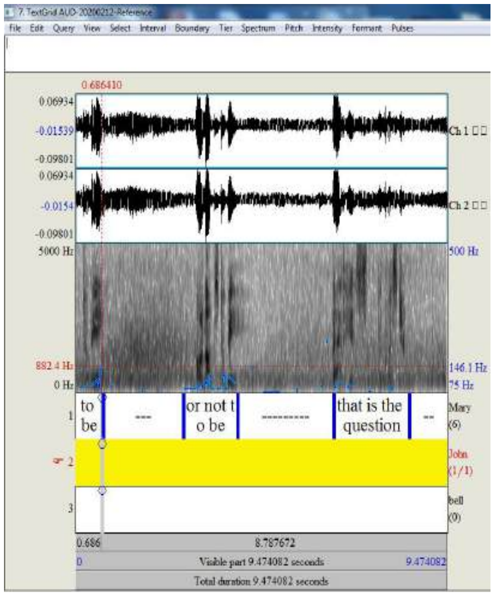
أما فيما يتعلق بشدة الصوت عند ميل جيبسون فلم يقم بتقسيم الجملة، وإنما قام بأداءها جملة واحدة، وجاء قياسها 54 ديسيبل، وهو كذلك يتنافى مع المعنى التنويع المراد توصيله، فكل مقطع من الجملة لابد أن يتم التركيز فيه على حرف، أو تمييز كلمة عن أخرى.

ويظل مقياس شدة الصوت ليس معيارا للحكم على طريقة الأداء ولكنه مرحلة أولية لابد من القيام بها، ومن ثم وجب القيام بقياس النبر ارتكازا على قياس شدة الصوت.

رابعا: قياس النبر:

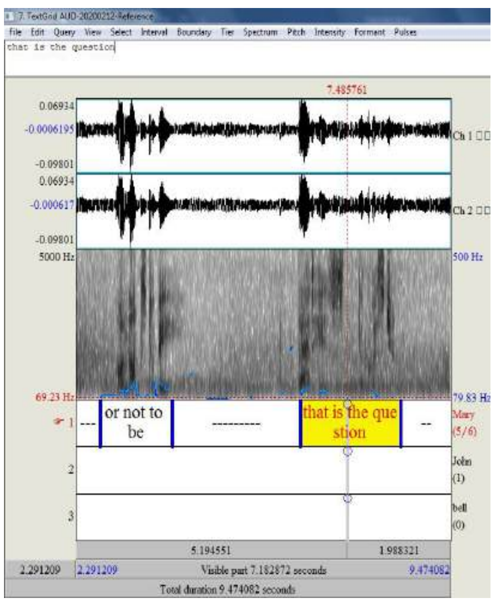
وتبنى هذه المرحلة على المراحل السابقة، ويتحدد عن طريقه التنغيم intonation ، "فالتنغيم هو من الفونيمات فوق التركيبية أو الإضافية التى تصاحب نطقنا للكلمات والجمل، ويرتبط الارتفاع والانخفاض بتذبذب الوترين الصوتيين الذان يحدثان النغمة الموسيقية" ، وفى اللغة إجمالا يمكن تقسيم وظائف التنغيم إلى ثلاث وظائف : أولهما وظيفة نحوية تحدد الإثبات أو النفى فى الجملة، ووظيفة صوتية ويعنى بها خلق نسق صوتى مميز، تشترك فيه النغمة والمسار اللحنى فى تحديد موسيقى الكلام، وثالثا: وظيفة دلالية، أى الكشف عن المعنى والمشاعر من خلال علو الصوت وانخفاضه." [37] وينتج عنه بالضرورة- فى الأداء التمثيلى- إيصال المعنى للسامع، وذلك عن طريق النبر على كلمات محددة دون غيرها فى المقطع الصوتى وفيها يتم تحديد أعلى درجة للنبر، وأقل درجة له فى كل مقطع صوتى، وذلك لمعرفة مناطق التنغيم فى المقطع، ومدى توافقه مع النموذج القياسى من عدمه. وسوف تعرض الباحثة قياسات لكل نموذج من النماذج المختارة.

أولا: نموذج القياس R لورانس أوليفييه:



شكل (7)

قياس أعلى نبر فى نموذج القياس



شكل (8)

قياس أدنى نبر فى نموذج القياس

يظهر الشكل أنه على مدار الجملة كلها فإن أعلى قمة للنبر peak of intonation كانت في مقطع "to be" حيث يمكن قياس head of intonation عند 882.4 HZ

كما أن فإن أدنى قيمة للنبر في الشكل الثاني لنفس الجملة كانت في مقطع "that is the question" حيث يمكن قياس intonationعند 69.23HZ

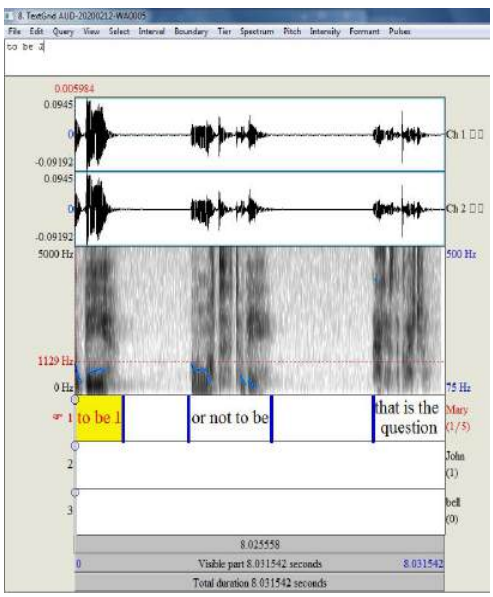
	To be	Or not to be	That is the question
أعلى نبر	HZ 882.4	660.7 Hz	414.2 HZ
أقل نبر	192.4 HZ	93.87 HZ	69.23HZ

جدول رقم(3) قياس النبر عند لورانس أوليفييه

وبقياس منحنى النبر للجملة فى إجمالى الفترة الزمنية وهى 9.4 ثانية، وجدنا أن أعلى نبر جاء فى كلمة(to be) حيث بلغ 882.4 هرتز، وجاء أقل نبر فى مقطع (that is the question) حيث سجل 69.2 هرتز. ومن ثم فإن التنغيم هنا إذا عبرنا عنه بمنحنى، فيظهر أعلى درجة له فى المقطع الأول، وأدنى درجة فى المقطع الأخير. وهذا يعبر بصفة أساسية على أن الكلام انتهى؛ فالانعطاف الصوتى يقصد به الاتمرار أو عدم الاستمرار بالصوت حتى نهاية الكلمة أو الجملة، فإذا كان الاستمرار بالصوت بالاتجاه الصاعد، يعنى أن للكلام بقية حتى لو توقفنا بعد ذلك، وإذا اتجهنا بالصوت إلى النزول فذلك يعنى أن ليس للكلام بقية، وأن الفكرة قد انتهت.[38] ويتمثل منحنى النبر هنا فى (صاعد-هابط-هابط) مما يدل على اكتمال الفكرة.

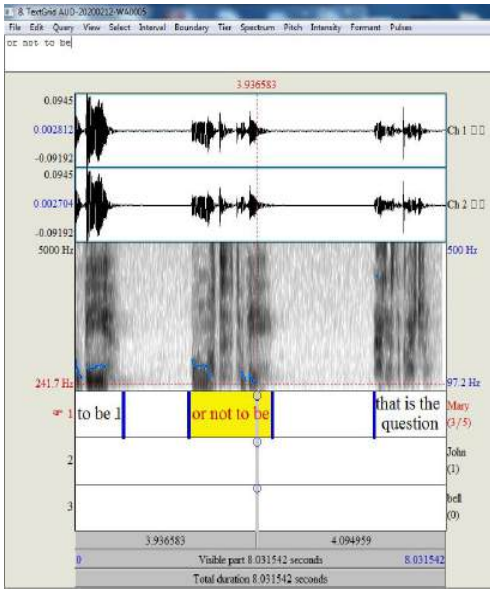
وهو ما يتناسب مع طريقة الأداء الصوتى لتساؤل هاملت (To be or not to be that is the question) ففى هذه الجملة لابد أن تتضح ملامح القلق والحيرة، والغوص فى النفس، والتساؤلات الحائرة، ففكرة كينونته فى كلمة (to be) هى التى لابد أن يكون التركيز علي هذه الجملة من حيث أنه يريد أن يكون وأن يعيش، وقد اتخذ فترة صمت قيمتها 2.1 ثانية لينتقل إلى المقطع الثانى من الجملة، أما كلمة (or not to be) فأتت شدتها أقل، وهو ما يتسق مع حيرته، وخوفه من كونه (لا يكون) ولزم له فترة صمت أخرى 2.3 ثانية تمهيدا للانخفاض التدريجى فى المقطع الأخير (that is the question) ليجسد المشكلة الحقيقية التى يعانى منها هاملت، وخوفه الشديد من تلك الحياة، ومن أفعاله التى يقدم عليها ، ومن ثم فإنه يغوص أكثر داخل نفسه وكذلك يشعر بالرعب الداخلى لما هو مقدم عليه، ومن ثم تأتى أقل شدة من سابقتها من الجمل. ووفقا للشكلين السابقين، فإن التنغيم يتكون فى الأساس من: شدة الصوت (التى انخفضت تدريجيا) ، ثم اختلاف درجة الصوت (التى تباينت بين المقاطع الثلاث من الجملة) هذا التنغيم قد أعطى مضمونا للمعنى المراد التعبير عنه.

2 - كينيث براناه:



شكل رقم (9)

قياس أعلى نبر عند كينيث براناه



شكل رقم (10)

قياس أدنى نبر عند كينيث براناه

يظهر الشكل أنه على مدار الجملة كلها فإن أعلى قمة للنبر peak of intonation كانت في مقطع "to be" حيث يمكن قياس head of intonationعند 1129HZولكن في بداية المقطع كما أنه، فإن أدنى قيمة للنبرفي الشكل الثاني لنفس الجملة كانت في مقطع "or not to be" حيث يمكن قياس النبر عند 241.7HZ. بينما القطع الأخير لا يكاد يظهر له نبر من ضعفه . فمنحنى النبر هنا (صاعد –هابط- لا قياس)

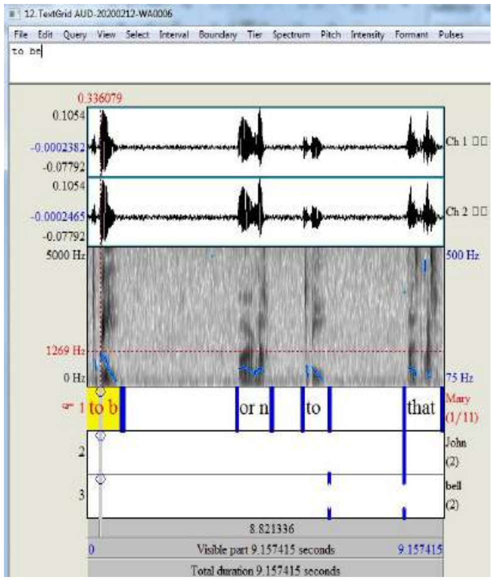
That is the question	Or not to be	To be	
---	241.7HZ	1129HZ	**أعلى نبر**
----	HZ 266.4	HZ 398.6	**أقل نبر**

جدول رقم(4) قياس النبر عند كينيث براناه

وبناء على ذلك ففى اجمالى الفترة الزمنية 8.03 ثانية نجد أن كينيث براناه قد اتفق مع لورانس أوليفيه فى تسجيل أعلى نبر فى المقطع الأول، حيث بلغ 1129 هرتز. ولكنه اختلف معه فى باقى مقاطع الجملة. فالتساؤل الحائر فى نهاية الجملة لم يظهر، وهذا يعنى عدم اكتمال إيصال معنى هذا التساؤل للمستمع.

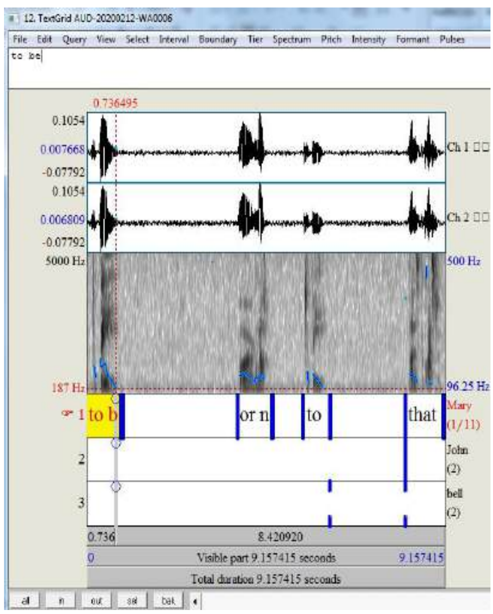
3- أندرو سكوت:

يظهر في الشكل أنه على مدار الجملة كلها فإن أعلى قمة للنبر peak of intonation كانت في مقطع "to be" حيث يمكن قياس head of intonationعند 1269HZ



شكل رقم (11)

أعلى نبر عند أندرو سكوت



شكل رقم (12)

أقل نبر عند أندور سكوت

يظهر في الشكل أنه على مدار الجملة كلها فإن أعلى قمة للنبر peak of intonation كانت في مقطع "to be" حيث يمكن قياس head of intonationعند 1269HZ

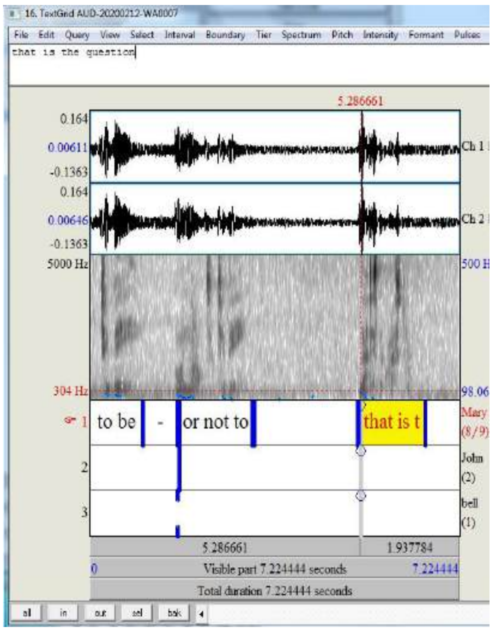
كما أن فإن أدنى قيمة للنبر في الشكل الثاني لنفس الجملة كانت في مقطع "to be" حيث يمكن قياس intonationعند 187HZ

	To be	Or not	to be	That is the question
أعلى نبر	1269 HZ	742.6 Hz	1014 Hz	760.7 HZ
أقل نبر	187 HZ	420.9 HZ	279.8	298.7HZ

جدول رقم(5) قياس النبر عند أندرو سكوت

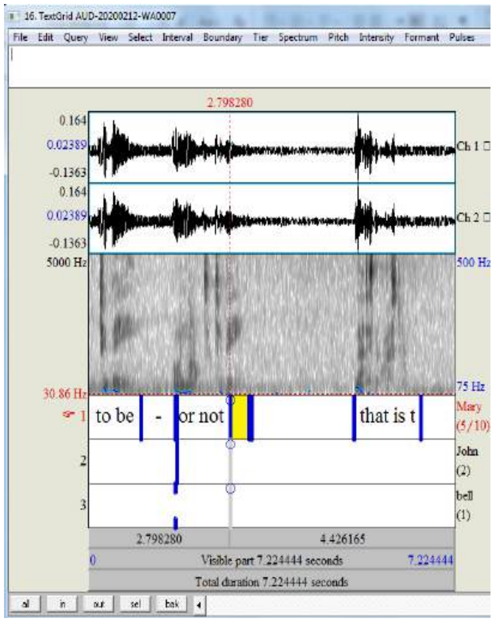
ومن خلال ذلك ففى اجمالى فترة زمنية 9.1 ثانية، نجد أن أندرو سكوت قد اتفق مع نموذج القياس فى تسجيل أعلى نبر فى المقطع الأول حيث بلغ 1269 هرتز ولكنه اختلف عنه فى نهاية المقطع نفسه حيث سجل 187 هرتز، وبتتبع باقى الجملة نجد أن منحنى التنغيم يتركز فى كلمة to be، سواء فى المقطع الأول أو المقطع الثانى. مما يعطيهما القدر نفسه من الأهمية، ومنحنى التنغيم هنا (صاعد-هابط-صاعد-هابط) كما كانت فترات الصمت مبالغ فيها، خاصة بعد المقطع الأول دون مبرر واضح.

4- أدريان ليستر:



شكل رقم (13)

قياس أعلى نبر عند أدريان ليستر



شكل رقم (14)

قياس أقل نبر عند أدريان ليستر

يظهر الشكل أنه على مدار الجملة كلها فإن أعلى قمة للنبر peak of intonation كانت في مقطع "that is the question" حيث يمكن قياس head of intonation عند 304HZ

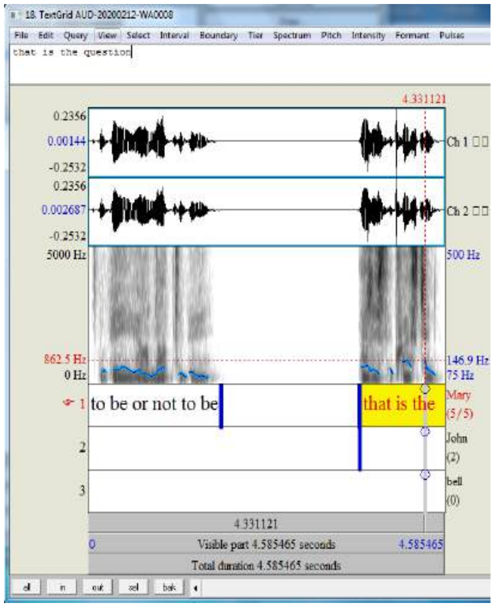
كما أن فإن أدنى قيمة للنبرفي الشكل الثاني لنفس الجملة كانت في مقطع "or not to be" حيث يمكن قياس intonation عند 30.86HZ. وجاء منحنى النبر (صاعد- هابط- أكثر صعودا)

That is the question	Or not to be	To be	
304	209.1	298.2	**أعلى نبر**
60.56	30.86	90.26	**أقل نبر**

جدول رقم(6) قياس النبر عند أدريان ليستر

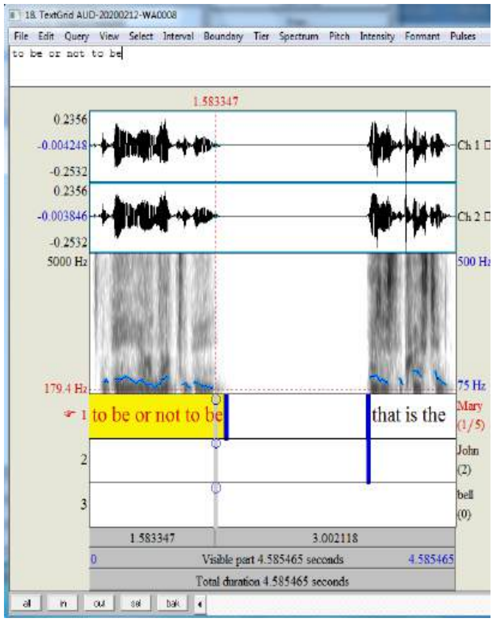
إجمالى الفترة الزمنية 7.2 ثانية، نجد أن أدريان ليستر قد اختلف تماما عن لورانس أوليفيه، حيث سجل أعلى نبر عنده فى المقطع الأخير من الجملة وقد بلغ 304 هرتز. وكان أقل نبر فى الجملة فى المقطع الثانى or not to be حيث سجل 30.8 هرتز. وهنا أدريان ليستر لم يعطى أهمية لفعل الكينونة، وكأن المعنى المراد توصيله هنا هو التركيز على التساؤل، وليس الفعل. وهذا بدوره يختلف عن المعنى الأصلى المراد توصيله. كما يدل على عدم اكتمال المعنى المراد توصيله، فالنهايات الصاعدة – كما ذكرت الباحثة آنفا- تعنى أن معنى الجملة لم يكتمل. ولابد من استكماله بكلمات أخرى.

5-دافيد تانينت:



شكل رقم (15)

قياس أعلى نبر عند دايفيد تانينت



شكل رقم (16)

قياس أدنى نبر عند دايفيد تانينت

يظهر الشكل أنه على مدار الجملة كلها فإن أعلى قمة للنبر peak of intonation كانت في مقطع "that is the question" حيث يمكن قياس head of intonation عند 862.5HZ

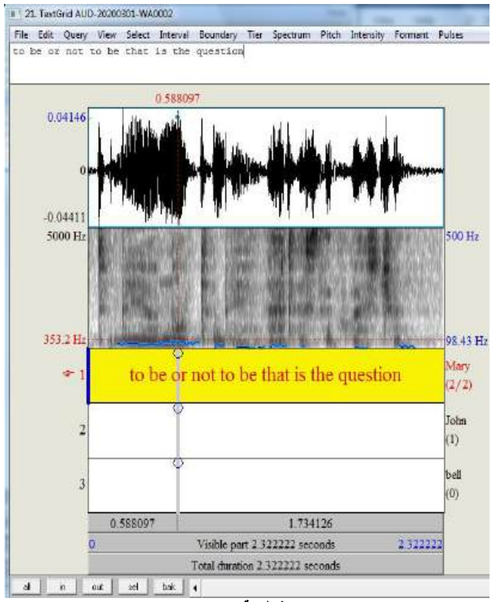
كما أن فإن أدنى قيمة للنبرفي الشكل الثاني لنفس الجملة كانت في مقطع "to be or not to be" حيث يمكن قياس intonation عند 179.4HZ. ومنحنى النبر هنا (هابط - صاعد)

	To beOr not to be	That is the question
أعلى نبر	654.5	862.5
أقل نبر	179.4	357.6

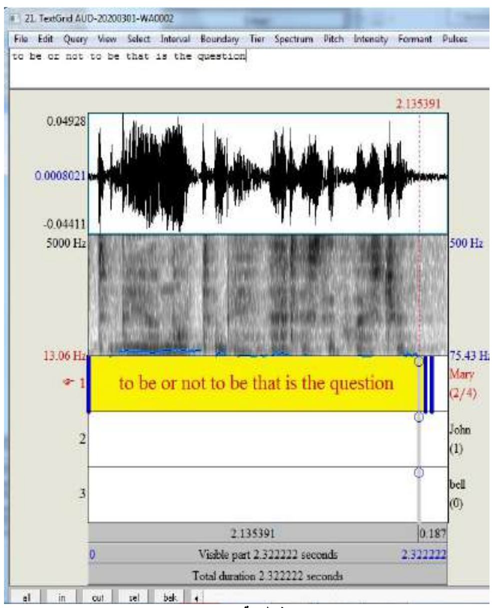
جدول رقم(7) قياس النبر عند دايفيد تانينت

ومن خلال القياس السابق، الذى سجل تقريبا نصف الفترة الزمنية التى وجدت فى نموذج القياس، فإجمالى الفترة الزمنية هنا 4.5 ثانية، وقد قسمت الجملة إلى مقطعين هما to be or not to beK و that is the question. وقد سجل دايفيد تانينت أعلى نبر فى المقطع الثانى من الجملة، وهى التساؤل حيث بلغ 862.5 هرتز، وجاء أقل نبر فى المقطه الأول حيث سجل 179.4 هرتز. أى أن منحنى التنغيم عنده سار عكس ما قام به النموذج القياسى فقد بدأ منخفضا، ثم ارتفع بشكل مفاجىء فى المقطع الثانى، مما يدل على عدم اكتمال المعنى، فهل هاملت هنا كان غارقا فى التفكير؟ أم أن دلالة ذلك هو استيقاظه فجأة وكأنه وجد إجابة الحيرة لديه؟ بتحليل المونولوج الأساسي كما كتبه شكسبير، فإن هذه الجملة دون غيرها لابد أن تظهر مدى عمق التفكير، والاستغراق فيه، ومن ثم تظهر مدى الضعف الذى يعانى منه هاملت. لذا يمكننا القول أن القياس هنا أظهر معنى آخر للجملة ، وهو مالا يتفق مع الهدف الأساسى من الجملة الأساسية.

7- ميل جيبسون:



شكل رقم (17)
قياس أعلى نبر عند ميل جيبسون



شكل رقم (18)
قياس أقل نبر عند ميل جيبسون

تظهر الجملة في تتابع واحد دون فصل وأعلى نبر 353.2 HZ وأدنى قيمة له عند النهاية بلغت 13.06 HZ .

سجلت الفترة الزمنية هنا أقل فترة زمنية للجملة، حيث استغرقت 2.3 ثانية، سجلت أعلى نبر فى بدايتها حيث بلغ 353.2 هرتز، وأقل نبر فى نهايتها 13.06 هرتز. وقد اختلفت من حيث الشكل أو المضمون عن الجملة الأساسية، فإيقاعها جاء متسارعا، ولم يرتكز النبر على جزء أو مقطع دون غيره، ومن ثم فإن دلالة ذلك أن المعنى الذى تم توصيله، لا ينطوى على

تأمل، أو استغراق فى التفكير. فلم يهدف ميل جيبسون هنا إلى طرح المشكلة على السامع، وإشراكه فيها. ولكن يبدو أنه قد وصل إلى النتيجة مسبقا.

أهم النتائج والتوصيات

وبناء على ما سبق يمكننا القول إن أهم نتائج القياسات جاءت كما يلى:

أولا:إمكانية إخضاع الانفعالات الصوتية للممثل للقياس، بل خلصت الباحثة أن ذلك يعتبر ضروريا سواء فى عملية اختيار الممثل لشخصية بعينها، أو التدريب عليها، وصولا لمرحلة العرض.

ثانيا: تراوحت تصنيف الأصوات ما بين الأصوات المتوسطة والغليظة. ولإذا كان هذا المعيار ليس حاكما فى قياس الانفعالات المنوط بالممثل تأديتها، إلا أنه معيار مهم لفى ختيار الممثلين لشخصيات بعينها، فشخصية هاملت وفق الرؤية الكلاسيكية لا ينبغى أن يؤديها ممثل ذو طبقة صوتية حادة؛ حيث لا يتناسب ذلك مع كونه يمثل الأمير الرومانسى الحالم (النموذج المثالى للأمير وفق مقاييس العصر الرومانسى وتحقيقا لمواصفات البطل التراجيدى) .

ثالثا: فيما يتعلق بقياس شدة الصوت إلى أن كل من كينيث براناه K، و أندرو سكوت S، وأدريان ليستر S يتشابهون مع لورانس أوليفيه R، ليس بنفس درجة قوة الصوت، ولكن فى طريقة التناول. فكل منهم قد بدأ عاليا فى المقطع الأول، ثم انخفض تدريجيا حتى وصل إلى أقل شدة فى المقطع الأخير. وقد اختلف عنهم كل من دايفيد تانينت D الذى لم يستخدم التنويع إلا فى فترة الصمت، وجاءت ا لجملة بنفس درجة القوة أما ميل جيبسون M فجاءت شدة الصوت ثابتة فى الجملة كاملة.

رابعا: اختلفت قياسات النبر لكل نموذج من النماذج المختارة عن نموذج القياس، مما أدى ذلك بدوره إلى تغيير المعنى المراد توصيله ، فقياس منحنى النبر عند كينيث براناه أثبت إخفاقه فى إيصال معنى مكتمل للجملة، حيث أنه بالقياسات العلمية اختفى النبر تماما فى المقطع الثالث للجملة، مما يعنى ذلك عدم اكتمال معنى للجملة. أما أندرو سكوت فبالرغم من تنوع شدة صوته، إلا أنها لم تأت متوافقة مع النموذج الأساسى، فقد سجل أعلى نبر فى مقطعى To be (الأول والثانى) بالرغم من تناقض المعنى بينهما فالأولى كانت للإثبات، والثانية للنفى ومن ثم من خلال القياس لم يظهر هذا الاختلاف. وتأتى قياسات أدريان ليستر معاكسة تماما لنموذج القياس، حيث بدأ هابطا، ثم تصاعد تدريجيا وانتهت الجملة بمنحنى صاعد، مما يعنى ذلك أن تركيز الممثل لم يكن فى إبراز فعل الكينونة، ولكن فى التساؤل فى نهاية الجملة، الأمر الذى لابد أن يستتبعه توضيح حيث إن المعنى لم يكتمل. ويتكرر الأمر نفسه عند دايفيد تينانت حيث انتهى صاعدا . أما النموذج الأخير ميل جيبسون فقد ابتعد تماما عن التنوع فى الشدة أو التركيز على كلمة بعينها، وجاء نطق الجملة على وتيرة واحدة.

خامسا: تنوع استخدام فترات الصمت للتمهيد للانفعال التالى، صعودا أم هبوطا.

سادسا: العمل ببرامج التحليل الصوتى للممثلين هو ضرورة لا غنى عنها، حيث تخضع عملية التقييم إلى أسس علمية معيارية ينبغى القياس عليها، فالانطباع الذى يخلقه الصوت فى نفس المدرب هو معيار لا يمكن توحيده، أو القياس عليه.

التوصيات:

1- ضروة استخدام الوسائل العلمية لقياس الأداء الصوتى للممثل، للحكم على طريقة التدريب أو الآداء.
2- ضرورة تزويد قاعات التدريب التمثيلى بفنى صوتيات لمتابعة عملية التقدم فى التدريبات الصوتية .
3- العمل على توجيه مدربى التمثيل فى الأكاديميات المختلفة لعمل بحوث ودراسات معملية للصوت وتدريبه بالاشتراك مع أقسام الصوتيات فى الجامعات المختلفة.
4- نشر المعرفة لدى مدربى التمثيل ببرامج تحليل الصوت الحديثة، كخطوة للانطلاق فى استخدامها فى تحليل الصوت.

المصادر والمراجع

1- *أرسطو طاليس "فن الشعر" ترجمة: شكرى محمد عياد (القاهرة، دار الكتاب العربى للطباعة والنشر، 1967) ص ص 19 :20*
2- *موريس فيشمان "تدريب الممثل" ترجمة: نور الدين مصطفى (القاهرة، الدار المصرية للتأليف والنشر، ب ت) ص 12*
3- *عطا درغام "فن التمثيل والممثل-2" (الحوار المتمدن- ع 4504- 2014)*
4- *مارى إلياس، وحنان قصاب حسن "المعجم المسرحى- مفاهيم ومصطلحات وفنون العرض " (لبنان، مكتبة لبنان ناشرون، 1997) ص 57*

5- ينظر فى : عبد الوارث عسر "فن الإلقاء والخطابة" (القاهرة، الهيئة المصرية العامة للكتاب،1992)
6- ينظر فى نجاة على "فن الإلقاء- بين النظرية والتطبيق" (الدار المصرية اللبنانية، ط4، 2008)
7- ينظر فى: سامى عبد الحميد و بيدى حسون " فن الإلقاء" (بغداد، جامعة بغداد، ط4، 1982)
8- فرحان بلبل" أصول الإلقاء والإلقاء المسرحى" (القاهرة، مكتبة مدبولى، 1996) ص 88
9- ويليم شكسبير"هاملت" ترجمة: عبد القادر القط (لبنان، دار الأندلس، 1982)
10- ألكسندر دين" أسس الإخراج المسرحى" ترجمة سعدية غنيم (القاهرة، دار المعارف المصرية، القاهرة ، 1972) ص 357
11- قنسطنطين ستانيسلافسكى " إعداد الممثل" ترجمة: الدكتور محمد زكى العشماوى، ومحمود مرسى (لبنان، بيروت، دار النهضة العربية للطباعة والنشر، ب ت) ص 233
12- -ينظر فى : ابراهيم أنيس "الأصوات اللغوية" (القاهرة، مكتبة الأنجلو المصرية، ط5، 1975)
13- ينظر فى: بشير كمال "علم اللغة العام- الأصوات" (القاهرة، دار المعارف، 1980)
14- عمر مهديوى" الهندسة اللغوية والترجمة الآلية- المفهوم والوظيفة" (بحث مقدم للمنظمة العربية للترجمة، حول الترجمة والحاسوب، 2014) ص 13
15- ينظر فى: د.أحمد على على لقم "تطبيقات هندسة اللغويات العربية-واقع وآفاق" (حولية كلية اللغة العربية- إيتاى البارود، ع31)
16- محمد السكران" الهندسة اللغوية وتنمية العربية" (جريدة الأهرام، ع 47526، 19 يناير 2017)
17- لينوكس موريس" شركات تستعين بتكنولوجيا تحليل الصوت البشرى لاختيار موظفيها" فبراير، 2017 https://www.bbc.com/arabic/
18- وكالة أنباء الشرق الأوسط " الذكاء الاصطناعى يشخص اضطراب ما بعد الصدمة عبر تحليل صوت المريض" ابريل 2019
19- أحمد الجعفرى" الخبراء الفنيين وحل ألغاز الجرائم.." يونيو 2018 https://www.youm7.com/story/2018/6/16
20- - سعاد الهرمزى" التحليل العلمى لصوت العندليب" ابريل 2017. https://www.almadasupplements.com/news.php
21- David Crystal "A Dictionary of linguistics and phonatics" (UK, Blackwell publishing, 2008) p.540
22- ينظر فى : عصام نور الدين "علم الأصوات- الفونوتيكا" (بيروت، دار الفكر اللبنانى، ط1، 1992) ص 18
23- د.كبير بن عيسى "دليل مستعمل تطبيق تحليل الإشارات الصوتية ومعالجتها- برات PRAAT" (الجزائر، مركز البحث العلمى والتتقنى لتطوير اللغة العربية، ع9، 2019) ص ص 20: 21
24- ينظر فى: خرما نايف "أضواء على الدراسات اللغوية المعاصرة" (الكويت، عالم المعرفة، 1979)
25- ينظر فى: بسام بركة "علم الأصوات العام- أصوات اللغة العربية" (لبنان، مركز الإنماء القومى، ب ت)
26- ينظر فى: عبد الجليل عبد القادر"الأصوات اللغوية" (عمان، دار الصفاء، ط1، 1998)
27- د.كبير بن عيسى "دليل مستعمل تطبيق الإشارات الصوتية ومعالجتها" سبق ذكره ص 5
28- برنامج برات بآخر تحديث https://phonetics-acoustics.blogspot.com/2016/04/praat-2016.html
29- هاملت أحد أشهر مسرحيات وليام شكسبير (1564- 1616) إن لم تكن أشهرها على الإطلاق وتصنف كمأساة، أو ما يسمى بـ(مأساة الانتقام) وهي نوع من المسرحيات كان شائعاً في ذلك الوقت . وتدور حول هاملت أمير الدانمارك، الذي يموت أبوه على يد عمه، ويتزوج عمه من أمه ويظهر الشبح لهاملت ويخبره بحقيقة موته، وهنا يبدأ هاملت بالانتقام من عمه ولكن تردده الدائم لا يمكنه من ذلك، وتنتهى المسرحية بموت الجميع.
30- ممثل مسرح وسينما إنجليزى، أول ممثل يمنح لقب سير، لعب عدد من الأدوار المختلفة من التراجيديا اليونانية ،وعصر النهضة، إلى الأنواع المختلفة من الدراما الانجليزية والأمريكية. مقاطع صوتية لمونولج To be or not to be Laurence Olivier's wonderful 1948 film 1948 www.youtube.com
31- ممثل بريطانى مواليد 1968 ، مقاطع صوتية لمونولج To be or not to be Adrian Lester speaks Hamlet's soliloquy from act III, scene 1 2016. www.youtube.com
32- ممثل ومخرج أمريكى مواليد 1956 حصل على عدد من جوائز الأوسكار، مقاطع صوتية لمونولج To be or not to be – Mel gebson movie of Hamlet (1990) www.youtube.com
33- ممثل اسكتلدى مواليد 1971، بدأ مسيرته الفنية عام 1987، وحصل على جائزة إيمى. مقاطع صوتية لمونولج To be or not to be David Tennant as Hamlet in a film of the Royal Shakespeare Company's award-winning production of Shakespeare's greatest play. Directed by Gregory Doran2010 www.youtube.com
34- ممثل أيرلندى مواليد 1976، حصل على عدد من الجوائز ومنها جائزة سير لورانس أوليفييه، مقاطع صوتية لمونولج To be or not to be Andrew scott 2018www.youtube.com
35- ممثل ومخرج بريطانى من أيرلندا الشمالية، مواليد 1960، مثل وأخرج العديد من أعمال شكسبير رشح لعدد من جوائز الأوسكار وجولدن جلوب وغيرها، مقاطع صوتية لمونولج "To be or not to be" From Hamlet, by Kenneth Branagh 1996 www.youtube.com

Summary

Through the The research aimed to evaluate the basis of the creation under controlled by modern technology. scientific analysis of voice using modern programs, so we can evaluate that the actor controls his or her voice in various emotions as a basic step in reaching the creation.

The problem of the study is measuring the emotions that the actor expresses, is achieved by analyzing the sound without affecting the creation. Therefore, the study divided into three sections, the first section, which is entitled "The Art of Acting and its Relationship to voice" it deals with the didacticism and its uses in the training of the actor .

In the second section, "the voice and the language engineering ",it deals with the most important terms which used in the field of language engineering.

And the third section is the applied part . I choice an analog audio performance model from the play Hamlet composed by Willem Shakespeare, and the performance of a number of different actors chronologically and temporally. The sounds were analyzed using the PRAAT program. The research ends with the most important .results, discussing them, recommendations, and finally the list of resources and references.

صديقة لاشين

- **أستاذ التمثيل المساعد بقسم الدراسات المسرحية -جامعة الإسكندرية.**
- **مدرب دولى معتمد من مؤسسة هورايزن الدولية لريادة الأعمال 2019.**
- **عضو لجنة تحكيم مسابقة مهرجان الفنون المسرحية فى الجامعة وخارجها.**
- **مدير مهرجان قسم المسرح المحلى والدولى في دورته الأولي والثانية والثالثة منذ 2011**
- **عدد من الأبحاث المحلية والدولية.**

<u>النشاط الفنى</u>

- **مدرب تمثيل بمركز الفنون- التابع لوزارة الشباب والرياضة بجامعة الإسكندرية منذ 2018**
- **المشاركة بلجنة تحكيم نصوص مسابقة المدارس " مهرجان شكسبير" 2016**
- **مدرب للتمثيل في مشروع القراءة الكبري بمكتبة الإسكندرية 2015.**
- **عضو مشارك للتدريب بالورشة الدائمة لتدريب الممثلين في الورشة الدائمة بالهيئة العامة لقصور الثقافة.**
- **عضو لجنة الندوات بمهرجان نوادي المسرح بالإسكندرية .**
- **عضو لجنة مشاهدة العروض المسرحية في مهرجان التذوق المسرحي الثالث .**
- **عضو لجنة تحكيم اختيار عروض (6 مين) بمركز جوته الألماني .**
- **ممثلة معتمدة باتحاد الإذاعة والتلفزيون منذ 2009.**
- **عمل ورش تدريبية للإعداد الجسدي للممثل، والارتجال، والأداء التمثيلى، منذ 2009 حتى تاريخه**
- **المشاركة بالأداء التمثيلى في عروض مسرح الجامعة وقصور الثقافة منذ عام 1992، وحصلت على العديد من الجوائز.**

The Acoustic Characteristics of Read and Spontaneous Colloquial Arabic Speech Corpora: A Pilot Study

Rudyna Ahmed

Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt

rudynaahmed23@gmail.com

Abstract— **Speaking styles are built by the social environment, dialect, gender and educational level. The speaker characteristics have to be known as a full description of the speaker. The ultimate goal for all speech technology applications is to produce high quality applications. Currently available speech technology applications are not known by a great amount of flexibility, especially not when it comes to different speaking styles. The main aim of this paper is to establish a precise and systematic description of the acoustic characteristics of different speaking styles which are divided into reading speech and conversational speech corpora through spectrographic analysis to clarify the significant differences between them. This study examined the production of Arabic three emphasized vowels [iː], [uː], [aː] in the speech of six native Egyptian speakers, either having a conversation with someone or reading a script. The words containing the target vowels were elicited using two ways of sampling spontaneous interactions and prompted monologues. Fundamental frequency, speaking rate, vowel space area and vowel duration were measured and compared. The result further reveals that there are significant differences between the two types of speech corpora on the acoustic level. Reading speech yielded higher F2, greater vowel space expansion for vowel /i:/. Moreover, vowels in conversational speech were shorter in duration than those in reading speech. Speaking rate was significantly fast in conversational speech. There was no significant change in the fundamental frequency.**

Keywords: **Acoustic analysis, Speech corpora, Colloquial Arabic, Speaking styles, Reading and Conversational Speech**

1 Introduction

The most important factor in corpus design is the intended use for this corpus. The purpose is to determine whether a corpus will be written or spoken language, or both, and what the exact varieties that will be used in it. Biber et al. [1] clarify a corpus as “a large and principled collection of natural texts” and he divided corpus into two types: written and spoken (speech). The highlighted area in this paper is speech corpora which is compiled for specific purposes, especially developing consumer applications.

Corpus is the vital source for research towards specific applications such as text-to-speech, automatic speech recognition and automatic dialogue systems which clearly influence the specific kind of data that will be needed. Automatic speech recognition systems are most effective when the language content is highly constrained, the environment free from extraneous noise, and when the system has been trained on an individual speaker [2]. There are many situations, however, in which these conditions are hard to be applied. Most recognition and text-to-speech systems have to operate in the real world outside the sound-proofed laboratory, and the real world is often noisy. They must deal with real people, whose speaking styles display wide variation.

As the range of interests in corpus collection vary to a great extent, then the methods used for speaking style will also vary. Differences include whether the speech occurs naturally or is elicited for the purpose of data collection. Consequently Speech corpus is divided into two types, firstly: dialogue corpus, which is defined linguistically as spontaneous, unscripted conversation. In this conversation the speakers must sustain the conversation. Another type of dialogue may be characterized as prompted monologues because of the long duration such as: broadcast interviews, the participant roles are asymmetrical. Like dialogue the term of monologue covers a wide variety of speech types that include reading aloud, unscripted but prepared speeches and story-telling [2].

Currently the great majority of text-to-speech and automatic speech recognition applications are not characterized by its variability, especially when it comes to the specific voice or speaking style. On the opposite, the focus has been on reading unrelated sentences. There is, however, a very practical need for different speaking styles which is used in a variety of applications. The range of applications ask for a variation close to that found in human speakers. Apart from these practical needs, there is the scientific interest in formulating our understanding of human speech variability in explicit models. Therefore, it is necessary to know whether there is a significant difference between the various speaking styles or not in order to be scaled and taken into account. In this paper we address this point in terms of acoustic phonetics and whether there are differences to be noticed in the acoustic variables that may differentiate and make a notable distinction between the various speaking styles especially reading and conversational speech.
Acoustic analysis can be informative because it affords quantitative and qualitative analyses that carry potential for speech description and for determining the correlates of perceptual judgments of intelligibility, quality, rate and prosody [3]. Differences in languages and dialects point to the need for further research in this area. The underlying motivation for this study is to examine distinctive acoustic variables that distinguish reading speech from conversational speech (spontaneous) in the colloquial Arabic which is as mentioned above a less explored area expanding this area of research beyond work on other languages.

2 LITERATURE REVIEW

Previous researches on speaking styles has identified a wide range of acoustic features that characterize the conversational and clear speech. These include features that serve to improve the overall acoustic salience of the signal such that it is more resistant to the adverse effects of background noise or a listener-related perceptual deficit [4]. The most studied features that may affect speaking styles are: speaking rate, fundamental frequency, duration and vowel properties.

A. Speaking rate

The study [5] claimed that the speaking rates for conversational speech is 160 to 205 word/ minute and 90 to 100 word/minute for clear speech. He also reported that the slower speaking rate for the clear speech was related to increases in the occurrence and average duration of pauses, and to increases in the duration of many sound segments. A study by [6] mentioned that the overall sentence duration increases of 51% for males and 116% for females, respectively in clear speaking style. Another study by [7] carried on five subjects focusing on clear and conversational speaking styles revealed that the speaking rates of clear speech is 144 to 200 wpm (average, 174 wpm) but in the conversational speech is 140 to 204 wpm (average, 179 wpm).

B. Fundamental frequency

Most of previous studies reported that clear speech has higher F0 and a larger range in F0, and this is due to the larger amounts of laryngeal tension that the clear speech requires. A study by [6] carried on two subjects one male and one female revealed that the mean F0 was increased by 1.1 and 5.4 semitones, and F0 range was increased by 6.2 and 5.8 semitones. However, study by [8] reported that changes in F0 are not consistent across the speakers.

C. Duration

Clear speech produced without a constraint on speaking rate generally has increased durations of speech segments, though not by the same amount or by the same percentage for all speech sounds [9][5][6]. In a study carried by [6] the researcher claimed that the vowel lengthening in clear speech for both English and Spanish relative to conversational speech, although that the amount of lengthening is less in Spanish.

D. Vowel properties

The great majority of studies stated that in clear speech the formant frequencies of vowels generally span a larger space F1 and F2 than conversational speech vowels [5] [6] [10] [11] [12] [13]. Bradlow in his research

[6] found a similar amounts of vowel expansion in clear speech in both English and Spanish vowels. Krause and Braida [14] figured out a less consistent result in their comparison of formants of clear speech and conversational speech, in that only the tense vowels of one talker showed a larger vowel space. Clear speech has increased rates of F2 transitions [12], narrower formant bandwidths [14] and longer durations of formant transitions [8] but these results are not always consistent across all vowel classes.

3 METHODOLOGY

The lack of available Arabic speech corpora, especially Egyptian Arabic, was a serious problem in this research, there was no available data to carry out the analysis. Accordingly a hand-tailored speech corpus was recorded to fit our analysis.

Q. Subjects

To elicit the reading and the conversation speech, six Egyptian Arabic native speakers, encompassing no regional accents participated in the study. All speakers live in Alexandria, their age ranged from 21 to 24 years old. The subjects were matched pairs in gender as in (table 1).

TABLE XIII
SUBJECTS

Number of subjects	**Gender**	**Age**
1	Male	21.
2	Male	22
3	Male	24
4	Female	21
5	Female	22
6	Female	24

R. Corpus

Approximately 20 sentences were read by every subject to obtain the speech sample in which every sentence must contain a word from the three content words that contain three Egyptian Arabic long emphasized vowels ([ɑ:], [i:], [u:]) (/ʔɪstwaːna/’cylinder’, /mʊsṭɑtiːl/ ’triangle’, /mæxruːṭ/’cone’). Also, the conversational speech sample was obtained by focusing on repeating the previous three content words by managing a natural conversation with the subjects. All vowels in the three content words appeared in stressed syllables.

S. Corpus Elicitation

In order to obtain comparable samples of the target vowels and sentences across speech styles, firstly spontaneous speech sample was elicited by giving the subjects picture in which there was geometrical shapes (e.g. the blue triangle is above the red cone and to the left of the green cylinder) which one of the two participants describes so that the other can re-create the same pattern. We have initiated a conversation about these geometrical shapes with the subjects trying not to deviate from the context. [15], in this way of elicitation the speech is spontaneous (unscripted), but very limited both in vocabulary. Secondly: reading speech sample task was elicited by asking the subjects to read the written orthographic transcription targeting approximately 20 sentences. The speech sample was recorded by Praat, with a sampling rate 44,100 Hz.

T. Acoustic Analysis

A total of 120 words and 120 sentences were acoustically analyzed in both conditions reading and conversational speech. Acoustic measurements were carried out in Praat. The analyzed acoustic variables are as follows:

10) Fundamental Frequency: Fundamental frequency is global talker characteristic that varies across genders. [5] claimed that, clear speech is characterized by a somewhat wider range of F0 with a slight bias towards higher F0 than conversational speech. In [6] study of clear speech, average (mean) F0 was increased by 1.1 and 5.4 semitones, and F0 range was increased by 6.2 and 5.8 semitones, respectively, for the one male and one female talker.

All F0 analysis were carried out by Praat. For each vowel produced by each subject the mean, minimum and maximum F0 were calculated ten times in the reading and conversational speech for each speaker. Vowel F0 were obtained from vowel midpoint.

11) Speaking Rate: The overall speaking rate is also a global talker characteristic, although it is not a voice quality characteristic it is one of the most important features that distinguishes clear speech and conversational speech within individuals [16].

Speaking rate was measured for each of the subjects for ten sentences in both reading and conversational speech. Speaking rate is expressed as the mean number of phonemes per second.

12) *Vowel Space Area*: In order to calculate each speaker vowel space area, we selected ten occurrences of each of the three vowels ([ɑ:], [i:], [u:]), from the sentences of our corpus. All words were content words and none was the final word in the sentence. F1 and F2 were measured using Praat. F1 and F2 were obtained from vowel midpoint. Although it is known that steady-state values for formant transitions are not found simultaneously for F1 and F2 [17], if at all, various proposals have been made for measuring steady-state values for formants. They were converted to the psychoacoustic Bark scale [18]:

$Z= \{26.81/ (1+1960/f\}-0.53$, where Z is the critical band value of a formant in Bark and f is a formant's frequency in Hertz.

Vowel space area was measured according to the following formula [19]:

$\text{Distance } (F1,F2)= \{\sum_i (F1_i- F2_i)^2\}^{½}$

13) *Vowel Duration*: In order to calculate each speaker vowel duration, we also selected ten occurrences of each of the three vowels (/ɑ:/, /i:/, /u:/). Vowel duration was calculated from the waveform and spectrogram by Praat.

4 RESULTS AND DISCUSSION

U. Global results

Global results are shown in Table 2.

TABLE 2
GLOBAL RESULTS

	Reading Speech	**Conversational Speech**
F0 mean (HZ)	185.95	190.12
F0 range (Hz)	70.53	66.33
Speaking rate (phoneme/sec)	12.19	18.24667

Vowel duration	0.119539	0.09236
Vowel space area mean value of /a:/ (Barks)	3.15	3.83
Vowel space area mean value of /i:/ (Barks)	10.63	9.36
Vowel space area mean value of /u:/ (Barks)	4.38	4.25

1) *Fundamental Frequency:* For vowel F0, there was a no significant effect of speech styles. Vowels in conversational speech were higher in F0 by 4.17 Hz.

2) *Speaking Rate:* The statistical analysis revealed a significant effect of speech styles on the speaking rate. Speaking rate in conversational speech were remarkably faster than reading speech by 26.66%.

3) *Vowel Space Area:* The overall mean vowel space area of conversational speech (6.02 Barks) was larger than that of reading speech (5.96 Barks), but there was no significant effect. However, the expansion of the vowel space area in reading speech was driven by a change in the F2 rather than the F1 dimension, especially in the F2 of conversational speech which was higher by 1.27 Barks for all speakers.

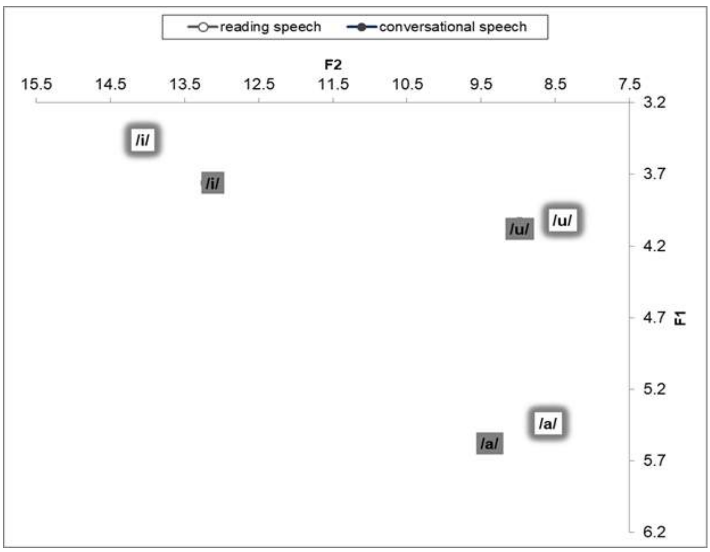


Figure 1: Vowel Space Area of Egyptian Arabic for Reading and Conversational Speech

4) *Vowel Duration:* The statistical analysis revealed a significant difference between reading speech and conversational speech with regard to vowel duration. Vowels in conversational speech were found to be shorter by an average 0.026 ms.

V. Individual results

Male1, male 3, female1 and female2 (Figure 1) correspond to a great extent to the global results but their average fundamental frequency in conversational speech is significantly high that may distinguish conversational speech from reading speech. On the contrary, the fundamental frequency of male 2 and female 3 is higher in reading speech. The other acoustic variables almost correspond to the global results.

TABLE3
INDIVIDUAL RESULTS OF CONVERSATIONAL SPEECH

Gender	Vowel	F0 Mean (HZ)	F0 Range (HZ)	F1 (HZ)	F2 (HZ)	Vowel Duration	Speech Rate (phoneme/sec.)
Male 1	/a:/	161.688	51	540.577	1113.391	0.0478	17.08
	/i:/			369.208	2046.312	0.0897	
	/u:/			375.647	1065.857	0.0817	
Male 2	/a:/	118.639	38	679.216	1328.074	0.0580	18.14
	/i:/			349.360	2028.851	0.1065	
	/u:/			407.043	1081.565	0.0792	
Male 3	/a:/	102.842	35	519.202	954.9057	0.0899	17.12
	/i:/			290.047	2039.266	0.0882	
	/u:/			403.348	972.9897	0.1207	
Female 1	/a:/	257.323	91	679.216	1328.074	0.0968	18.78
	/i:/			349.360	2028.851	0.1337	
	/u:/			407.043	1081.565	0.0809	
Female 2	/a:/	247.684	99	496.882	1149.673	0.1018	19.02
	/i:/			374.647	1978.328	0.1078	
	/u:/			379.977	1132.457	0.1027	
Female 3	/a:/	252.565	84	560.457	1030.569	0.0356	19.34
	/i:/			513.276	2088.497	0.0544	
	/u:/			475.339	1136.528	0.4784	

TABLE4

INDIVIDUAL RESULTS OF READING SPEECH

Gender	Vowel	F0 Mean (HZ)	F0 Range (HZ)	F1 (HZ)	F2 (HZ)	Vowel Duration	Speech Rate (phoneme/sec)
Male 1	/a:/	149.192	53	564.620	1005.286	0.1112	11.82
	/i:/			305.862	2220.534	0.1414	
	/u:/			359.143	1010.481	0.1468	
Male 2	/a:/	135.3293	40	638.207	1058.991	0.0742	15.12
	/i:/			327.755	2492.63	0.1112	
	/u:/			381.963	957.6422	0.0829	
Male 3	/a:/	113.947	44	513.543	917.4883	0.1060	10.83
	/i:/			273.001	2157.643	0.1273	
	/u:/			358.630	868.7661	0.1274	
Female 1	/a:/	220.623	117	638.207	1058.991	0.1103	10.6
	/i:/			327.755	2492.635	0.1288	
	/u:/			381.963	957.6422	0.1442	
Female 2	/a:/	224.384	96	482.033	1038.341	0.1096	10.72
	/i:/			353.546	2276.129	0.1537	
	/u:/			399.751	1034.272	0.1354	
Female 3	/a:/	272.226	73	536.695	987.8427	0.0964	14.05
	/i:/			472.736	2421.009	0.1284	
	/u:/			523.158	1050.382	0.1165	

5 CONCLUSIONS

The acoustic variables that were measured clarified that there is a difference between speaking styles and showed that there is need to work in this area in the future. We look forward to test our results by manipulating speech synthesis and evaluating the perception of the results. Many more studies, involving other style changes, relations and dialects other than Colloquial Arabic also need to be carried out in order to aid our understanding of the limits of individual variability.

REFERENCES

[1] Biber, Douglas/Conrad, Susan/Reppen, Randi (1998), Corpus Linguistics. Investigating Language Structure and Use. Cambridge: Cambridge University Press.

[2] Lüdeling, A. (Ed.) & Kytö, M. (Ed.) (2008). *Volume 1*. Berlin, Boston: De Gruyter Mouton.

[3] Kent, R. D. and Kim, Y. (2008). Acoustic analysis of speech. In M. J. Ball, M. R. Perkins, N. Müller, and S. Howard (eds), The Handbook of Clinical Linguistics. Oxford

[4] Bradlow, Ann & Bent, Tessa. (2002). The Clear Speech Effect for Non-Native Listeners. The Journal of the Acoustical Society of America. 112. 272-84. 10.1121/1.1487837.

[5] Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. Journal of Speech & Hearing Research, 29, 434–46.

[6] Bradlow, A. R. (2003). Confluent talker and listener-related forces in clear speech production. In C. Gussenhoven & N. Warner (eds.), Laboratory Phonology, 7 (pp. 241–73). Berlin & New York: Mouton de Gruyter.

[7] Krause, J. C. & Braida, L. D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. Journal of the Acoustical Society of America, 112, 2165–72.

[8] Krause, J. C. (2001). Properties of naturally produced clear speech at normal rates and implications for intelligibility enhancement. Unpublished doctoral thesis, MIT, Cambridge, MA.

[9] Ferguson, S. H. & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal hearing and hearing-impaired listeners. Journal of the Acoustical Society of America, 112, 259–71.

[10] Chen, F. R. (1980). Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level. Unpublished master's thesis, MIT, Cambridge, MA.

[11] Ferguson, S. H. & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal hearing and hearing-impaired listeners. Journal of the Acoustical Society of America, 112, 259–71.

[12] Moon, S.-J. & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. Journal of the Acoustical Society of America, 96, 40–55.

[13] Kuhl, P. K. et al. (1997). Cross-language analysis of phonetic units in language addressed to infants. Science, 277, 684–6.

[14] Krause, J. C. & Braida, L. D. (2003). Acoustic properties of naturally produced clear speech at normal speaking rates. Journal of the Acoustical Society of America, 115, 362–78.

[15] Swerts, M./Collier, R. (1992), On the Controlled Elicitation of Spontaneous Speech. In: Speech Communication 11, 463 468.

[16] Ann R. Bradlow, Gina M. Torretta, David B. Pisoni. (1996) Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. Speech Communication 20:3-4, 255-272.

[17] Di Benedetto, M.-G. (1989). Vowel representation: Some observations on temporal and spectral properties of the first formant frequency. The Journal of the Acoustical Society of America, 86, 55–66.

[18] Traunmüller, H. 1990. Analytical expressions for the tonotopic sensory scale. J. Acoustic Society of America, 88, 97–100

[19] Bradlow, Ann & Torretta, Gina & Pisoni, David. (1996). Intelligibility of Normal Speech I: Global and Fine-Grained Acoustic-Phonetic Talker Characteristics. Speech Comm. 20. 255-272. 10.1016/S0167-6393(96)00063-5.

ACKNOWLEDGMENT

This work was supported by Dr. Sameh Alansary. I thank him for his insightful comments.

BIOGRAPHY

Rudyna: *Master Degree student at Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt*

She graduated from Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, (2019), her main areas of interest include Acoustic Analysis, Prosodic Analysis and Speech Corpora

الخصائص الاكوستيكية للمدونات اللغوية للكلام المقروء و المحادثة في العامية المصرية

Rudyna Ahmed

Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt

rudynaahmed23@gmail.com

ملخص — **اساليب الكلام تبنى من خلال البيئة الاجتماعية ، واللهجة ، والجنس ، والمستوى التعليمي وأيضًا حسب موقف ما معين يمر به المتحدث. الهدف النهائي لجميع تطبيقات تكنولوجيا الكلام هو إنتاج تطبيقات عالية الجودة. لا تُعرف تطبيقات تكنولوجيا الكلام المتاحة حاليًا بقدر كبير من المرونة ، لا سيما عندما يتعلق الأمر بأساليب التحدث المختلفة. الهدف الرئيسي من هذا البحث هو وضع وصف دقيق ومنهجي للخصائص الصوتية المميزة لأنماط التحدث المختلفة و التي تنقسم الى كلام مقروء و محادثة و ذلك من خلال التحليل الاكوستي الطيفي. تناولت هذه الدراسة نطق ثلاثة صوائت مفخمة و هم كالاتي: [iː]، [uː]،[aː] و يتم تسجيلهم لستة متحدثين للعربية المصرية ، و ذلك عن طريق اقامة محادثة و قراءة نص. تم قياس ومقارنة التردد الأساسي ، ومعدل الكلام ، و مساحة الفراغ الاكوستي و مدة لكل صائت. كشفت النتيجة عن وجود فروق بين الكلام المقروء و المحادثة على المستوى الاكوستي. أسفرت النتائج عن ارتفاع الشكل الثاني للصائت /iː/ ، و طول المدة الزمنية للكلام المقروء عن طول المدة الزمنية للمحادثة، و ان معدل الكلام ايضاً كان مرتفع عن المحادثة اكثر من الكلام المقروء. لم يتم ملاحظة اي تاثير ملحوظ من التردد الاساسي ليميز بين المحادثة والكلام المقروء.**

الكلمات الدالة: **التحليل الاكوستيكي، المدونات اللغوية، العامية المصرية، اساليب الكلام ، الكلام المقروء و المحادثة.**

Syllables Classification of ASR using Hybrid Visual Features in Fixed HMM

Doaa A. Lehabik [*1], Amr M. Gody [*2], Mohamed H. Merzban[*3], Sameh F. Saad [**4]

[*]*Electrical Engineering Department, Fayoum University, Fayoum, EGYPT*
[1]da1174@ fayoum.edu.eg
[2]amg00@fayoum.edu.eg
[3]mhm00@fayoum.edu.eg

[**]*Modern Sciences and Arts University, 6 October City, Giza, Egypt*
[4]dr.sam.far@gmail.com

Abstract: This paper presents a different approach to classifying speech phonemes. Two-hybrid techniques are used to emphasize the principle idea of this paper. The first hybrid model is constructed of fixed state, structured Hidden Markov Model, Gaussian Mixture, Mel scaled Best Tree, Convolution Neural network, Vector Quantization (FS-HMM-GM-MBT-CNN-VQ). The second hybrid model is constructed of fixed state, structured Hidden Markov Model, Gaussian Mixture, Mel scaled Best Tree, Convolution Neural. TIMIT database is used in this paper. All phones are classified into five classes and segregated into Vowels, Stops, Fricatives, Nasals, and Silences. The results show that using (FS-HMM-GM-MBT-CNN-VQ) is an available method for classification of phonemes, with the potential for use in applications such as automatic speech recognition and automatic language identification.

Key words: ***Automatic Speech Recognition, Classification technique, HTK, wavelet packet, Convolution Neural Network, Vector Quantization, Hidden Markov Model***

7 INTRODUCTION

Automatic speech recognition (ASR) is recognized as the independent computer-driven transcription which changes talked language into legible text. ASR lets a computer to get the words from a person that speaks into a microphone and alter them into written text. The phonetic features and articulation of various sounds are necessary for the classification into separate categories. This classification of sounds can be applied for applications like speaking rate evaluation, speech recognition, phone recognition, and language recognition. The broad phone classes are usually known as vowels, nasals, fricatives, stops, and silence. This categorization can improve speech recognition and hence categorization techniques were attempted. The current research presents two classification methods. The hybrid features model of MBT, CNN and or without VQ with fixed states of the Hidden Markov Model is used. Various Gaussian mixture numbers are used to get a greater rate of recognition. Now, Section 2 explains related work but, section 3 discusses each block in the proposed model. The experiment environment includes a database and experiment procedure in section 4. The results discussion would introduce in section 5 and conclusions would be shown in section 6.

.

8 RELATED WORK

Using classification techniques in ASR is the best preprocessing tasks to increase the recognition speech rate. Classification means that you had some categories and observed units and you want to assign these observed units to these categories. Classification is performed by matching feature vectors from various categories with the observed units.

In [1], expert classifiers are used for each broad phonetic class which classifies speech signals into vowels, stops, fricatives, and nasals and used the TIMIT database. The highest obtained phone accuracy was 74.2%. In [2], two feature extraction techniques to classify speech signals into five classes in the TIMIT database.

These features are MFCC and time-frequency reassigned cepstral coefficients (TFRCC). For stops, the highest success rate (SR) is 53.74% by TFRCC but other classes achieved high SR when using MFCC. In [3], the authors classify the speech phonemes based on histograms of the reconstructed phase spaces into three classes Which are fricative, vowel, and nasal in the TIMIT database. The results achieved overall recognition rates of 61.59%, 34.49% and 30.21% for fricative, vowel, and nasal phonemes respectively. Using the Malayalam speech signal in [4], that used a two-stage system to spot the boundaries of vowels, nasals, and approximants. In the first stage, a speech signal is classified into six broad phoneme classes using an ANN but for second stages, frequency domain parameter named spectral peak frequency is suggested for accurate verification of nasals. Sonorant and non-syllabic features are used for verifying approximants and the syllabic feature is used for locating vowels. In [5], the authors classified speech signals into 19 classes which were 8 vowels and 11 consonants. This classification was made using two techniques that were Procrustes analysis and support vector machine (SVM). Procrustes analysis performed 91.67%, 91.37% accuracy for vowel and consonants respectively but SVM performed 89.05%, 88.94% accuracy for vowel and consonants respectively. In [6], support vector machine (SVM) is used to match phonemes into 6 classes in Gujarati language and the accuracy is 95.70 %. Producing in [7] new classification for broad phoneme by features that obtained immediately from a speech at the level of this signal. Broad phoneme classes comprise vowels, nasals, fricatives, stops, approximants, and silence. This classification is applied to three systems, each system is applied to three tests and results are 54%, 61%, and 46% for the combination on TEST 1, TEST 2 and TEST 3 respectively. Based on the pattern matching in [8], there is a method for accurate spotting of plosives which tested using TIMIT corpus. Results showed that the accuracy of spotting the plosives using the presented approach was very high. The distribution of insertions for the different classes is 34% for silences, 28% for fricatives, 23% for vowels, 7% for glides, 6.5% for nasals and 1.5% for affricates.

9 PROPOSED MODEL

In this model as indicated in Fig. 1, the input speech signal was resampled into 10 kHz to best distribute the wavelet tree structure through the significant band. Then framing it into small frames (20 ms). Mel Best Tree (MBT) images that are obtained from speech signal enter on Convolution Neural Network (CNN) then to Vector Quantization (VQ) to extract the features. This feature enters on HMM with various GMM to analyze speech signals into five classes; Vowels (V), Stops (S), Fricatives (F), Nasals (N) and Silences (Si). The state's number of HMM is fixed for each class. Each block of this model would be discussed in more detail

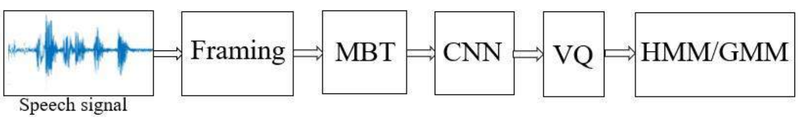


Figure 1: Block diagram of the proposed model

in the latter subsection.

W. Feature Extraction

1) Mel Best Tree

The idea of Best Tree Encoding should be introduced. The algorithm of creating the BTE feature file [9] starts by converting the speech signal into a collection of short frames. Then, the entropy of the Wavelet Packet Decomposition (WPD) coefficients is applied as a projection of these frames of the speech signal power into defined filter banks. The best tree that contains the significant signal power, using the acquired entropy, is obtained. The dynamic range of the BTE is normalized using the encoding algorithm in [10]. The new direction is considered in the version of BTE (BTE with Mel-filter). The algorithm of estimating the best tree is targeted in this version of BTE. It extracts the Mel frequency wavelet packet Best Tree Encoding

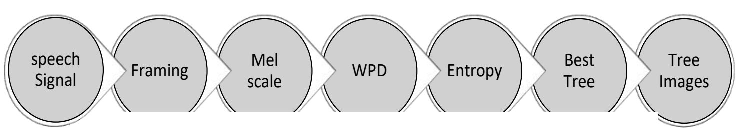


Figure 2: Block diagram of creating MBT

(MFBC) features from the WAV file. It considers applying Mel-scale before the conversion [11]. It also resamples the input to 10 kHz. In this research, we used Mel best tree images not encoding as in Fig. 2.

2) *Convolution Neural Network*

It is a branch of deep learning that used for image recognition and rarely in speech recognition [12] and we choose deep residual network (Resnet50) that act as a type of CNN with 50 layers. This network can organize images into 1000 object classifications. Subsequently, the network can learn rich features for a broad range of images. It has a size for an input image of [224 224 3]. The output features vector is included in 1000 features.

3) Vector Quantization

Vector quantization (VQ) is an efficient source-coding model. Vector quantization is a method that encodes the input vector into an integer number with an entry of a group of reproduction vectors. The reproduction vector is defined to be closest to the input vector. The coding efficiency is originated in the process of converting the vector into a compact integer representation. The performance of the vector quantizer, however, depends on whether the set of reproduction vectors, which are usually called code words, is accurately chosen such that the distortion is minimized.

X. Hidden Markov Model with Gaussian Mixture Model

Hidden Markov Model (HMM) is the ablest method used in automatic speech recognition. This system is produced for the Markov process with private parameters and we want to distinguish the hidden parameters from the observation [13]. The states are hidden, and the probability distribution for each is known as the variable which affects the states. Temporal data and states are usually identified as separate GMMs [14] in the HMM model. The transition matrix learns from training data and it is a known transition of state to another [15].

10 EXPERIMENT ENVIRONMENT

The later subsections explain the type of database that is used in this paper. the process of converting raw speech into features used in this classification and verification. MATLAB 2018b and visual studio 2015 are used as lab environment. The specifications of the laptop that used in the experiment are 8.00 GB RAM, 64-bit operating system, Intel(R) Core (TM) i5-8250U CPU @1.60 GHz 1.80 GHz and NVIDIA GeForce MX150 with 8061 MB Memory @4 GHz.

Y. Database

The continuous corpus of TIMIT [16] is an acoustic-sounding speech made of English, recorded by a microphone at 16 kHz and 16-bit resolution. This database holds 6300 sentences (5.4 hours) in 630 speakers from 8 regional dialects of the United States (US). Each speaker articulated 10 sentences and all the sentences were identified with its phone level. The main version of TIMIT includes 61 phonics. The database is prepared to modify transcription files for the character recognition objective of this research. Vowels (V), Stops (S), Friction (F), Nasal (N) and Silence (Si) [17]. The following table presents the phone assigned to each classification.

TABLE 14

PHONES CLASSIFIERS

Classifiers	TIMIT Labels
Vowels (V)	aa, ae, ah, ao, ax, ax-h, axr, ay, aw, eh, el, er, ey, ih, ix, iy, l, ow, oy, r, uh, uw, ux, w, y
Stops (S)	p, t, k, b, d, g, jh, ch
Fricatives (F)	s, sh, z, zh, f, th, v, dh, hh, hv
Nasals (N)	m, em, n, nx, ng, eng, en
Silences (Si)	h#, epi, pau, bcl, dcl, gcl, pcl, tcl, kcl, q, dx

Z. The Procedure of the proposed model

In this model, First, speech signal enters on MBT block. In this block; read speech signal and resample signal into 10 kHz. These samples are framed into 20 ms. The Mel-scale curve is implemented to convert all frequencies to Mel frequencies. Then apply wavelet packet decomposition and extract the best tree images. Second, Mel Best Tree images are entering as input on CNN block. In that extracting features from MBT images by using Resnet50. In this case; the features output vector is containing 1000 components. Third, these features enter as input to vector quantization block which features output vector is one component. Fourth, these features enter on HMM/GMM block. In this block; all classifiers are trained to utilize a fixed number of HMM states. In this model, we used four emitting states and two non-emitting states as in Fig.3 (The non-emitting states are required to identify the entry and the exit state in HMM model). GMM is used to act as spatial distribution probability density functions of attribute vectors of N-Dimensions. The mixture count is variable in this research. The system is tested with various Gaussian mixture counts (1, 2, 3, 4, 5, 6 and 7). HTK tool is utilized to build HMM-based speech processing tools. There is another model that is not used by VQ. In this, we used feature vector as 1000 components that were extracted from CNN block, then

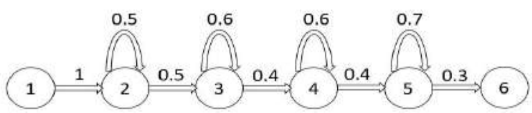


Figure 3: FS-HMM-VQ for all classifiers model

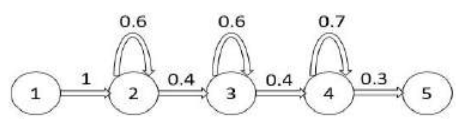


Figure 4: FS-HMM for all classifiers model

these features enter on HMM/GMM block which used five a fixed state (three emitting states and tow non-emitting states) as in Fig. 4.

11 RESULTS AND DISCUSSIONS

There are two models of features vectors that are applied: First, (MBT-CNN-VQ) is used and this vector consists of one component. Second, (MBT-CNN) is used and this vector is 1000 components as indicated in session 4. The results of these two proposed models (MBT-CNN-VQ) and (MBT-CNN) are illustrated in this section. Success Rate (SR) of each syllable is utilized for evaluating the proposed model. The comparison study is implemented to show the details and the key power in each specific feature set. We calculate the success rate (SR) in each case as in Table 2. The success rate can be defined by equation 1 and the result is shown as in Fig. 5 that shows the value of each SR against the Gaussian mixture (GM). In this equation: (D denotes as deletions), (s denotes as substitution) and (n denotes as the number of phones in the expected transcription). The following tables show the result in two cases with various numbers of GMM (1,2,3,4,5,6 and 7). Case1: Using (MBT, CNN, and VQ) features with fixed states number of HMM as in Fig. 6. Case 2: Using (MBT and CNN) features with fixed states number of HMM as in Fig. 7.

$$SR = \frac{N - D - S}{N} \qquad (1)$$

TABLE 15

EXPERIMENT RESULTS OF SUCCESS RATE

Mixture Count	Case 1 SR%	Case 2 SR%
1	43.42%	42.44%
2	47.36%	43.13%
3	50.21%	46.78%
4	51.79%	46.95%
5	54.97%	45.59%
6	**57.49%**	43.08%
7	57.07%	**47.76%**

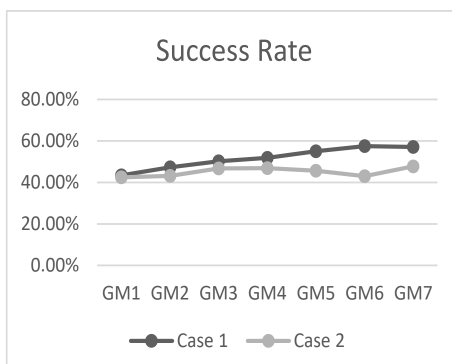


Figure 5: Success rate results

Table 2 shows that the best successive rate in two modules that achieved 57.49% when applying (MBT-CNN-VQ) features extraction with fixed state Hidden Markov model and Gaussian mixture number 6. Table 3 shows that the greatest success rate for each classifier. In case 1: vowels achieved 81.1% using a gaussian number 6, stops achieved 93% using gaussian number 1, fricatives achieved 70.5% using gaussian number 7, nasals achieved 94.7% using a gaussian number 4 and silences achieved 47.6% using gaussian number 7 as in Fig. 6. In case 2: vowels achieved 67.5% using a gaussian number 3, stops achieved 93.3% using gaussian number 1, fricatives achieved 87% using gaussian number 4, nasals achieved 72% using a gaussian number 2 and silences achieved 92.4% using gaussian number 6 as in Fig. 7.

TABLE 16

SUCCESS RATE FOR EACH CLASSIFIER IN EACH CASE

	SR for each classifier									
	Fixed states HMM and using vector quantization					Fixed states HMM and without using vector quantization				
GMM	V	S	F	N	Si	V	S	F	N	Si
1	53	**93**	0	91.8	29	65.6	**93.3**	28.4	45.1	63.2
2	63.9	87.1	3.5	92.1	30.3	64.6	89	52.8	**72**	53.1
3	71.9	76	15.9	93.9	31.6	**67.5**	92.1	60.1	61.7	63.9
4	73	58.5	48.4	**94.7**	35.2	59.6	76.2	**87**	66.4	72.9
5	78.3	57.5	58.7	90.8	41.4	49	75.1	72.4	61.1	89.9
6	**81.1**	67.7	66.9	81.7	45.5	39.8	75	64.7	65	**92.4**
7	78.2	77.9	**70.5**	70.9	**47.6**	65.1	75.1	71.4	51.3	90.4

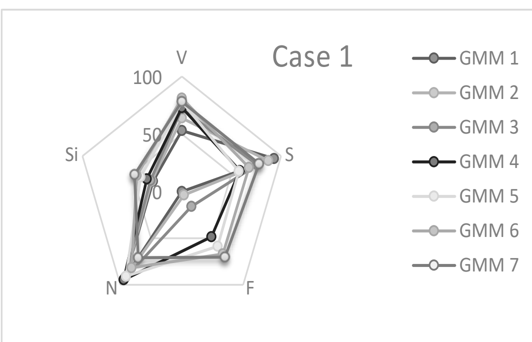


Figure 6: SR of classes in Fixed states HMM and using VQ

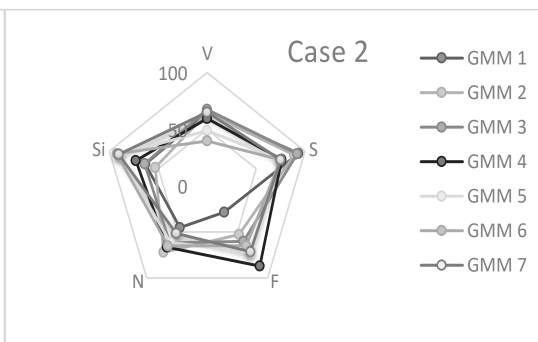


Figure 7: SR of classes in Fixed states HMM and without using VQ

The confusion matrix (error matrix) results of all experiments were showed in table 4 exhibit which is a table layout applied for a classification model, a classifier, or a recognizer as in our case. The performance summary is depended on test data, where its actual values are well identified. This table relates to the probability (success rate of the class) of the best decision. Vowels are the usual occurring sound group in the speech data. In GMM 1, 2 and 3 vowels are not recognized as silence. Vowels are recognized more in the first model than second. Stops and fricatives have small appearances in the TIMIT database. Stops are very far from silence in small gaussian, but they very near in nasals in the first model. Fricatives are very bad in the first model with small gaussian because they were recognized as stops and nasals more. Fricatives have increased in successive rate with the higher Gaussian mixture. Nasals are ideally recognized as nasals in the first model and good in the second model. For silence when using vector quantization, a result of successive rate is small because this class contain [17] (h#, epi, pau, bcl, dcl, gcl, pcl, tcl, kcl, q, dx) that (h#) had high duration in time but (bcl) had short time. So, vector quantization is not suitable for this class. But in the second model, we used all features vector (1000 vector size) not compressed vector (1 vector size). So, the result of this class is perfect with the high gaussian mixture.

TABLE 17

CONFUSION MATRIX

GMM		Confusion Matrix									
		Fixed states HMM and using vector quantization					Fixed states HMM and without using vector quantization				
		V	S	F	N	Si	V	S	F	N	Si
1	V	**13340**	6340		5492	0	**13466**	5642	684	430	302
	S	125	**6759**		358	0	63	**5832**	306	39	13
	F	353	2943		2539	0	109	2297	**1087**	201	130
	N	84	307		**4394**	0	44	1121	188	**1174**	78
	Si	590	3717		3929	**3365**	243	2477	308	265	**5662**
2	V	**16386**	3940	27	5310	0	**12933**	4079	1494	1403	123
	S	221	**6091**	4	680	0	70	**5438**	433	163	5
	F	540	2071	**193**	2686	0	134	1464	**2485**	563	56
	N	139	236	0	**4348**	0	41	526	330	**2368**	23
	Si	922	2701	23	4085	**3360**	298	2043	670	918	**4440**
3	V	**18540**	1896	158	5194	1	**13784**	4179	1103	945	417
	S	348	**5037**	33	1213	0	113	**5547**	189	140	36
	F	732	1165	**878**	2751	0	141	1321	**2958**	354	147
	N	158	123	6	**4391**	0	88	674	266	**1827**	105
	Si	1266	1544	145	4324	**3363**	340	1846	530	619	**5894**
4	V	**18665**	1020	683	5173	19	**11030**	1879	3322	1452	837
	S	498	**3604**	270	1786	0	140	**3821**	719	274	63
	F	569	485	**2981**	2117	5	91	332	**5648**	284	139
	N	130	62	54	**4393**	0	71	289	585	**2133**	134
	Si	1100	779	585	4122	**3576**	229	559	1384	608	**7486**
5	V	**19965**	947	929	3577	92	**8326**	1867	2326	1256	3214
	S	650	**3395**	380	1475	1	101	**3611**	471	232	394
	F	648	436	**3704**	1491	29	85	434	**4168**	288	782
	N	222	93	92	**4082**	5	50	270	428	**1936**	484
	Si	1278	652	844	3058	**4112**	100	265	675	223	**11203**
6	V	**20574**	1298	1223	2137	131	**6297**	1935	2059	1578	3939
	S	642	**4073**	506	786	5	64	**3627**	417	196	532
	F	682	548	**4251**	815	57	44	490	**3546**	370	1031
	N	343	217	190	**3411**	15	26	221	381	**2113**	511

	Si	1522	1019	1121	1817	**4565**	48	206	519	218	**12051**
7	V	**19441**	2169	1609	1474	167	**11969**	1649	1952	680	2125
	S	495	**4953**	459	441	7	157	**3267**	444	161	321
	F	590	773	**4602**	506	57	149	465	**3633**	211	627
	N	373	456	267	**2752**	31	86	310	413	**1294**	418
	Si	1304	1547	1289	1214	**4860**	179	287	500	149	**10471**

12 CONCLUSIONS

It has been indicated that the automatic speech recognition success rate is improved using a hybrid method of acoustic-phonetic approach and pattern recognition approach. The acoustic-phonetic approach is a statistical procedure, which depends on the HMM. The methodology of mixing (MBT-CNN-VQ) gives a greater success rate of correctness than the second model (MBT-CNN). The vector quantization technique also acts as a good role in achieves real results. The result was improved by using a various number of gaussian mixtures. To be specified in terms of particular class classification performance, the highest success rates are produced, using (FS-HMM-GM-MBT-CNN-VQ) as of approximately 81.1% for vowels at (GM 6) and as of 94.7% for nasals at (GM 4). When using (FS-HMM-GM-MBT-CNN), the highest success rates are achieved as of 93.3% for stops at (GM 1), as of 87% for fricatives at (GM 4) and as of 92.4% for silence at (GM 6). By comparing with the result in [1] that classifies speech signals into vowels, stops, fricatives, and nasals with 60.5%, 83.3%, 81.4%, and 75.9% success rate respectively. The successive rate (SR) of vowels and nasals are more than in [1] when used case 1 and more than 1 for stops and fricatives when used case 2. The results are developed by renewing the key parameters of the hybrid model. The key parameters are the features, HMM model and the numbers of the gaussian mixture. HMM, models can be renewed for most often occurring models by tying the states of seldom appearing phones. The silent class can be also altered for better HMM designs. A future work, concerning the modification of the HMM states from fixed state to variable states for each classifier and using RNN instead of CNN to obtain a higher recognition rate.

REFERENCES

[1] P. Scanlon, D. P. Ellis, and R. B. Reilly, "Using broad phonetic group experts for improved speech recognition," *IEEE transactions on audio, speech, and language processing,* vol. 15, no. 3, pp. 803-812, 2007.

[2] G. Tryfou, M. Pellin, and M. Omologo, "Time-frequency reassigned cepstral coefficients for phone-level speech segmentation," in *2014 22nd European Signal Processing Conference (EUSIPCO)*, 2014: IEEE, pp. 2060-2064.

[3] J. Ye, R. J. Povinelli, and M. T. Johnson, "Phoneme classification using naive bayes classifier in reconstructed phase space," in *Proceedings of 2002 IEEE 10th Digital Signal Processing Workshop, 2002 and the 2nd Signal Processing Education Workshop.*, 2002: IEEE, pp. 37-40.

[4] S. Salim, G. Deekshitha, A. George, and L. Mary, "Automatic Spotting of Vowels, Nasals and Approximants from Speech Signals," in *2018 International CET Conference on Control, Communication, and Computing (IC4)*, 2018: IEEE, pp. 272-277.

[5] J. Wang, J. R. Green, A. Samal, and Y. Yunusova, "Articulatory distinctiveness of vowels and consonants: A data-driven approach," *Journal of Speech, Language, and Hearing Research,* 2013.

[6] A. Chittora and H. A. Patil, "Classification of phonemes using modulation spectrogram based features for Gujarati language," in *2014 International Conference on Asian Language Processing (IALP)*, 2014: IEEE, pp. 46-49.

[7] G. Deekshitha and L. Mary, "Broad phoneme classification using signal based features," *International Journal on Soft Computing,* vol. 5, no. 3, p. 1, 2014.

[8] J. Keshet, D. Chazan, and B.-Z. Bobrovsky, "Plosive spotting with margin classifiers," in *Seventh European Conference on Speech Communication and Technology*, 2001.

[9] A. M. Gody, "Wavelet Packets Best Tree 4 Points Encoded (BTE) Features," in *The Eighth Conference on Language Engineering, Ain-Shams University, Cairo, Egypt*, 2008, pp. 189-198.

[10] A. M. Gody, R. A. AbulSeoud, and M. M. Ibraheem, "Hybrid Model Design for Baseline-Context-Independent-Mono-Phone Automatic Speech Recognition," *International Journal of Engineering Trends and Technology (IJETT)–Volume,* vol. 27.

[11] A. Gody, R. Abul Seoud, and M. Ezz El-Din, "Using Mel-Mapped Best Tree Encoding for Baseline-Context-Independent-Mono-Phone Automatic Speech Recognition," *The Egyptian Journal of Language Engineering,* vol. 2, no. 1, pp. 10-24, 2015.

[12] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458,* 2015.

[13] P. Bansal, A. Kant, S. Kumar, A. Sharda, and S. Gupta, "Improved hybrid model of HMM/GMM for speech recognition," 2008.

[14] G. Xuan, W. Zhang, and P. Chai, "EM algorithms of Gaussian mixture model and hidden Markov model," in *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*, 2001, vol. 1: IEEE, pp. 145-148.

[15] J. C. Brown and P. Smaragdis, "Hidden Markov and Gaussian mixture models for automatic call classification," *The Journal of the Acoustical Society of America,* vol. 125, no. 6, pp. EL221-EL224, 2009.

[16] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, "DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM. NIST speech disc 1-1.1," *NASA STI/Recon technical report n,* vol. 93, 1993.

[17] C. Lopes and F. Perdigao, "Phone recognition on the TIMIT database," *Speech Technologies/Book,* vol. 1, pp. 285-302, 2011.

BIOGRAPHY

Amr M. Gody obtained the B.Sc. M.Sc., and PhD. from the Faculty of Engineering, Cairo University, Egypt, in 1991, 1995 and 1999 sequentially. He met the teaching staff of the Electrical Engineering Department, Faculty of Engineering, Fayoum University, Egypt in 1994. He is the Acting director of Electrical Engineering department, Fayoum University in 2010, 2012, 2013, 2014 and 2016. His contemporary research areas of interest involve speech processing, speech recognition and speech compression. He is author and co-author of several papers in national and global conference proceedings and journals such as Springer(International Journal of Speech Technology), the Egyptian Society of Language

Engineering (ESOLE) journal and conferences, International Journal of Engineering Trends and Technology (IJETT), Institute of Electrical and Electronics Engineers (IEEE), International Conference of Signal Processing And Technology (ICSPAT), National Radio Science Conference(NRSC), International Conference on Computer Engineering &System (ICCES) & Conference of Language Engineering(CLE).

Doaa A. Lehabik obtained the B.Sc. degree in Electrical Engineering – Communications and Electronics Department with very good with honor degree, from the Faculty of Engineering - Fayoum University in 2014. She met the M.Sc. program in Fayoum University - Communications and Electronics Department in 2015. She got the Pre-Master degree from Fayoum University with very good with honor degree, in 2016. Her areas of interest involve Speech recognition and Image recognition.

Mohamed H. Merzban obtained the B.Sc. from Fayoum university, M.Sc. from Cairo university, and PhD. from Egypt-Japan university for science and technology (E-JUST). He joined the teaching staff of the Electrical Engineering Department, Faculty of Engineering, Fayoum University, Egypt in 2006. He is now a lecturer at Fayoum university. His current research areas include computer vision and its applications to Robotics.

Sameh F. Saad obtained the B.Sc. degree in mechatronics engineering from High Institute of Engineering, Giza, Egypt, in 2000, the M.S. degree in mechatronics engineering from Mechatronics Laboratory, Paderborn University, NRW, Germany, in 2004 and the Ph.D. degree in electrical engineering at Cairo University, Giza, Egypt in 2012. From 2016 to 2018, he was a Researcher and a Lecturer with the October University for Modern Sciences and Arts, Giza, Egypt. His research interest includes the development of mobile robots and autonomous vehicle, automation of aquaponics ecological system, and automation of mechatronics systems.

التعرف التلقائي على الكلام بأستخدام تصنيف المقاطع الصوتية عن طريق نموذج ماركوف ذو الهيكلة الثابتة

دعاء أحمد لحيبك [1*] ,عمرو محمد جودي[2*] , محمد حمدي مرزبان[3*], سامح فريد[4**]
[*] قسم الهندسة الكهربية, جامعة الفيوم , مصر
[1]da1174@ fayoum.edu.eg
[2]amg00@fayoum.edu.eg
[3]mhm00@fayoum.edu.eg
[**] جامعة العلوم والاداب , 6 اكتوبر , الجيزة , مصر
4dr.sam.far@gmail.com

ملخص
يقدم هذا البحث طريقتين مختلفتين لعملية التصنيف المستخدمة فى عملية التعرف التلقائى على الكلام. اول طريقة تتكون من هذه الخصائص المميزة للصوت (تطبيق نظام الميل سكيل على افضل شجرة من ال WPD مع شبكة التداخل العصبيه مع VQ) وثاني طريقة تتكون هذه الخصائص (تطبيق نظام الميل على افضل شجرة من ال WPD مع شبكة التداخل العصبيه) . تم استخدام بنيات مختلفة لنموذج ماركوف الخفى ذو الهيكلة الثابتة. تم تصنيف المقاطع الصوتية الى 5 مقاطع وهي حروف متحركة (V) و حروف لا تحتوي على كلام (S) و حروف احتكاكية (F) و حروف انفية (N) . وصامت التي لا تحتوي على اي كلام (Si) . تم استخدام قاعدة البيانات TIMIT في هذا البحث . وتم استخدام عدد مختلف من GM الذي يتكون من(1 او 2 او 3 او 4 او 5 او 6 او 7). وحيث أن نجاح عملية التعرف التلقائى على الكلام يعتمد بشكل كبير على إجراء عملية التصنيف بطريقة صحيحة. لذا فإن هذا البحث يركز على الوصول لعملية التصنيف الصحيحة لرفع الكفأة لنظام التعرف التلقائي على الكلام عن طريق تحسين مصفوفة الانتقال TM وتغيير عدد GM. اول طريقة هي التي تعطينا افضل نتائج 57.49% . ونستخدم ال HTK وذلك لشهرتها الواسعة في مجال ال ASR.

الكلمات الدالة
التعرف التلقائي على الكلام, تقنية التصنيف, تحليل حزمة الموجيات (WPD), شبكة التداخل العصبية, المتجهات الكمي (VQ) , نموذج ماركوف الخفي

The Perception of Arabic Vowel Length by Native and Non-native Listeners An Experimental Investigation

Eman Kassem *1, Lamyaa Tawfik**2

*Phonetics and linguistics Department, Faculty of Arts, Alexandria University
Alshatby, Alexandria, Egypt
1Eman.kassem@alexu.edu.eg

** Phonetics and linguistics Department, Faculty of Arts, Alexandria University
Alshatby, Alexandria, Egypt
2lamyaatawfik@hotmail.com

Abstract— **The length of a vowel - in many languages- is considered a distinguishing phonemic aspect; a listener could perceive an utterance in two different manners depending on the length and the quality of a specific vowel. [1] highlights the difference between duration and length stating that vowel duration is the actual timing an utterance takes, whereas length refers to the time of perceiving a vowel.**

Since a vowel length difference in Arabic leads to changes in meaning, the relative length of a vowel turns to be a distinguishing aspect in the Arabic language. However, to what extent there could be a perceptual and acoustic difference between long and short vowels? And what are the acoustic features upon which the native listeners and nonnative Arabic learners [Indonesian listeners) may depend to differentiate between short and long vowels?

This research aims at studying the perception of both native Arabic listeners and foreign listeners of the vowel duration in the Arabic language based on an experimental approach; it also clarifies their dependence on their cognitive relevance of the vowel length, through their responses on the processed vowel length.

Keywords: **vowel length perception, shortening, lengthening, merging, native and non-native listeners, psychoacoustic study.**

1 INTRODUCTION

With respect to all psycho-acoustic relations, it is worth noting that acoustic variation does not necessarily lead to an equivalent perceptive change [1]. Moreover, since the quality of a vowel conveys a great significance in communicating information, there could be – in general - certain differences in the quality of a long vowel and its short counterpart. [2] Stated that There could be both quantitative and qualitative differences between certain long and short pairs of vowels, whereas for other pairs, only quantitative ones occur .

Arabic includes six basic vowels (Fat<u>h</u>a/Kasra/<u>d</u>amma) /æ/ /i/ /u/ and the long vowels (*alif*, *yaa* and *waw*) / æ:/ /i:/ /u:/. Classifying Arabic vowels into *fatha*, *Kasra* and *damma*, does not imply that Arabic has no other vowels. The two vowels /ɛ/ and /o/ have been added along with the other basic vowels as they occur in a great number of Arabic dialects [3]. Such two vowels and the aforementioned emphatic vowel / ɑ /have been added to the present study as they do exist in the Egyptian dialects; they also play a main functional role in identifying meaning.

In this research , it was noticed that meaning change could also occur as a result of length change as well; If the vowel / ɛ / in /bɛ:t/ is shortened, it gives another word with different meaning, such as /bet/ meaning "a girl" as pronounced in Egyptian colloquial Arabic [4]. Thus, the vowels /ɑ/ , /o/ and /ɛ/ are dealt with as independent ones in the present study.

Giving due concern to vowels from a phonological perspective, certain Arabic long vowels do have short counterparts, such as /æ/, / ɑ / and /ɛ/, whereas others do not such as /i , u/. Vowel quality changes might be occurred with some vowels when they turned into short one; for example the word /di:b/ which means " fox" in Arabic , when it is pronounced with short vowel , it turns to vowel / ɛ/ /d ɛb/, also vowel /u/ in the word /ku:b/ . For the latter, a change in the quality of the vowel may occur as it is clear in /i/ and /u/.

Ancient linguists were mainly concerned with the quantity of vowels rather than the quality. The contemporary linguists hold different views concerning the relation between the quantitative and qualitative aspects of a vowel in what concerns the limits differentiating long and short vowels, and whether such limits are restricted to duration or other acoustic features related to the quality of a vowel (Intensity, fundamental frequency and the formants).

The contemporary linguists are divided into two main groups: *The first group* deems that the differences between a long vowel and its short counterpart, from a quantitative perspective, dictate a change in quality.

According to a study made by Omar,A.M [5] indicates that both a long and a short vowel have independent phonemes functionally causing a change in meaning and having minute differences in acoustic features concerning the length or the shortness of the same vowel. This is also manifested by a number of studies, such as that of (6], [4], (7]- [8].

Such conclusions are manifest through a previous experimental study by Al-Ani, S. [9] in which he represented the difference between the long and short vowels through their measured formants as follows:

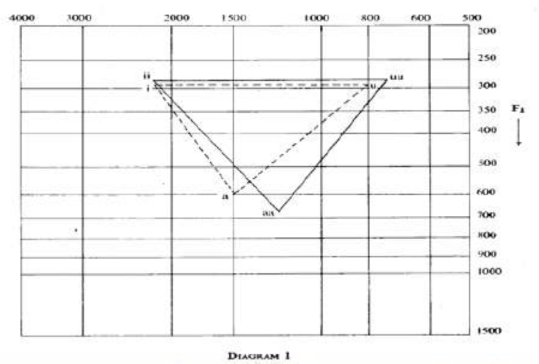


Figure [1] indicates the difference between the short vowels and their long counterparts through measuring F1 and F2 [9].

The second group deems that the differences between a long vowel and its short counterpart are a quantitative one; they state that no qualitative difference occurs, and if any, they are considered to be mere secondary. A study made by [10] also stresses the quantitative difference; he remarks that since fatha, kasra, damma are parts of their long counterparts *alif, jaa* and *waw*, as mentioned by the old grammarian [11.], they thus possess the same quality features. As for the quality, there is no difference whatsoever between the two groups.

This is also manifested by a number of studies, such as that of . [12],[13]- [14].

In Arabic, the relative length of short and long vowels have been measured by some linguists; [9] found that short vowels duration are in the range between 100-150 msec., while long ones are about 225 – 350 msec. [4] are about 95 msec. for short and about 112 msec. for long vowels in monosyllabic words. [15] found that the short vowels are about [117 msec.) while the long vowels about (282 msec.) in monosyllabic words.

In other languages, some of the experiments showed the importance of vowel duration as a prominent cue to differentiate the vowel qualities, whereas it has a little weight compared with the other spectral information.

In Holland language, [16] has conducted an experiment (by changing the spectral information of the vowel, while its durational information hold the same.) its results indicated that native language tend to depend on spectral information, while learners of English as second language depend on durational information.

Tsukada, K. [17 & 18] proved that listeners depend on spectral characteristics beside vowel duration in their perception of vowel length by examining the perception of vowel length pair of words in three languages; Arabic, Japanese and Thai. In Japanese language, [19] found that duration is the main cue in the perception of vowel length, while F0 is a secondary feature. In French language, [20] found that native French language does not depend on the difference in vowel durations as a perceptual cue, while the French learners of English native language do.

In English,. [21] found that the change in vowel durations has a little weight as a perceptual cue in identifying the quality of the vowel. This was an opposite result of. [22]' study who found that duration is an important perceptual cue.

Some studies like [23] found that vowel type (backness and height) may affect its duration. They found that low back vowels are affected by change in duration, while high front vowels were difficult to be affected.

.[24] came to the conclusion that Indonesian Language is syllable-timed language, they usually do not have lexical stress nor do they have vowel length contrast -as opposite to Arabic language- while the latter is stressed-timed language having a vowel length contrast.

2 EXPERIMENTAL STUDY

The present study thoroughly examines the approach of the experiment adapted. The approach is tackled at two main parts; the first one is manifested into three processes done on recorded words: (a) lengthening the short vowel, (b) shortening the long vowel, and finally (c) merging a part of the long vowel into the short one to make the later long. Acoustic measurements of the basic features (duration, F0, F1, F2, F3 and jitter/shimmer features) of the vowels are performed before and after the above mentioned three processes. The second part is a subjective–perceptual test for the results of the three processes.

A.. Procedures:

-1 Selecting the Linguistic Samples:

The material selected for the experiment is made up of six pairs of monosyllabic words /CVC/ /CVVC/ where the /V/ and the /VV/ represent both short and long basic Arabic vowels / æ / /i/ /u/ and the /ˤ/ /ɛ/ /o/ as used in colloquial Egyptian; the difference in the length among such vowels do have a great role in differentiating meaning.

Vowel	Long V: CVVC		Short V: CVC	
	Word	Meaning	Word	Meaning
/æ/	\dææb\	Melted	\dæ bb\	To step feet
/ɛ/	\bɛɛt\	House	\bɛtt\	A girl
/i/	\diib\	Bear	\dib\	No meaning
/o/	\toob\	A dress	\tob\	The Arabic infinitive for "to ask forgiveness"
/u/	\kuub\	A glass	\Kub\	No meaning
/ˤ/	\tˤˤb\	To be well done	\t̲ˤbb\	To fall down (in colloquial Egyptian)

Table II- I

- Closed consonants /b t d k/ are tested at initial and final positions of syllables.
- Choosing word pairs with long high vowels /u/ and /i/ are reported to have no short counterparts in Arabic except when the quality is changed. Accordingly, subjects were trained to pronounce the short counterpart of the vowels in the same quality form even when the resulting words make no sense, e.g., /dib/ and /kub/.
- The words were pronounced by five male subjects between 22- 35 years old with no speech deficiency.
- The researcher explained the experiment to the subjects and trained them on how to pronounce the words before the actual recording; they were asked to pronounce each pair several times, and only correct pronunciations in context are selected.

-2 Recording the Linguistic Data:

The words were recorded in the department of phonetics and linguistics, University of Alexandria using a Computerized Speech Lab (CSL), 4500 kay elemetrics. The material of the experiment is hundred and eighty samples for the aforementioned words, including 5 subjects X 3 times of repetition for each word X 12 words (6 short vowels and 6 long ones).

-3 Acoustic Analysis:

- Programs used for Analysis:

Acoustic analysis of the selected recordings was done using the following speech acoustic analysis programs:

- ➢ Computerized speech lab (CSL), 4500 kay elemetric [being the main program of recording and analysis]
- ➢ Praat version 6005- win 32 for making the three processes, shortening, lengthening and merging by using copy , cut, and past options.
- ➢ E section, speech filling systems [SFS] for the LPC analysis.
- ➢ Winsnoori speech analyzer, version 1.34. is adapted in this respect to test the clarity of the merging process for the listener.

- Analysis Steps and Measurements:

Some measurements for the vowels were done before applying the aforementioned three processes: (duration, I, F0, F1, F2, F3 and jitter/shimmer).

Basic Vowels (short/ long/ duration measurements):

- Measuring the average duration of the basic short vowel in /dæb/ [60 – 100 msec.] of all the subjects with their repetitions (for all the vowels).
- Measuring the average duration of the basic long vowel in /dæ:b/ [160 – 250 msec.] of all the subjects with their repetitions (for all the vowels). (see Figures: 1-2 A/ 1-2 B)
- **P.S.:** The average (100 msec.) of short vowels and (250 msec.) of long vowels were chosen to help applications on shortening and lengthening processes.

Spectrogram analysis of long vowel in the word /dæ:b/ and the short one in /dæb/

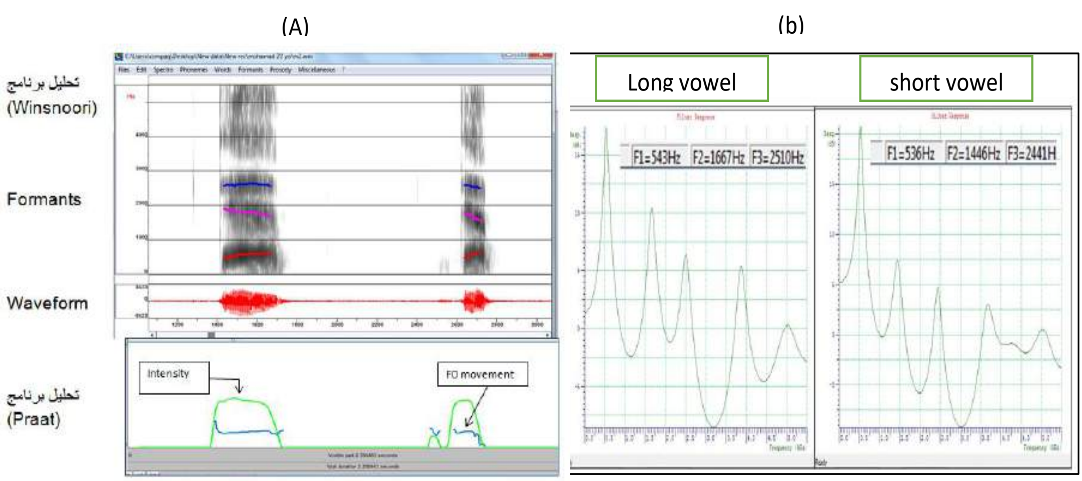


Figure 2-1 A - spectrogram and the waveform of the long and the short basic forms of the vowel /æ/. According to the steps followed, the duration of the short vowel and its fundamental frequency and [F1/F2/F3] in /dæb/ as well as that of /dæ:b/ were measured. Differences in the intensity of the two vowels and the FO were measured by Praat. Figure 2-1-B: LPC spectrum of the long vowel in /dæ:b/(right) and the short one in /dæb/ (left as measured by EsectionSFS)

The first process: *Shortening a Long Vowel* Reaching the Length of a Short One:

This is done by deleting a part of the steady state of the long vowel to equal the length of the short one (see Figure 2-2):

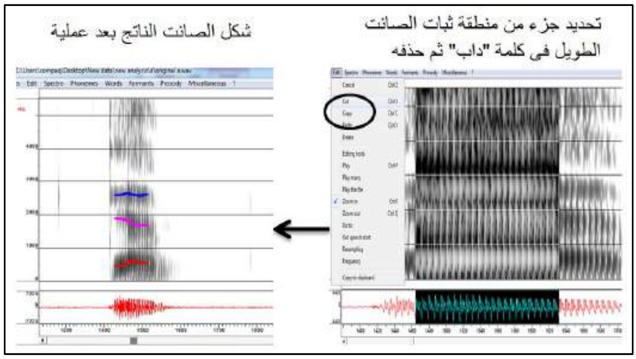


Figure: 2-2 indicates the steps of identifying, cutting a part of the steady state of the long vowel (right), and the result of the shortening process (left).

Comparing both the original short vowel and the result of the shortening process. See the following Figures 2-3 A (spectrogram, waveform, and F0/Intensity). See also Figure 2-3 B which shows the LPC spectrum.

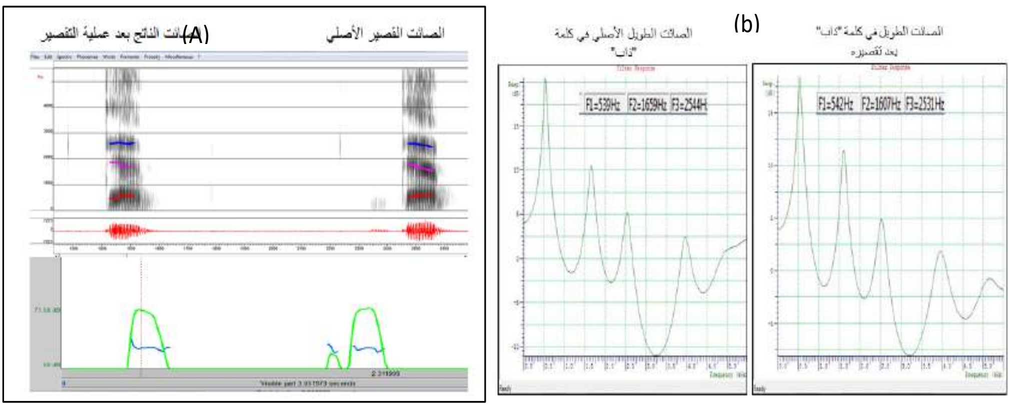


Figure: 2-3 A represents a comparison between the original short vowel /a/ (right), and the shortened vowel (left) in spectrogram, waveform and the (F0/intensity). Figure 2-3 B represents the LPC analysis of the original short vowel /a/ (left), and the shortened one (right).

In cases of pasting, adding or copying a part of the waveform, identifying the steady state of a vowel was done first; this part starts and ends with a full wave; cutting, pasting, adding and copying were not, hence, randomly proceeded in order to avoid sound clicks (see Figure 2-4).

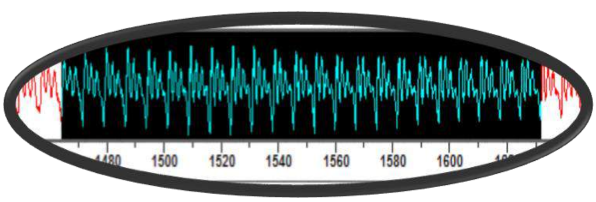


figure 2.4 select wave that will be deleted. this part starts and ends with a full wave; cutting, pasting, adding and copying were not, hence, randomly proceeded in order to avoid sound clicks.

<u>The second process: Lengthening the Short Vowel in its Same Wave:</u>

The short vowel /æ/ in /dæb/ was lengthened by pasting a section of steady state of its band ***several times*** to equal the length of the original long vowel. This procedure follows the coming steps:

- Identifying the section of the short vowel needed to be copied.
- Copying the selected section and pasting it in the same place several times (see Figures: 2-5/2-6 A, B)

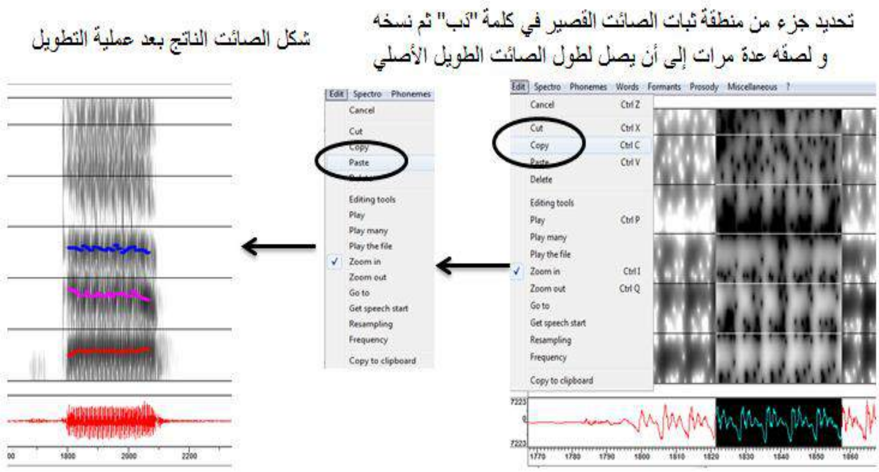
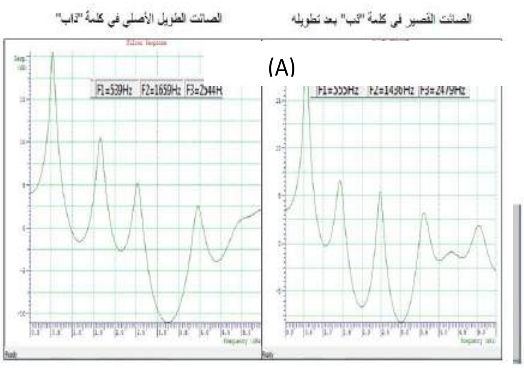


Figure: 2-5 indicates the process of identifying a part of the steady state of the short vowel /a/ for copying and pasting it into the same place of the short vowel several times. The left spectrogram represents the result.

(b)



(B)

1.
2.
3.
4.
5.
6.
7.
8.
9.

Figure: 2-6 A represents the spectrogram of the vowel /a/ after the process of lengthening (right), and the original short vowel (left) attached with their intensity and F0 movement. Figure: 2-6 B LPC spectrum of both vowels; after lengthening (right) and the original short one (left)

<u>The third process: Merging Part of the Wave of the Long Vowel to the Wave of the Short One to Make the Later long:</u>

The procedure was done as: Copying a part of the steady state of the long vowel and Paste it to the wave of the short one to make it long.

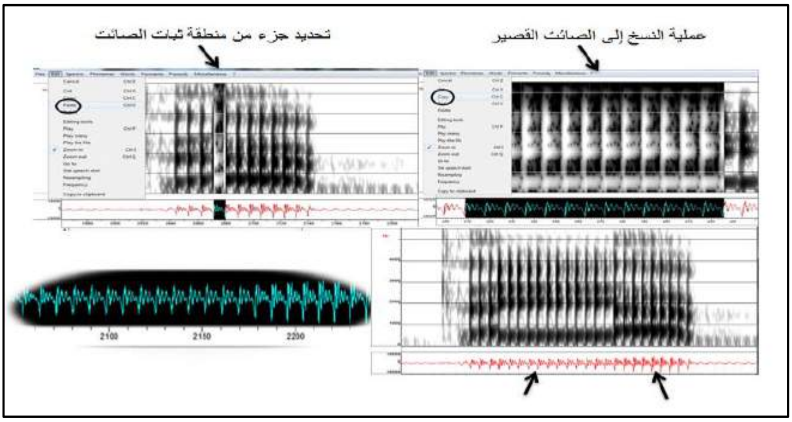


Figure: 2-7 represents the following steps of the merging processes: copying a part of the steady state of the long vowel (left) and paste it into the short vowel (right). The result of lengthening the short vowel is seen in the fourth spectrogram (down/ right), the difference between the two waveforms is noticed by the two arrows.

The following is the comparison between the original long vowel and the result of merging process.

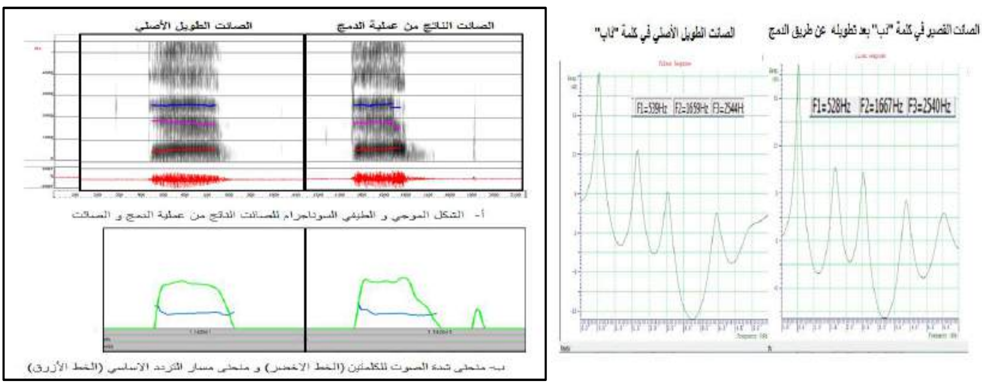


Figure: 2-8 A represents the comparison between: - The sonogram of the original long vowel (left) and the result of the merging process (right). – The intensity and F0 curves (the lower diagrams) for the two words /dæ:b/ /dæb/ and Figure 2-8 B represents LPC spectrum of both vowels; after lengthening by merging (right) and the original long one (left).

B. The second part of the experiment is the Perceptual Test:

-1 Subjects

- **Native Arabic listeners:**

Fifty participants between (20-35 years old), 25 females and 25 males, from undergraduate and postgraduate students at the department of phonetics were selected. The listening experiment was then done on the spoken material resulted from the previous acoustic procedure.

-2. Presenting Recorded Material to Native Listeners:

- The listening experiment took place in the studio of the department of phonetics, faculty of Arts, University of Alexandria. Each person was tested individually.
- Each pair of words was presented to the listeners, without pictures, using PowerPoint. The meaning of the words is given under each pair .
- The researcher explained the recorded material for the listeners, indicating that the difference lies in the length of the vowel, such as /dæ:b/ and /dæb/.
- The testing sheet of the listener includes six pairs of words. The participants were asked to choose and tick the word corresponding to what they had heard. Table (2-2)
- In applying all the previous procedures, the material was presented to the listeners in random order.

➢ **Nonnative Arabic listeners**

Data was presented to 40 nonnative Arabic listeners who were Indonesians (20 male and 20 female). They were 17 – 18 years old, their native language was Bohasa. They were Arabic learners at “The secondary extension institute of Arabic and Islamic studies in Egypt". They were at different levels (from level 1 to level 6). It is noteworthy that their native language does not have vowel length contrast. As the recorded material presented to native listeners (See: 2.2.2), aforementioned recorded material steps & instructions were presented to Indonesian listeners

Vowels		**بيت**	**بِت**
/ɛ/		**/bɛ:t/**	**/bɛt/**
		√	
O	**1**		√
S	**2**		√
L	**3**	√	
M	**4**	√	√

Table II-II: The testing sheet includes 4 cases of recording: Original record (O), Shortening (S), Lengthening (L) Merging (M).The listener was guided to consider numbers (1, 2, 3, 4) rather than symbols (O, S, L, M). The participants were asked to choose and tick the word corresponding to what they had heard.

3 RESULTS

This research is considered as a psychoacoustics study since it includes two main directions: perceptual and acoustical aspects. The first is the human ability to distinguish and classify what he/she hears in an abstract manner. the second aspect is the acoustical measurements and changes of the acoustic features in the production which is the concrete evidence for listener`s perception.

The research is based on an experimental approach; it clarifies the cognitive relevance of the vowel length of the listeners; natives and Arabic learners (whose mother tongue have no difference in vowel length), through their responses on processed vowel durations.

A. Acoustical results

-1 Durational features:

The results of this study showed that, the relative length of the short vowel was 70 - 90 msec. and the long vowel was 170-220 msec. with the average length of short to long was approximately 1: 2.5, while the results of several studies in Arabic showed that the length of the long vowel was twice as long as the short vowel [25], [26], [9], [13], [27]-[17]

The measurements of the relative vowel length differ from a study to another.

1. [9] study showed that the length of long vowel was is 300-350 m / s and the short vowel was 130-150 msec.
2. [4] The results of showed that the relative length of the short and long 95 m long was 112 msec.
3. [15] while the relative length of vowels was 117 msec. for short and 282 msec. for long.

The relative length of the different studies varies according to the spoken people and the vocal context of the spoken pronunciation as well as the different dialect and gender differences.

-2 Spectral information (vowel quality):

The study [13] showed that some long vowels have short counterpart that differ in quantity and in quality, the latter is a secondary feature

[28] study showed that vowels duration may affect vowel quality in some vowels, for example high closed vowels /i, u/ do not change their quality according to vowel length. While some other vowels such as /ʌ, ɛ, æ, ɑ / are affected by the length in their quality. This result was confirmed by the [29]' study applied to the Hungarian language. The study of [30] also examined the extent to which duration affects the listeners' perception of the length and quality of the vowel in Swedish language, through the process of shortening

some long vowels. The results showed that vowel duration is a distinct element in the Swedish language, and also there are some vowels are affected in their quality.

[9], studied duration and components of short and long vowels. His results showed that the change in the quality of the long and short (æ / æ æ) was more than the change in the U / UU and (I / II).

The present study enhancing these results, native listeners observed a quality change in some vowels in the shortening process i.e. as they described that the long vowel / æ / in the word [dæ:b] was changed to the vowel [ᵋ] when it was shortened so the resulted word was [dᵋb], while high vowels / i and u/ qualities kept on.

-3 fundamental frequency f0 for vowels:

Some studies have demonstrated the idea of the intrinsic pitch of vowels, [31] confirmed that the values of f0 vary somewhat by changing the quality of the pattern. High vowels [u, i] have higher f0 values than the rest of vowels.

[18] study, in the Japanese language, showed that there is a relationship between the fundamental frequency and the temporal absorption in the perception of vowel length, since the low fundamental frequency is associated with the long vowels.

Results of this study showed that Short vowels are higher in f0 values than long, especially high vowels [u, i]. even in shortening and lengthening processed vowels but the merged processed vowels were the lowest in f0 values.

Acoustic measurements of Jitter and Shimmer (which measure the amount of irregularity of cycle to cycle F0 and Intensity) indicated significant values of irregularity
of merging waveforms. (Figure: 3.1)

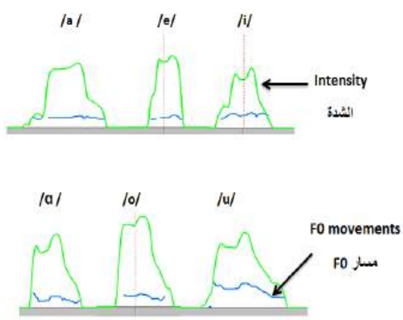


Figure: 3-1 Short vowels are higher in f0 values than long, especially high vowels [u, i]. even in shortening and lengthening processed vowels but the merged processed vowels were the lowest in f0 values, also F_0 contour shows irregularity

-4 F1 – F2 distance and the cognitive space:

[31] mentioned that [F1-F2 plot] is considered as a mental map that helps listeners to categorize the vowels precisely.

- In this study, formant frequencies of the perceived **original** short and long vowels were represented in the F1-F2 plot (see Figure: 3.2), where the long vowels are located on the edges of the F1-F2 plot map, while the short vowels were in the center.

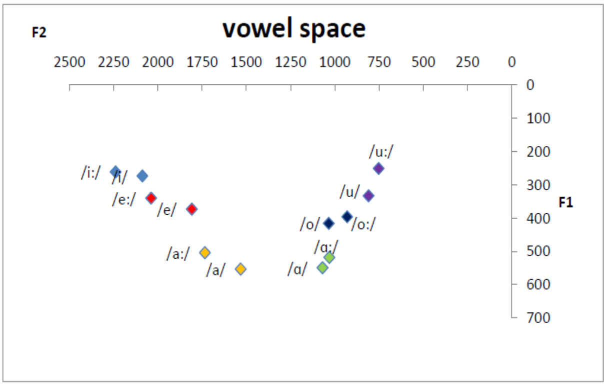


Figure:3.2 represents F1-F2 vowel space of the average values of the original short and long vowels.

- ➢ ***F1-F2 vowel space of the processed vowels; shortening, lengthening and merging* processes, it shows that:**
- ➢ *Long and lengthened vowels* specially the front vowels [æ, ɛ, i], are located on the edges of the F1-F2 plot map, while short, shortened and merged vowels tend to be in the center of the F1-F2 vowel space chart.
- ➢ *Back vowels* [u, o, ɑ], original and their processed ones tend to be close to each other with slight differences.

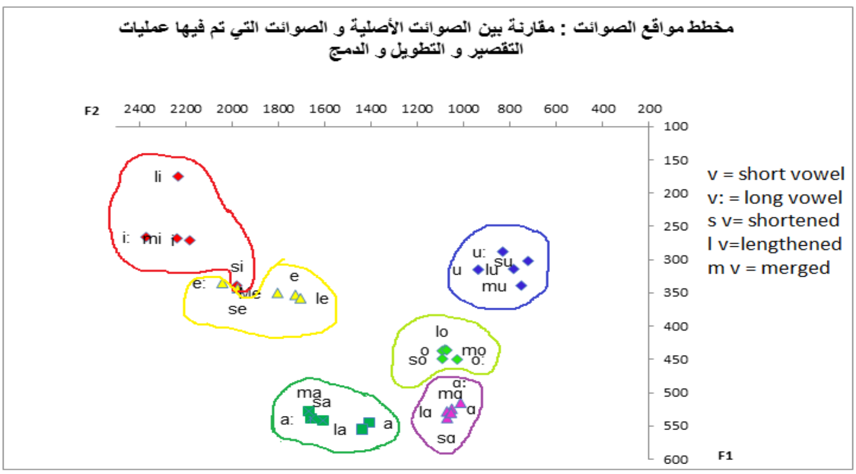


Figure: 3-3 indicates F1-F2 vowel space of original and processed vowels.

<u>*B. Perceptual Results*</u>

-1 Perceptual results Native Arabic listener

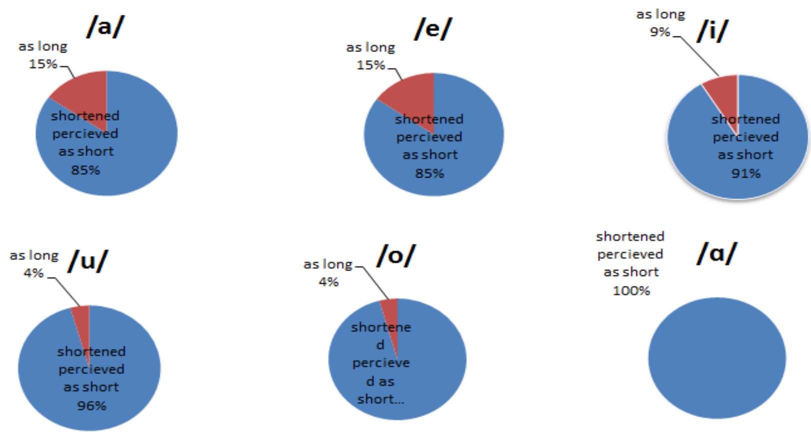
- *Shortening of the long vowels:*

In respect to the perceived ***length***; the shortened vowels were perceived as short ones with the percentage of about (96 – 100%) for the back vowels, and (85 – 91%) for front vowels.

In respect to the perceived ***quality***; 42% of the listeners perceived a change in quality when the shortened vowels compared with the original ones, especially the quality of /æ/ vowel. (See Figure: 3-5)

Perceptual results for shortening process

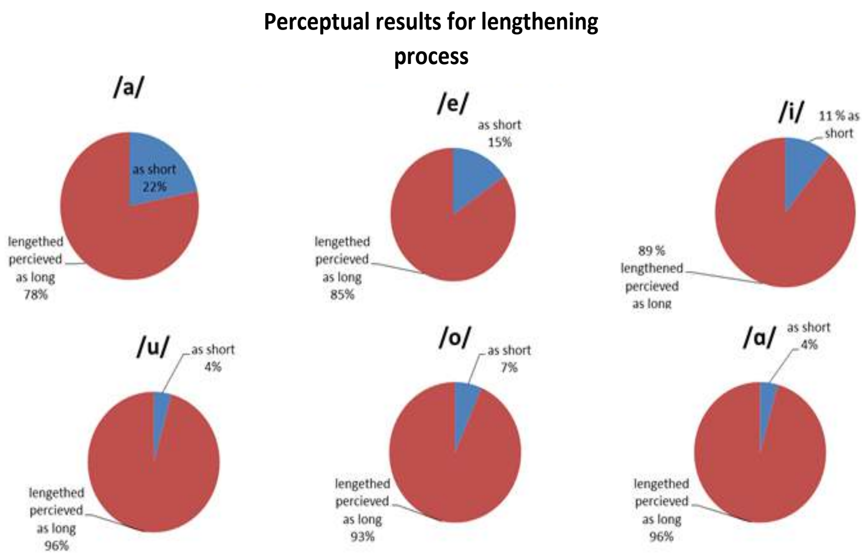
أ- النتائج الإدراكية لعملية تقصير الصائت



➢ *Lengthening of the short vowels*

In respect to the perceived ***length***; the lengthened vowels were perceived as long with high percentage values of about (78 – 96%). Back vowels were higher percentage than front ones.

In respect to the perceived ***quality;*** 40% of the listeners perceived a significant change between the lengthened vowels and the original long ones, especially for the front vowels. (see figure: 3-6)



➢ *Merging process*;

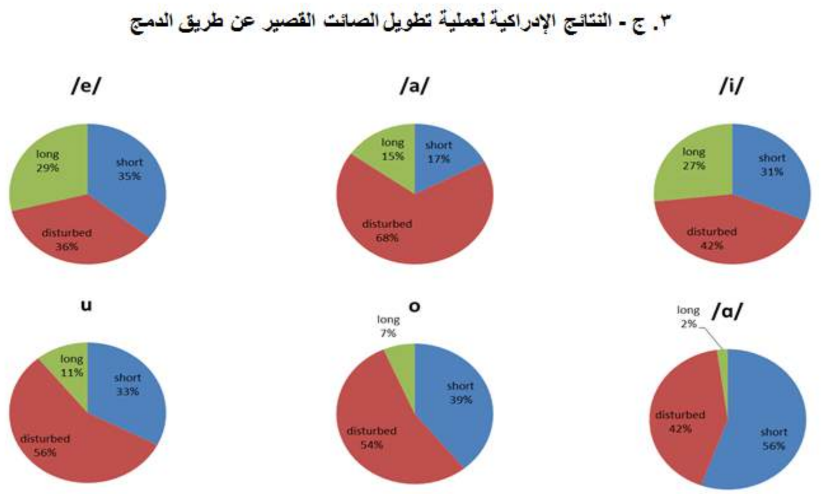
The listeners' judgments were at significant variance from the merged vowels quantity as well as their quality.

In respect to the perceived ***length;*** back merged vowels were perceived as long with percentage about (2%- 4%) and were perceived as short about (33% – 56%).

The front merged vowels were perceived as long about (15% – 29%) and as short about (17% – 35%). (See figure 3-7).

In respect to the perceived ***quality;*** all the merged vowels were perceived as disturbed and unclear- as the listeners' observations – with about (36 – 68%):

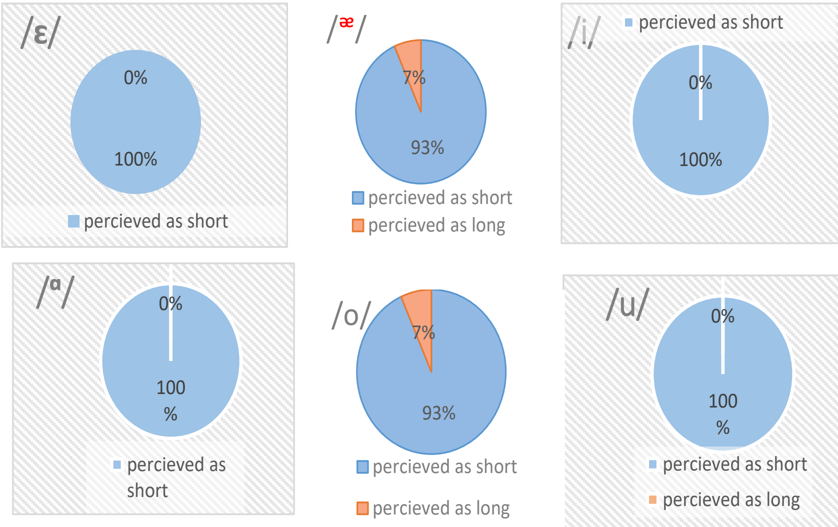
- The vowel /æ/ was the highest percentage, 68% of the listeners judged it as disturbed and described it as if a sequence of two different vowels.
- 56% of the listeners judged the back vowels [/u/ , /o/ , /ɑ/] as disturbed and described them as if there is a sound like glottal stop inside the vowel, e.g., / $t^{\alpha\alpha}$b / ------→ / t^{α} ? ab /, / kuub / ----→ / ku ? ub /.
- 42% of the listeners perceived the front vowels / i / as disturbed and described them as if there is a glottal stop inside the vowel.
- For the vowel / ɛ / about 29% judged it as long, 35% judged it as short, and 36% judged it as disturbed.



-2 Perceptual results for nonnative listeners

➢ *Shortening of the long vowels:*

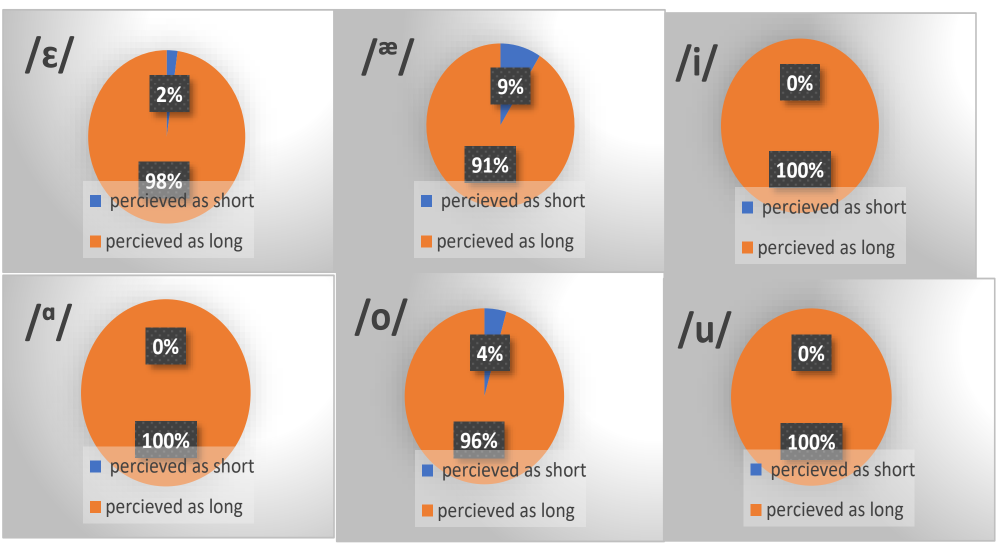
In respect to the perceived ***length***; nonnative listeners perceived the shortened vowels as short ones with high percentage values of about (93% - 100%). In respect to the perceived ***quality***; none of the nonnative listeners perceived change in quality when the shortened vowels compared with the original ones, (See Figure: 3.8).



- *Lengthening of the short vowels*

In respect to the perceived ***length***; nonnative listeners perceived the lengthened vowels as long ones with high percentage values of about (91 - 100%).

In respect to the perceived ***quality***; none of the nonnative listeners perceived any change in quality when the shortened vowels compared with the original ones, (see figure: 3-9).

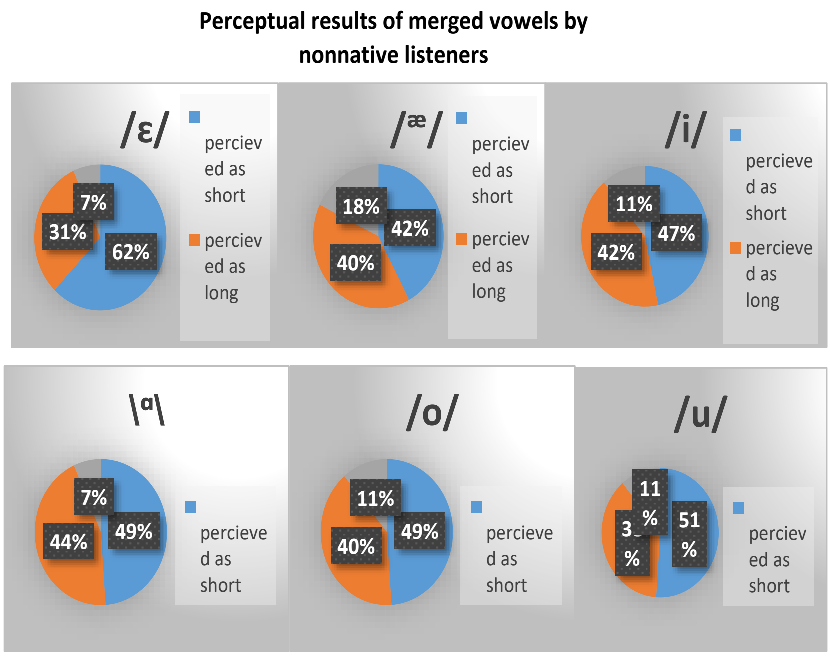


- *Merging process;*

In respect to the perceived ***length;***

- ✓ High percentage values range from (42% - 62%) of nonnative listeners who perceived all vowels lengthened by merging as a ***short*** vowel.
- ✓ Less percentage values range from (31%-44%) of nonnative listeners who perceived all vowels lengthened by merging as a long vowel (see figure 3-9)

In respect to the perceived ***quality;*** the lowest values percentage of nonnative listeners range from (7%-18%) who perceived all the merged vowels as disturbed and unclear- as the listeners' observations .



4 DISCUSSION

This part aims to correlate the results of acoustical measurements (i.e durational and spectral information) with listeners response to vowel length. As was mentioned in results, measurements proved that long and short vowels are different in their spectral features (f_0, F_1, F_2, intensity, pitch contour and waveform). Also, in the processes of shortening and lengthening, a slight change was occurred in these spectral features. For perceptual results, responses of native and nonnative listeners are naturally different.

A. Native Arabic listeners perception

When native Arabic listeners heard the original short and long vowels they could easily discriminate between them, this is because They have a prototype of long and short vowels stored in their minds according to vowel spectral features in their native phonetic system of Arabic.

- *In case of proceeded sample (shortening and lengthening)*
  - When we made the shortening of the vowel by deleting a part of its steady state, and also lengthening short vowel by reduplicating length of its own wave, some acoustic information was changed. Though this change was a slight change, but it had a perceptual correlate.
  - It was noticed that the Arabic native listener perceived back vowels differently from front vowels. In back vowels, most of them perceived proceeded shortened vowels as a short one in high percentages (96-100%) specially proceeded shortened vowel \ᵅ\ which was perceived as a short vowel by all native listeners. While in front vowels, lower percentage of native listeners (85- 91%) perceived proceeded short vowel as a short one. In case of shortening long vowels, perception of proceeded front vowels which are shortened is more resistant in their perception than proceeded back shortened vowel. This is because that when we made shortening of the front vowels, their quality has been changed. Unlike back vowels which keep its original

features. This is may be due to back vowels are reinforced by articulatory features like rounding, backness and emphasis for vowel \ɑ\, these features did not change during shortening or lengthening of the back vowel. While front vowels are characterized by easy articulation specially vowels \e, a\ which involve relaxation of articulators, and accordingly they are easily changeable in their quality

- ❖ Most of Arabic native listeners (85-100%) perceived proceeded shortened vowel as a short one and lengthened vowel as a long one i.e. they depended on duration for discrimination of vowel length. But many of them (40%) noticed a *quality change*. These results agreed with the studies [4], [5], [6], [7], [8] – [9] who cited that listener depend on spectral information (qualitative) as well as durational aspect (quantitative).
- ❖ Also, some of them perceived the proceeded shortened vowel as its origin long vowel but it is pronounced in a fast manner. For the proceeded lengthened vowel few of them described it as a short vowel but were pronounced in a prominent or slow manner.

- ➢ *In case of lengthening by merging two different waves:*
  - ❖ (lengthened short vowel was done by merging its wave to a wave of original long vowel), it was noticed that differences were resulted acoustically and perceptually.
  - ❖ Acoustically, the resulted merged vowel is characterized by disturbance in shimmer and jitter, pitch contour, intensity contour, waveform and spectrographic analysis.
  - ❖ Accordingly this disturbance was perceived by native listeners as a disturbed vowel i.e. the highest percentage of listeners described the merged vowel as a disturbed vowel in all six vowels.

Hence;

- ✓ The difference between long and short vowel is not only a difference in duration, but also it involves difference in spectral characteristics
- ✓ Arabic native listeners do not depend only on duration in perceiving vowel length, but also, they depend on spectral information too.
- ✓ This was strongly proved in the process of merging long with short vowel, in which listeners made what is called perceptual cue weighting. They tended to depend on spectral characteristics more than durational characteristics in their perception of the proceeded merged vowel. And accordingly, because the disturbance that happened in the spectral features of the proceeded merged vowel, they perceived it as a disturbed vowel.

.

B. Nonnative Arabic listeners

Perception of nonnative Arabic listeners (Indonesian listeners) was different from perception of native Arabic listeners. they depended basically on durational differences in their perception of vowel length.

- ❖ They perceived proceeded short vowel as a short vowel
- ❖ And perceived proceeded long vowel as a long vowel
- ❖ In merging vowel, the highest percentage of them perceived proceeded long merged vowel as a short vowel, also a lower percentage of them were perceived it as a long vowel. while the lowest percentage of them perceived it as a disturbed vowel

This was due to that Indonesian language has no long vowel in its phonetic system. It has only short vowel [32]. So, vowel length is not a distinctive phonemic feature in the Indonesian language.

They could perceive long and short Arabic vowels (and also shortened and lengthened vowels) correctly by depending only on duration of the vowel. They just hear a long timing and a short timing of the same vowel, without any notice that long vowel is a pattern that differed from the short vowel. unlike native speaker who depended on spectral information and noticed a quality change in some vowels and also noticed that the vowel identity did not change by shortening (as they described it was pronounced in a fast manner when it was shortened) and also in lengthening vowel (as they described it was pronounced in a stress or slow manner when it was lengthened.)

When nonnative listeners heard merged vowels, firstly most of them perceived it as a short vowel because they enclose what they hear to their native phonetic system, and because they have no long aspect of vowels in their native language, they perceived it as a short vowel. Secondly other group of them who perceived merged vowel as a long vowel, this was because that they only depend on duration and heard just a long timing vowel. The lowest percentage of them perceived it as a disturbed vowel – unlike the previous two cases – they could notice some spectral features that is presented in the disturbance that occurred in the process of merging. These results agreed with [33] study who investigated the acquisition of long and short vowels of Japanese language by 20 Indonesian speakers, results showed that They made few errors in listening to short vowels, however, many errors in long one, and final positioning was very difficult. [32] assumes that Indonesian perhaps will generalize the English vowels with the Indonesian vowels that they have where only short vowels exist.

Indonesian has only six vowels, which are [I], [ə], [ɑ], [o], [u], [e], and three diphthongs and [32] showed that Pronunciation problems faced by foreign language learners are caused by differences found between the learners" language and the target language. The students have tendency to pronounce English vowels as the way they pronounce vowels in their mother tongue. The students also tend to pronounce the word longer if the word has double same vowels, even though the word truly is not pronounce long. The students also tend to pronounce every word shortly if they have only one vowel. Likewise; in this research when Indonesians listeners heard long duration, they respond it as long vowel, and when they heard short duration they respond it as a short one, but in the case of listening to the merged vowels they responded it as a short vowel because only short vowel exists in their native phonetic inventory.

5 CONCLUSIONS:

Conclusions of this study could be summarized as follows:

- ✓ ***In Arabic, the acoustic differences between long and short vowels are not only durational features, but also in spectral information. i.e., quantity and quality.***
- ✓ ***Perceptually, however, native Arabic listeners tend to depend on length of the vowels as well as spectral information.***
- ✓ ***Vowel length is not a distinctive phonemic feature in some languages like Indonesian language. Accordingly, nonnative listeners of Arabic could discriminate between long and short vowel by depending only on duration (unlike native listeners of Arabic).***
- ✓ ***From the afore-mentioned perceptual percentages, shortening and lengthening of the vowel from their own waveform showed a significant change in quality by natives, but no significant values by non-natives.***
- ✓ ***Merging two different wave qualities showed high significant values of perceptual percentages by natives, while non-natives showed insignificant ones.***
- ✓ ***The high vowels /i/ and /u/ are more resistant in changing than other vowels by natives and nonnatives. This is proved in the original vowels as well as all kinds of processed vowels.***
- ✓ ***Finally, in second language acquisition, it is very important to consider influence of native language on the listener who learn another foreign language***

ACKNOWLEDGMENT

Great thanks to Phonetics and linguistics department specially to our professor Dr. Mervat Fashal for her guidance and regular support. And Dr. Sameh Alansary - the head of the department- for his permanent encouragement . Another thanks to students of phonetics department who participated in this this experiment, and Indonesians students in "The secondary extension institute of Arabic and Islamic studies in Egypt".

BIOGRAPHY

Eman Mohammed Mohammed Kassem

- ***Demonstrator*** in phonetics and linguistics department 2003, **M.A in Experimental Phonetics "Influence of punctuation marks stemming of suprasegmental features in understanding read-text " Psycho-acoustic study 2010.**
- ***Assistant lecturer*** in phonetics and linguistics department 2011
- **PHD in Experimental Phonetics "Contributions of Nasals and Vowels to Speaker Identification in Arabic"2014**
- ***Lecturer*** in phonetics and linguistics department 2015
- ***Academic Advisor*** in the Forensic Phonetic and linguistic Program in 2018.
- ***Academic supervisor*** in the Forensic Phonetic and linguistic Program from September 2019 untill now.
- ***Teaching experience in*** different subjects; Experimental Phonetics, Speech Perception, Acoustic Phonetics, Physiological Phonetics and Auditory Phonetics.

Lamiaa Mohammad Tawfik Aldawy

- ***Postgraduate researcher*** in phonetics and linguistics department
- **B.A in phonetics and linguistics department in 2010**
- **M.A in Experimental Phonetics "perception of durational differences of vowel length in Arabic an experimental study "2016.**

الملخص العربي

إدراك طول الصوائت فى اللغة العربية عند المتحدث الأصلى وعند متعلمى اللغة العربية من غير الناطقين بها

دراسة تجريبية

*1 إيمان محمد قاسم, **2 لمياء محمد توفيق

1*قسم الصوتيات و اللسانيات , كلية الآداب, الشاطبي , جامعة الإسكندرية

1Eman.kassem@alexu.edu.eg

2**قسم الصوتيات و اللسانيات , كلية الآداب, الشاطبي , جامعة الإسكندرية

2lamyaatawfik@hotmail.com

الملخص

يعد طول الصائت المدْرك - في العديد من اللغات - من العلامات الفارقة للتمييز الصوتي؛ حيث يتضح ذلك في إدراك المستمع لنفس المنطوق بطريقتين مختلفتين إعتماداً على الطول الزمني للصائت ونوعية الصائت.

قام فراي (1956) بإلقاء الضوء على الفرق بين الطول الزمني والطول الإدراكي للصائت، حيث أن الطول الزمني للصائت هو الوقت الفعلي الذي يستغرقه الصائت أثناء نطقه، بينما الطول الإدراكي للصائت هو الطول المدرك للصائت من قبل المستمع.

إن اختلاف الطول المدرك للصائت في اللغة العربية يؤدي إلى تغييرات في المعنى؛ ولذلك تعتبر علامة فارقة في لغتنا العربية. ولكن إلى أي مدى يمكن أن يكون هناك إختلافات إدراكية وفيزيائية (قياسية) بين الصوائت القصيرة والطويلة؟ وما هي السمات الصوتية التي قد يعتمد عليها المستمعون العرب والمستمعون الإندونيسيون للتمييز بين الصوائت القصيرة والطويلة؟

يهدف هذا البحث إلى دراسة إدراك كل من مستمعي اللغة العربية الأصليين والمستمعين الأجانب لطول الصائت في اللغة العربية من خلال دراسة تجريبية؛ وتوضح الدراسة أيضاً مدى إعتمادهم على الصلة المعرفية المخزنة/ المدركة لديهم لطول الصائت، من خلال تسجيل إستجاباتهم على طول الصائت المعالج.

REFERENCES

[1] *Fry, D.B. (1956). 'Duration and intensity as physical correlates of linguistic stress'. Journal of the Acoustic Society of America 27, 765-768*.

[16] *Escudero, P. (2001). The role of the input in the development of L1 and L2 sound contrasts: language-specific cue weighting for vowels. Proceedings of the 25th Annual Boston University Conference on Language Development, Cascadilla, pp. 250-261.*

[17] Tsukada, K., (2009). An acoustic comparison of vowel length contrasts in Arabic, Japanese, Thai: Durational and Spectral data, International Journal on Asian Language Processing; 19:127-138.

[18] Tsukada, K., (2013).Vowel length categorization in Arabic and Japanese: Comparison of native and non-native Japanese perception, Speech, Language and Hearing, 16(4):187-196.

[19] Kinoshita, K., Behne, D., M., and Takayuki, A., (2002). Duration and F_0 as perceptual cues to Japanese vowel quantity, 7th international conference on spoken language processing, Denver, Colorado, USA: 757-760.

[20] Gottfried, T., L and Beddor (1988). Patrice speeter: Perception of temporal and spectral information in French vowels, Haskins laboratories

[21] Sawusch. J., R., (1996). Effects of duration and formant movement on vowel perception, ICSLP, vol.4: 2482- 2485,

[22] Kassem, E., M. (2014), Contributions of nasals and vowels to speaker identification, PhD thesis, Alexandria university.

[23] Abramson, A.S. & Ren, N. (1990). Distinctive vowel length: Duration versus spectrum in Thai. Journal of Phonetics, 18, 79-92.

[24] Dauer,R.M.(1983), "stress-timing and syllable-timing reanalyzed", Journal of Phonetics,11:51-62.(in reference at the end)

[25] *Wahba.k, Sh., (1988) ,The acoustic analysis of colloquial egyptain Arabic vowels: An experimental study, M.A thesis, Alexandria*
university.

[26] Saadah, E (2011). The production of Arabic vowels by English L2 learners and heritage speakers of Arabic. PHD.

[28] Hillenbrand, J., M., Clarkm, M.J., Houde, R.A. (2000). Some effects of duration on vowel recognition, J Acoust Soc Am; 108(6):3013-22.

[26] Mady, K & White, L (2008). The long and the short and the final: Phonological vowel length and prosodic timing in Hungarian, Brazil, Speech Prosody, ISCA Archive: 363-366.

[30] Hadding-Koch, K. & Abramson, A. S. (1964). Duration versus spectrum in Swedish vowels: Some perceptual experiments. Studia Linguistica, XVIII, 94-107.

[31] Hayward, K., (2000). Experimental Phonetics -. England, pearson Education Limeted.

[32] *Riadi, A, Rufinus, A, Novita, D, (2013). STUDENTS' PROBLEMS IN PRONOUNCING SHORT AND LONG ENGLISH VOWELS, English Education Study Program, Language and Art Education Department Teacher Training and Education Faculty of Tanjungpura University, Pontianak*.

[33] Franky R, YOKOYAMA, N, ISOMURA, k, USAMI,Y, KUBOTA, Y., (2012). The Acquisition of Japanese Vowel Length Contrast by Indonesian Native Speakers: Evidence from Perception and Production, Journal of the Phonetic Society of Japan, Volume 16 Issue 2 Pages 28-39 available at https://www.jstage.jst.go.jp/article/onseikenkyu/16/2/16_KJ00008228944/_article

المراجع العربية :

[2] بريتل مالمبرج. الصوتيات. \ ترجمة محمد حلمي هليل- الإسكندرية : عين للدراسات و البحوث الإنسانية و الإجتماعية, 1994

[3] محمد صالح الضالع. الصوتيات والفونولوجيا مقدمة معاصرة للقارئ العربي.- جامعة الإسكندرية, 2003

[4] سعد عبد العزيز مصلوح. التناسب الزمني بين الحركات القصيرة و الطويلة دراسة صوتية معملية في القافية العربية. مجلة معهد اللغة العربية.ع 2, 1984

[5] أحمد مختار عمر. دراسة الصوت اللغوي.- القاهرة: عالم الكتب 1991.

[6] غالب فاضل المطلبي. في الأصوات اللغوية دراسة في أصوات المد العربية.- العراق: دائرة الشئون الثقافية و النشر, 1984

[7] عبد الحميد زاهيد. حركات العربية دراسة صوتية للتراث العربي/ تقديم التهامي الراجي الهاشمي.- مراكش, المطبعة و الوراقة الوطنية, 2005

[8] محمد سالم الرجوبي. الحركة الإعرابية بين القيم الصوتية و القيم الدلالية. مجلة الجامعة الاسمرية, ع 20, ليبيا, 2014

[9] سلمان حسن العاني. التشكيل الصوتي في اللغة العربية فونولوجيا العربية.- جدة. النادي الأدبي الثقافي, 1983

[10]   عبد الفتاح عبد العليم البركاوي. مقدمة في علم الأصوات العربية. القاهرة – كلية اللغة العربية, 2004
[11] أبي الفتح عثمان ابن جني / تحقيق د. حسن هنداوي. سر صناعة الإعراب ,الجزء الاول.- دمشق: دار القلم 1993.
[12] كمال بشر. علم الأصوات.- القاهرة : دار غريب للطباعة والنشر والتوزيع , 2000
[13] منصور محمد الغامدي. الصوتيات العربية.- الرياض: مكتبة التوبة, 2001
[14] محمد احمد زكي. المَدُّ في العربية دراسة صوتية موجزة. مجلة جامعة بابل للعلوم الصرفة والتطبيقية. م 19, العراق, 2011
[15] يحيى علي أحمد. طول الحركة في اللغة العربية وعلاقته بالبنية المقطعية. مجلة جامعة دمشق, م 29, ع 3+4, 2013.
[27] إبراهيم أنيس. الأصوات اللغوية.- القاهرة : مكتبة انجلو, 1971

An Investigation of the Correlation between Perceived Pauses and Syntactic Structures

Israa Elhosiny*1, Mervat Fashal*2

*Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt

[1]israa_elsayed@yahoo.com

[2]Mervat.fashal@alexu.edu.eg

Abstract— **Natural-sounding speech synthesis requires close control over the temporal structure of the speech flow. Therefore, the relationship between prosodic phrase boundaries in terms of pausing and the syntactic structure has been investigated in read dialogue and continuant speech in Arabic. Both the speakers' production and the listeners' perception of pausing are considered and mapped to the syntactic structure. In order to describe the correlation between pauses and phrase boundaries in the proposed speaking styles, recall and precision rates were counted.**

Keywords: **speech production, speech perception, phrasal categories**.

1 INTRODUCTION

> ***"In the speech sound wave, one word runs into the next seamlessly; there are no little silences between spoken words the way there are white spaces between written words. We simply hallucinate word boundaries when we reach the end of a stretch of sound that matches some entry in our mental dictionary."***
>
> **(Pinker, 1995*)***

Pauses are useful for the speaker [1] and for the listener [2]. In fact, there is a positive correlation between the need for pause time during the speech sequence and the level of processing required by the task. Lists of digits and letters presented at fast rates or at low signal-to-noise ratios are recalled more accurately when pauses are inserted into the stimulus sequence. Similarly, speakers tend to increase the ratio of speech to pause time as the utterance becomes more complex.

Grammatical pauses appear to be more associated with different types of processing than the non-grammatical pauses. They occur between clauses and used for structural and semantic long-term planning in speech production. By contrast, non-grammatical or within-clause pauses are concerned with last-minute word selection [3].

Prosodic features are not sufficient, to analyze and generate the structure of the speech, but also grammatical analysis on many linguistic levels. Several researchers have discussed how prosody, morpho-syntax and discourse structure are related to each other. Linguistic structure has been shown to play a vital role in pausing strategies which signal information flow of the utterance, thereby helping the listener to interpret the message uttered by the speaker [4]. Swerts and Geluykens [5] found that speakers in monologues use pauses of various lengths to signal information flow in terms of topic structure. Shriberg et al [6] reported that new topics are often realized by some combination of silent pauses, low boundary tones and/or pitch range resets in English. Also, Hirschberg [7] argued that phrases introducing a new topic can be characterized by an initially wider pitch range preceded by a longer pause, and on average they are louder and slower than other phrases. Van Donzel [8] studied prosodic features of discourse boundaries for Dutch on the basis of clause, sentence and paragraph division, as well as the prosodic features of information structure in terms of the New–Given taxonomy. She found that discourse boundaries in spontaneous speech are realized by silent pauses and high boundary tones. These studies show that there is a relationship between prosody and (at least) higher linguistic structure, such as discourse in terms of topic, theme, and New–Given taxonomies.

Moreover, several researchers have investigated the relationship between morpho- syntactic structure and prosody. Most of the studies deal with the automatic prediction of prosodic phrase boundaries, given some linguistic information, used in text–to– speech systems. Some studies show that full syntactic analysis is not needed for the prediction of prosodic boundaries, while others claim the opposite. For example, in text–to–speech systems, phrase breaks are often predicted by distinguishing between content and function words. For prosodic phrase boundary detection, detailed but incomplete syntactic analyses were used by Bachenko and Fitzpatrick [9] by implementing the Phi rule-based algorithm developed by Gee and Grosjean [4]. Wang and Hirschberg [10] as well as Ostendorf and Veilleux [11] used PoS and syntactic constituent structure together with some acoustic information (such as pitch accent, phrase duration, and position to the last break) to predict phrase breaks. However, Ostendorf and Veilleux reported that good phrase prediction can be achieved without using any detailed PoS, or syntactic information. Taylor and Black [12] assigned phrase breaks on the basis of part-of-speech sequences only, although they suggested that syntactic parsers giving reliable parse trees might facilitate phrase break assignment. These studies show that prosodic phrase boundaries do not necessarily correspond to syntactic phrase boundaries. Most of the researchers agree that there is a relation between prosody and syntactic structure on one hand, and between prosody and discourse structure on the other hand. However, most of the studies performed on this topic investigate one of these relations either for non–spontaneous or for spontaneous speech.

The aim of this study is to investigate some aspects of the relation between the prosodic, and syntactic structure in spontaneous as well as in non–spontaneous speech. Additionally, both the speakers' production and the listeners' perception of pausing are considered and mapped to the linguistic structure. For spontaneous speech, we use a continuant speech, and for non-spontaneous speech, we study the acted dialogue for a children story. We investigated the pausing strategies in the speaking styles in terms of syntactic phrasal boundaries. Next, we will describe the data and method used for investigating the relationship between pausing and linguistic structure.

2 DATA AND METHOD

In order to investigate, primarily, the relation between pauses and syntactic structure, a modern standard Arabic speech sample is the most suitable material. The following subsections presents how the speech sample collected, prepared and analyzed.

A. Speech Material

As the Modern Standard Arabic; MSA, is the representative method used in Arab world, the researcher preferred to start the syntax-acoustic interfaced research by the MSA not by colloquial Arabic. MSA considered as more regular then colloquial. We selected an acted children story as a dialogue to represent the non-spontaneous speech. Moreover, a continuant speech has been selected to represent the spontaneous speech. The speech sample of dialogue is segmented into sentence; a sentence is defined as a turn. The first 10 sentences were selected which represent 106 words. For spontaneous speech, the first 9 sentences from a continuant speech have been selected. The sentence boundaries have been detected by the length of sentence; the maximum length for the sentence is 18 words. The number of words in spontaneous speech was 171 words.

B. Syntactic boundaries annotation

The research assumption is, pauses may split some phrasal units in Arabic. Speech sample has been written as a text to enable syntactic annotation. Table (1) expresses some annotated examples of the expected phrasal boundaries (bold font). The last word from each expected phrase has been marked beside its final word position. Around 30% of sentences' words in the overall data. Both speaking styles, have been marked manually as phrasal boundaries. The ratio between marked and unmarked phrasal boundaries has been represented as in figure (1).

TABLE 1: EXAMPLES FOR MARKED PHRASES.

phrase	*Function*	*example*
NP: Vocative particle + **NP**	Vocative style	أيتها الحمامة **الضعيفة**
VP: V +$_{obj\ pron}$ + **N**	Verbal phrase	أكلتكم **جميعا**
NP: N + Prep + **NP**	2nd Verb argument	القى إلى واحدا من **صغارك**
VP: V + **PP**	2nd Verb argument	صعدت **إليكم**
JP: J + **A**	absolute Object	ماهر **جدا**
PP: Prep + N + **NP**	Predicate argument	ماهر في تسلق **الأشجار**
VP: V + PP + **NP**	Temporal noun	آتي إليكم صباح **غد**

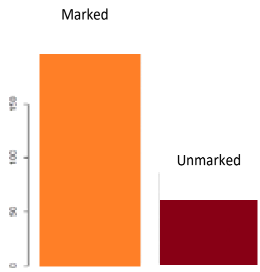


Figure 1: Marked vs unmarked phrase boundaries.

C. Acoustic Detection of Pauses

In the traditional linguistic definition, normal speech flow is considered to be interrupted by a physical pause whenever a brief silence can be observed in the acoustic signal (i.e., a segment with no significant amplitude). Which exact duration of the silence is considered sufficient for the constitution of a physical pause depends on its linguistic context [13]. Intra-segmental pauses are related to the occlusions of the vocal tract in normal speech production. Example: In the word "happy" as in Figure (2), the pause component of the Voice Onset Time (VOT) for the consonant /p/ corresponds to a silence of 96 ms [14].

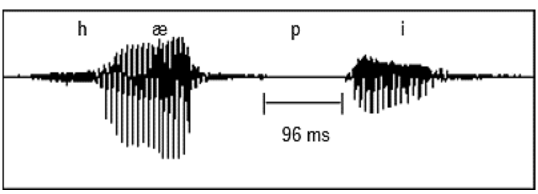


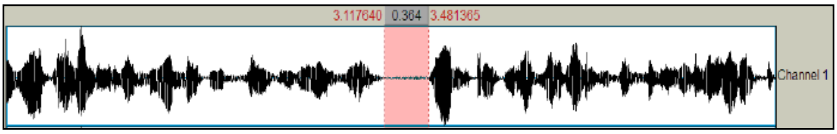
Figure 2: The acoustic signal of "happy" showing an intra-segmental pause.

Both speaking styles samples have been segmented into spoken sentences in isolated .MP3 formatted files. A database has been built to store sound files and their written sentences with their IDs. In addition, the acoustic signals for each sentence have been stored as a picture file. Sentences were segmented into written words to be stored in the data base. The database designed as it is the best way for data handling, retrieval, exploration and calculation. The words table inside the database is linked to sentences table to store information about phrasal boundary, silence after the stored words, etc.

Observed silence intervals; more than 100 ms, are measured by Praat for each sentence' words and stored beside word it proceeded as in figure (3).

words | sentences

ID	sound file	sentence	Pattern	wave form	high_pitch
1	Package	أيتها الحمامة الضعيفة ألقي إلي واحدا من صغارك وإلا صعدت إليكم وأكلتكم جميعا فأنا ماهر جدا في تسلق الأشجار.	NP VP VP[VP COO VP] NP[NP ADJP[ADJP PP]]	Package	368



words | sentences

ID	Word	duration_pause	constituent
2	الحمامة		
3	الضعيفة	0.364	NP
4	ألقي		
5	إلي		
6	واحدا		
7	من		
8	صغارك		NP – COO
9	وإلا		
10	صعدت		
11	إليكم		PP – short – EOF (VP)

Figure 3: database and speech information storage

3 Results and Discussion

In order to hold a comparison of pauses between the two different speaking styles, we investigate three different dimensions of communication; production, perception, and context. The following subsections discuss the relationship between the production of pauses and the linguistic context in which pauses appear, as well as the perception of the pauses and the linguistic environment in which people actually perceive them for the two speaking styles: acted dialogs and continuant speeches.

D. Production and linguistic features

In this section, the silent intervals characteristics detected in the two speaking styles will be described with special attention directed to the syntactic context in which pauses appear. The mean duration and word/pause ratio of silent intervals are calculated as in table (2). There are differences in the duration and frequency of pauses between the two speaking styles. Hence, the time it takes to pronounce a word on average differs between the styles suggesting greater variation in speech tempo. Our results, indicating the variation of pausing patterns across speaking styles.

TABLE 2: FEATURES OF ACOUSTIC PAUSES

Speaking style	*Pause Duration*	*Word/Pause Ratio*
Acted dialog	435.6 ms	4.25
Continuant speech	963 ms	3.24

For describing the correlation between silent intervals and supposed phrasal boundaries in the two speaking styles, recall and precision rates were counted. Precision and recall are calculated in terms of positive and negative classifiers. The predicted condition can be one of the following:

TN / True Negative: case was negative and predicted negative

TP / True Positive: case was positive and predicted positive

FN / False Negative: case was positive but predicted negative

FP / False Positive: case was negative but predicted positive

Precision and recall are then defined as in Definitions (1) and (2) [15]:

$$(1)\ \boldsymbol{Precesion} = \frac{TP}{TP+FP}$$

$$(2)\ \boldsymbol{Recall} = \frac{TP}{TP+FN}$$

Recall (R) describes the percentage of acoustic pauses that appear in boundary positions, see Definition (3). Precision (P), on the other hand, gives the percentage of phrasal boundaries that corresponds to pauses in the speaking styles, see Definition (4).

$$(3)\ \boldsymbol{P} = \frac{\textit{Turly predicted pauses}}{\textit{Truly predicted pauses} + \textit{wrongly predicted pauses}}$$

$$(4)\ \boldsymbol{R} = \frac{\textit{Turly predicted pauses}}{\textit{Truly predicted pauses} + \textit{unpredicted pauses}}$$

The correlation between silent intervals and phrasal boundaries for each speakingstyle is shown in Figure (4). It is obvious that phrasal boundaries and pauses do not always coincide, and we find clear differences between the speaking styles. In the acted dialog style, pauses often appear at a phrasal boundary positions (shown by the high recall = 100%). In the continuant speech, greater number of pauses existing in a non-phrasal boundary positions (lower recall= 93.6%). On the other hand, looking at the phrasal boundaries and their acoustic correlate in terms of silent intervals (i.e. the precision rates) we find that in dialog, 6 predicted phrasal boundaries were not followed by a pause (80% precision), although the continuant speech do this to a greater extent (88% precision)

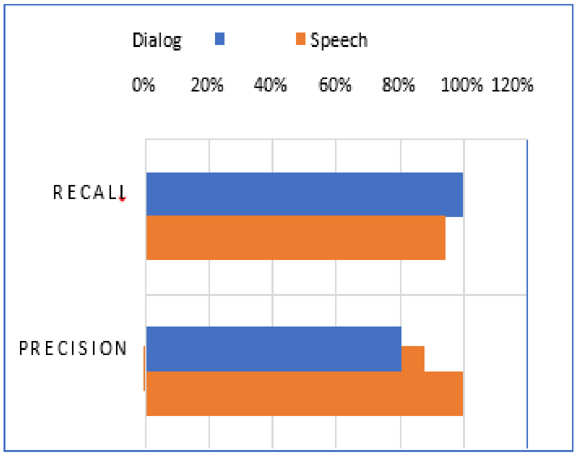


Figure 4: Recall and precision rates for acoustic pauses and phrasal boundaries in acted dialog and continuant speech.

For the Syntactic context of the acoustic pauses, the majority of silent intervals can be found at sentence boundaries. In acted dialog, pauses appear entirely at sentence boundaries and between phrases, e.g. in after noun phrases (21%), adverb phrases (6%), verb phrases (12%). While in continuant speech, after noun phrases (26%), adverb phrases (2.2%), verb phrases (24.4%) and prepositional phrases (17.7%).

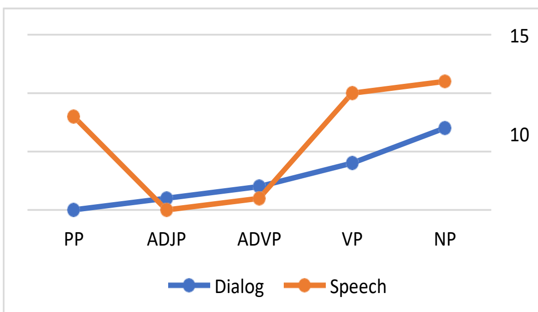


Figure 5: The ratio between occurrences of phrase boundaries in both speaking styles.

E. Perception and linguistic features

In order to investigate how often and in what linguistic context people actually perceive silent intervals, the frequency and position of the perceived pauses were examined. The distribution of the perceived pauses, labeled by the eleven subjects, are to a large extent evenly distributed across the speaking styles (as opposed to the distribution of silent intervals). The average "words per perceived pauses ratio" is highest in acted dialog (4.1) followed by the continuant speech (3.2) as in figure (6).

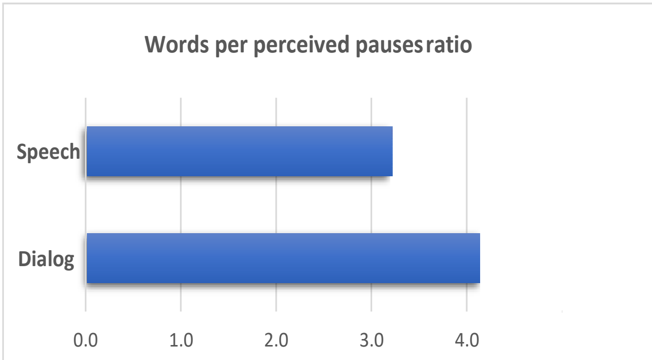


Figure 6: The words/perceived pauses ratio.

Concerning the relation between perceived pauses and phrasal boundaries, recall and precision rates are given in Figure (7) using definitions (1) and (2). The results reported here are based on instances where at least 7 of the 11 subjects perceived a pause.

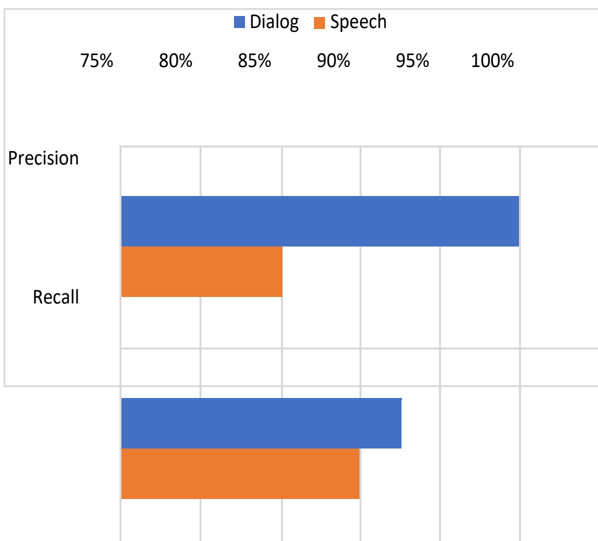


Figure 7: Recall and precision rates for perceived pauses and phrasal boundaries in acted dialogs and continuant speeches.

In the acted dialog style, the majority of the perceived pauses are located at phrasal boundaries (high recall = 93%), while in the continuant speech, we found the majority of perceived pauses located between words positions (low recall = 89%). It is observed that, the perceived pauses are rare at phrasal boundary positions in spontaneous speech, as shown by low precision (85.48%), while more frequently occurring in the acted dialog (100%). Additionally, in the acted dialog style, pauses are perceived at sentence and phrase boundaries. In continuant speech, pauses are located between phrases, e.g. in front of NPs, AdvPs, PPs, conjunctions or verbs, as well as within phrases. Even though the linguistic context of the perceived pauses is similar to the context described for the silent intervals for the various speaking styles, the acoustic and perceived pauses do not necessarily overlap [16].

F. Production and perception of Pauses

To give an overall picture of the correlation between the silent intervals and pauses perceived by each of the 11 subjects participating in the listening test, recall and precision are measured. Here, recall (R) describes the percentage of the acoustic pauses that were actually perceived, while precision (P) gives the percentage of perceived pauses that corresponds to acoustic silence.

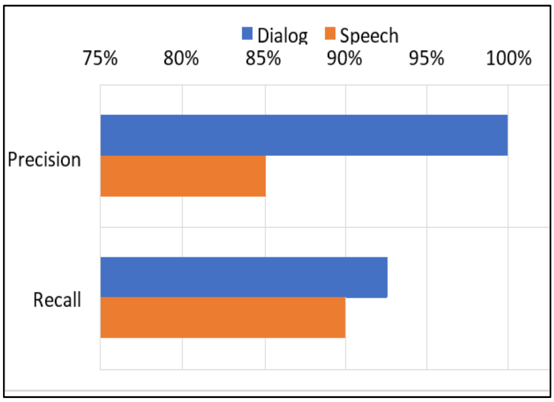


Figure 8: The correlation between acoustic and perceived pauses in the three speaking styles.

The results reported here are based on instances where at least 7 of the 11 subjects perceived a pause. It is clear that the correlation of the acoustic and perceived pauses varies across the speaking styles. In the acted dialog, a considerable number of pauses are perceived by the subjects (100% recall), but many of the perceived pauses do not have any correlates in acoustic silence (93% precision). In the continuant speech, the majority of the acoustic pauses are perceived by the listeners, and many of the perceived pauses actually have an acoustic correlate (represented by higher precision = 96.36%).

4 Conclusion and Future work

In this study, we investigated the phenomena of pausing in two different speaking styles in Arabic: acted dialogue, and continuant speech. Additionally, we examined the syntactic context that corresponded to intervals of acoustic silence and listener perceived pauses. Our results show large differences across the speaking styles. In the acted dialog, all acoustic silence intervals are perceived by the listeners, but there is a number of perceived pauses do not have an acoustic correlate in silence. In the continuant speech, the majority of the acoustic pauses are perceived by the listeners, and many of the perceived pauses actually have an acoustic correlate. Considering the syntactic environment in which the acoustic and perceptual pauses appear, we observed that small silence intervals is perceived: if it occurs in Connection to phrasal boundaries, while if this small silence is found in the middle of phrase, the listeners do not perceive those intervals as pauses. Not surprisingly, we also showed that pause length have an effect on the listeners' perception; the longer the silent intervals are, the better the chance that the perceived pause is actually an acoustic silent interval. Questions we find important to explore in future work concerning syntactic variation in connection to pausing structure. Since the speech sample is not large enough to enable predicting more syntactic boundaries necessary for pausing. Other fields for future work include the investigation of the relation between the duration of silence interval and the type of the phrasal boundaries. We propose in the end of this study, to complete working on enlarging and improving the speech-syntactic database and enhancing it with different kinds of linguistic information such as semantic and pragmatic. Integrated speech- linguistic database will open the way for better language understanding and speech synthesis in the future.

REFERENCESD.

[1] F.Goldman-Eisler, Psycholinguistics: Experiments in Spontaneous Speech. Academic Press, New York, 1968.

[2] D.Aaronson, *Temporal course of perception in an immediate recall task.* J. Exp. PsychoI. 76:129-140, 1968.

[3] F. Goldman-Eisler, *Speech production and the predictability of words in context*. Q. J. Exp. Psychol. 10:96, 1958.

[4] J. P. Gee, and F.Grosjean, *Performance Structures: A Psycholinguistic and Linguistic Appraisal*. Cognitive Psychology, pp. 411–458, 1983.

[5] M. Swerts, and R.Geluykens, *Prosody as a Marker of Information Flow in Spoken Discourse*, Language and Speech, vol. 37, pp. 21–45, 1994.

[6] E.Shriberg, A.Stolcke, D.Hakkani-T¨ur, *Prosody-Based Automatic Segmentation of Speech into Sentences and Topics,* Speech Communication, vol. 32, pp. 127–154. 2000.

[7] J. Hirschberg, *Communication and prosody: Functional aspects of prosody*. Speech Communication, *36*1, 31-43, 2002.

[8] M. E. V. Donzel, *Prosodic aspects of information structure in discourse*. 1999.

[9] J.Bachenko, E. Fitzpatrick, *computational grammar of discourse- neutral prosodic phrasing in English*. Computational linguistics, 155-170 . 1990.

[10] Wang, M. Q., and Hirschberg, J. Automatic classification of intonational phrase boundaries. *Computer Speech & Language*, *6*2), 175-196. 1992.

[11] M.Ostendorf and N.Veilleux, *A hierarchical stochastic model for automatic prediction of prosodic boundary location*. *Computational Linguistics,* 27-54. 1994.

[12] P.Taylor and A. W. Black, *Assigning phrase breaks from part-of-speech sequences*. 1998.

[13] A.Butcher, *Pause and syntactic structure*. In W. Dechert & M. Raupach (Eds.), Temporal variables in speech (pp. 86-90). Mouton. 1980.

[14] B.Zellner, *Pauses and the temporal structure of speech*. In *Zellner,* B. Pauses and the temporal structure of speech, in E. Keller (Ed.) Fundamentals of speech synthesis and speech recognition. *(pp. 41-62). Chichester: John Wiley.* (pp. 41-62). John Wiley. 1994.

[15] L. Olson, *Advanced Data Mining Techniques*, Springer, (1st edition February 1, 2008), page 138,

ISBN 3-540-76916-1. 2008

[16] B.Megyesi and S.Gustafson-Capkova, *Production and perception of pauses and their linguistic context in read and spontaneous speech in Swedish*. In INTERSPEECH. 2002.

[17] E. MacNeal and S.Pinker, *The Language Instinct: How the Mind Creates Language*. 1995.

BIOGRAPHY

Israa Elhosiny
Principal Grammar Developer in the Arabic Computational Linguistics Center Bibliotheca Alexandrina.
PhD. Student in the Department of Phonetics and Linguistics. Faculty of Arts, Alexandria University (2016).
She obtained her MA, from Department of Phonetics and Linguistics 2015.
She obtained her BA, from Department of Phonetics and Linguistics 2004.
She obtained the UNL certificates; CLEA250, CLEA750, CUP250, and CUP500. She attended the X UNL School organized by the UNDL foundation and Bibliotheca Alexandrina (7-11 October 2012).
She has an experience in Python programming, and obtained the New Horizon certificates in Java programming (OOP, JDBC and JSF) (2017).
She is also Principal Grammar Developer of Arabic Computational Linguistics Center in Bibliotheca Alexandrina. She is working in the Universal Networking Language project in Library of Alexandria since 1-6-2005 till now.
She has an experience in Arabic language analysis and generation and text tokenization. She participated in building grammars using UNL for library information system (LIS) and Knowledge Extraction sYStem (Keys).
She has an experience in automatic Arabic diacritization. She participated in building grammars, and developing a diacritization evaluation tool.
She is a member in many scientific organizations: (1) Egyptian Society of Language Engineering, Cairo, (2) Universal Networking Language foundation, United Nations, Geneva, Switzerland.

Mervat Fashal
Professor in Department of Phonetics and Linguistics, Faculty of Arts, Alexandria University, Alexandria, Egypt.
She obtained her BA, from Department of Phonetics and Linguistics in 1979.
She obtained her MA, from Department of Phonetics and Linguistics in 1986.
She obtained her PhD in 1991.
She is a member in many scientific organizations: (1) Egyptian Society of Language Engineering, Cairo, (2) Egyptian Society of Phonetics, Ain Shams University, Cairo.
She has articles in the field of acoustics, prosody, experimental phonetics, speech perception and speech recognition and speaker identification in the field of forensic phonetics.

دراسة الوقفات الصوتية المدركة وعلاقتها بالمركبات النحوية

إسراء الحسيني[1]، ميرفت فشل[2]

قسم الصوتيات واللسانيات، كلية الآداب، جامعة الإسكندرية

[1]israa_elsayed@yahoo.com

[2]Mervat.fashal@alexu.edu.eg

ملخص— إن التوليد الآلي للكلمات بصورة شبه طبيعية يتطلب التحكم في التركيب الزمني لإنتاج الكلام. لذلك فإن العلاقة بين حدود العبارات التطريزية من حيث الوقفات الصوتية والتركيب النحوي وجب دراستها في نوعين مختلفين من الكلام؛ المحادثة والكلام المتصل. تركز هذه الورقة البحثية على دراسة العلاقة بين كل من إنتاج المتكلم للكلام وإدراك المستمع له وعلاقتهما بحدود العبارات التركيبية في الجملة المنطوقة. من أجل وصف الترابط بين الوقفات الصوتية وحدود العبارات في كل من أساليب الكلام موضع الدراسة، اعتمد الباحث على حساب نسبة كل من الدقة والضبط **precision** والاسترجاع **recall.**

الكلمات المفتاحية: انتاج الكلام – فهم الكلام – تصنيف العبارات

Automatic Arabic Speaker Recognition Using Gaussian Mixture Model

, Amna Dheif[*1], Aya Nabil[*2,] Rehab Arafat[*3] Mervat Fashal[*4]

[*]*Phonetics and linguistics Department, Faculty of Arts, Alexandria University*
Alshatby, Alexandria, Egypt

[1]amnaadheif@gmail.com

[2]aya.nabil253@gmail.com

[3]rehab.arafat@gmail.com

[4]mervat.fashal@alexu.edu.eg

Abstract— **This article is presenting an experiment for Arabic speaker identification using the Gaussian Mixture Model (GMM). The speech signals of the speaker were described using the Mel-Frequency Cepstral Coefficients (MFCCs). Three experiments were held on a free open-source dataset, which consists of 20 words spoken by 50 male Arabic speakers. The fifteen speakers were used both in text-dependent and text-independent experiments. The first experiment yielded an accuracy of 84.48%. While the second experiment yielded an accuracy of 83.45%. The third experiment yielded an accuracy of 94.59%.**

Keywords: **speaker recognition (SR), Speaker Identification, Speaker verification, Mel-Frequency Cepstral Coefficients (MFCCs), GMM model, text-dependent recognition, text independent recognition**.

1 INTRODUCTION

Automatic Speech recognition is an interdisciplinary field that involves may methodologies that aim at the automatic recognition and identification of spoken language by means of applications. Automatic Speaker Recognition (ASR) is the process of recognizing the identity of persons based on their voices using an application. The process of speaker recognition can be categorized into **speaker identification** and **speaker verification** (ASV). [1] **Speaker verification** is the process of automatic verifying a person's identity from his voice. In other words, it decides if the speaker is whom he claims to be. This process can be used for verifying the identity of a person in applications like banking by telephone or voice mail. While the purpose of automatic speaker identification is to decide who the person is, without a priori identity speaker. It is the procedure of mapping of the speech signal of an unknown speaker to a database of known speakers.

There are some factors that must be taken in account that may lead to recognition and identifications errors. These factors could be misspoken or misread phrases. The verification process also may be affected be the emotional state of the speakers. Other environmental factors could be as the noise, the microphone placement or using different microphones for enrolment.

On the other hand, other factors could make the verification process a hard task. The variety of speaker's age, gender, mood and sickness may also affect the process and could make the recognition more complicated job and may lead to low rate of recognition. Also, accents will differ between speakers. The variety of the physiology of the organs of vocal tract of speakers will lead to variability in the speech signals.

The first speech recognition systems worked on isolated word or letter recognition and were speaker dependent. The next systems were to work on continuous speech, with a vocabulary of approximately a thousand words. Current state of the art systems is working on conversational or spontaneous speech in noisy and limited bandwidth domains. [2]

2 FEATURE EXTRACTION TECHNIQUES:

In the phase of extracting acoustic features form the speakers, the most commonly feature sets used in speech recognition are Mel frequency cepstral coefficients (MFCCs) and perceptual linear prediction (PLP) [2]. And

the Mel frequency cepstral coefficients (MFCCs) are the most used method to represent the speech spectrum in speaker recognition systems. The MFFCs will be used in this experiment to extract the features in the voice signal. MFCC focuses on series of calculation that uses Cepstrum with a nonlinear frequency axis following Mel scale. To obtain melcepstrum, the voice signal is windowed first using analysis window and then Discrete Fourier Transform is computed. The whole calculations will de discoed in the section 6.2.
Perceptual Linear Prediction (PLP) coefficients are improved spectral representation. It target is to model the psychoacoustics of hearing, by means of implementing three properties of the human auditory system: the nonlinear frequency response of the human ear; the critical bands in the cochlea; and the non-linear amplitude response. [2]

3 MACHINE LEARNING MODELS:

The classical machine learning models applied in the field of speech recognition somehow revolves around Hidden Markov Model (HMM) and Gaussian Mixture Model (GMM). [3] Hidden Markov models are generative models based on stochastic finite state networks. The acoustic model provides the likelihood of a set of acoustic vectors given a word sequence. Markov models are stochastic state machines with a finite set of N states. Given a pointer to the active state at time, the selection of the next state has a constant probability distribution. Thus, the sequence of states is a stationary stochastic process. It assumed that a given state depends on the existence of the previous states.
In the other hand, GMM can be thought of as a single state HMM [3]. GMM is a probabilistic model that supposes that all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. One can think of mixture models as generalizing k-means clustering to combine the information about the variant structure of the data as well as the centers of the latent Gaussians. GMM is preferably applied because it is more reliable than HMM, although HMM possess more accuracy. Also, GMM operates the output much faster than HMM.[3]
The GMM parameters can be thought as being analogous to a set of formant-like features. The component means correspond to the **formant locations**, the standard deviations to the **formant bandwidths** and the component energies to the **amplitudes**. The GMM parameters are extracted, then, they are ordered according to their frequency values. They are ordered as the component with the lowest mean is the first component and so on [2].
Usually, either of them is used independently or with a Deep Neural Network (DNN). However, sometimes they both can be used together in combination [3].

4 ARABIC EXISTING MODELS

(Tolba, 2011) Presented an approach for automatic speaker identification on Arabic speakers depended on Continuous Hidden Markov Models (CHMMs). The Mel-Frequency Cepstral Coefficients (MFCCs) were selected to describe the speech signal. The general Gaussian density distribution HMM is developed for the CHMM system. Ten Arabic speakers (Tolba, 2011)[4] Presented an approach for automatic speaker identification on Arabic speakers depended on Continuous Hidden Markov Models (CHMMs). The Mel-Frequency Cepstral Coefficients (MFCCs) were selected to describe the speech signal. The general Gaussian density distribution HMM is developed for the CHMM system. Ten Arabic speakers were participated in this experiment. The identification rate was 100% for text dependent experiments, but 80% for text-independent experiments.

Djemili, Bedda, & Bourouba (2007) [5] proposed a new hybrid approach that combines different models (statistical models and support vector machines). This technique gets benefit from the advantages of both models. Support vector machines (SVMs) are used in the training phase to divide the whole speakers' space into small subsets of speakers within a hierarchical tree structure. During the testing phase a speech token is assigned to its corresponding group and evaluated using gaussian mixture models (GMMs). They showed that the proposed method can significantly improve the performance of text independent speaker identification task. The improvements are up to 50% reduction in identification error rate compared to the baseline statistical model.

Al-Ani, Mohammed, & Aljebory (2007)[6] presented a hybrid approach for Arabic speaker identification, where the wavelet transform and neural networks are used together to form the. Features are extracted by applying a discrete wavelet transform (DWT), while a neural network (NN) is used for formulating the system database and for handling the task of decision making. Evaluation of the system depends on false acceptance ratio (FAR) and false rejection ratio (FRR) performance. The participated speakers were 25 randomly aged male and female speakers.

The evaluation criteria parameters obtained are; FAR=14.5% and FRR=24.5%.

Naseem & Deriche (2006) [7] discussed a new system using the Dempster Shafer theory of evidence. They depended on a combined classifier based on the Dempster-Shafer theory outperforms the individual LPCC and MFCC classifiers. They developed a new approach for combining the results of the two different to improve the classification results of the LPCC and the MFCC. Results of their work are reported in this following table.

Aldhaheri & Al-Saadi (2004) [8] presented a new technique for text-independent speaker identification for noisy speech is presented. They based on finding the ratio of the singular values of the feature vectors of the unknown speaker and each of the N reference features stored in the database. The i**th** reference feature that gives the largest ratio is considered the feature of the unknown speaker.

They archived an accuracy rate of 99.5% for clean speech and 77.5% for noisy speech.

Table (1): Results of Naseem & Deriche's work (2006)

Method	Training	Testing	Recogntion
MFCC	5	3	85%
LPCC	5	3	83.3%
NNEF	5	3	90%

5 CHALLENGES OF AUTOMATED SPEAKER RECOGNITION

In order to perform the task of speaker recognition, a series of acoustic features are extracted from the speech signal, and then pattern recognition algorithms are trained on these features. Thus, the quality of speech signals is critical for the system performance. The are many factors that may be considered as challenges in that audio dataset. Some factors could be back to speaker characteristics. The speech signal differs among speakers due to the different anatomy and physiology of the vocal tracts of the speakers. Variations in the speech signal also could be back to the gender of the speakers, the ager of the speakers, speaking at different rates or using different accents of the language. A procedure to deal with this variability is through the construction of speaker-dependent speech recognition systems. But this requires a new system for each speaker to be constructed [9]. *However*, *Speaker-independent* systems have some flexibility, in the point that these systems are designed to recognize any speaker. Practically, speaker-dependent systems will make fewer errors than a speaker-independent system.

Other factors that could also lead to variations in the speech signals are environmental factors, in other words, the acoustic environment in which the speech is recorded, along with any transmission channel. This may have a significant impact on the accuracy of a speech signal, and accordingly on the speaker recognition process.so in case of recording outside the laboratories, it is very important to separate different acoustic signals or noises found in the environment [9]. Also, the position the microphone to the speaker and the movements of the speaker's head relative to the microphone may lead to some variability of the signal. So, all these factors must be taken in account during the process of speaker recognition.

6 PROPOSED MODEL:

This section explains the methodology followed in the current study. Two experiments were conducted, the first and second experiment's goal is to build a text-independent speaker recognition system for 50 speaker and the third experiment's goal is to build text-dependent speaker recognition system also for 50 speakers.

The three experiments relied on the same features extracted from the speech corpus and also used the same machine learning model which is the Gaussian Mixture model (GMM) to build the speakers models . In the following sections, the speech corpus used, the features used to represent each speaker identity, and each experiment will be discussed in great detail. Figure (1) shows how the model works in the three experiments.

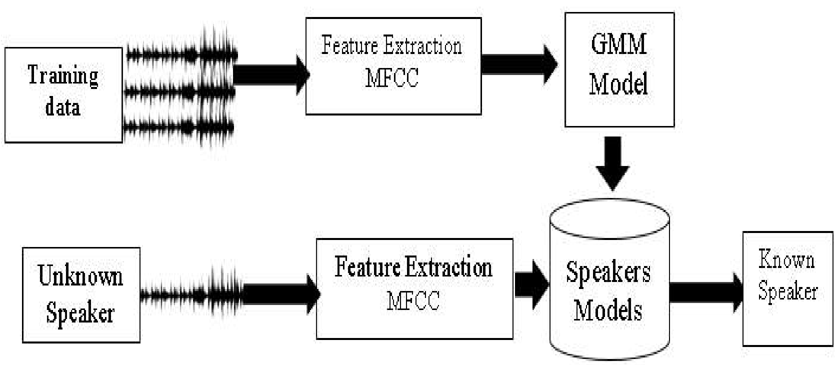


Figure (1) diagram for how the proposed model works in the three experiments

12.1DATASET

This dataset used in this experiment is taken from "Arabic Corpus of Isolated Words". It is a free open-source dataset created by University of Stirling in Central Belt of Scotland. It can be downloaded from the official website of the university60. It consists of 9992 utterances, of 20 words spoken by 50 native male Arabic speakers. The data was recorded with a 44100 Hz sampling rate in .wav format files. Each file contains only one voice. Table (1) shows the twenty words of the Arabic Corpus of Isolated Words along with their transcription.

Table 2: The twenty words of the Arabic Corpus of Isolated Words Along with their transcription

Word	Transcription
صِفْر	se.fer

[60] http://www.cs.stir.ac.uk/~lss/arabic/[11]

وَاحِدْ	wa.ħid
إِثْنَانِ	ʔiθ.naːn
ثَلَاثَة,	θa.laː.θah
أَربَعَة	ʔar.ba.ʕah
خَمْسَة	xam.sah
سِتَّة	sit.ta
سَبْعَة	sab.ʕah
ثَمَانِيَة	θa.maːniy.yah
تِسْعَة	tis.ʕah
التَّنْشِيطْ	ʔat.tan.ʃiːt̠
التَّحْوِيلْ	ʔat.taħ.wiːl
الرَّصِيد	ʔar.ra.s̠iːd
التَّسْدِيدْ	ʔat.tas.diːd
نَعَمْ	na.ʕam
لَا	La:
التَّمْوِيل	ʔat.tam.wiːl
الْبَيَانَاتْ	ʔal.ba.yaː.naːt
الْحِسَابْ	ʔal.ħi.saːb
إِنْهَاءْ	ʔin.haʔ

12.2FEATURE EXTRACTION

The general methodology of speaker recognition involves extracting discriminatory features from the audio data and feeding them to a training model that would use these features to segregate speakers from each other. The so-called *Mel-Frequency Cepstral Coefficients* (MFCC) are the most used in speech and speaker recognition.

The MFCC model imitates human perception processes, and therefore is considered the best model for speech/ and speaker identification. MFCC model have main stages a speech signal has to go through. The first stage is the pre-emphasis stage where high frequencies are amplified, and the noise is cut off from the signal. The Pre emphasis filter used in all the experiments was .97 filter. The second stage is framing where the speech signal is sliced into (overlapping) short frames. The size of frame in the current study was 25ms and the step between each two frames is 10ms. The third stage is windowing to keep the continuity of the first and the last points in the frame. The windowing function used in the current study is hamming window. The fourth stage is applying a fast Fourier transform is applied on each frame to identify the main frequencies present in each frame. Subsequently the filter banks are computed using the Mel scale which tells which frequencies humans discern as the difference between two closely spaced frequencies are not discerned as frequencies increase. Once the filter bank energies are computed, we take the logarithm of them. This imitates how humans perceive loudness of a signal. The last step is to apply the Discrete Cosine Transform (DCT) to the filter banks and eventually a multidimensional feature vector for every speech signal is yielded. In the current study delta coefficients are appended to the MFCC coefficients. The Following figure number (2) depicts the MFCC model steps.

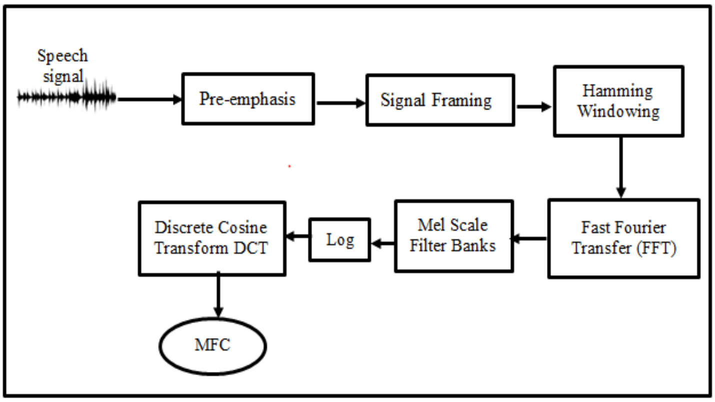


Figure (2): MFCCs Model Steps

12.3EXPERIMENTS:

The experiments proposed in this paper were based on with Gaussian Mixture Models (GMMs) in order to identify Arabic speakers automatically from their voices using the **sklearn.mixture** python package.

12.3.1 EXPERIMENT ONE:

Experiment one addressed the goal of building a text-independent speaker recognition system using the early proposed Gaussian Mixture Model (GMM). The previously discussed features where fed to The GMM where it built a model for each speaker. The first thirteen (13) words of the Arabic Corpus of Isolated Words of each speaker represented the training material for the GMM model. The model was trained on the ten trials of each of the thirteen words for each speaker, thus, the training data consisted of 6,500 speech signals. The other 7 words remaining in the corpus represented the testing material for the GMM model. The model was tested on the ten trials of each of the seven words for each speaker, thus, the testing data consisted of 3,500 speech signals.

12.3.2 EXPERIMENT TWO:

Experiment two was also conducted to build a text-independent speaker recognition system using GMM. The last thirteen (13) words of the Arabic Corpus of Isolated Words of each speaker were alternatively used as the training material in this experiment, while the first 7 words in the speech corpus represented the testing material for the model. The concept behind the second experiment is to find out if the training data is different, how the accuracy of the model shall differ.

12.3.3 EXPERIMENT THREE:

Experiment three alternatively addressed the goal of building a text-dependent speaker recognition system. The first seven trials of all the 20 words of the speech corpus of each speaker represented the training material for the GMM model, thus, the training data consisted of 7,000 speech signals. The other last three trials of the 20 words remaining in the corpus represented the testing material, thus, the testing data consisted of 3,000 speech signals. The results of the text dependent speaker recognition are assumed to be higher than the text independent two experiments.

7 RESULTS

The first experiment yielded an accuracy of 84.48%. Figure (3) depicts the number of times each testing word caused the system to misidentify the speaker.

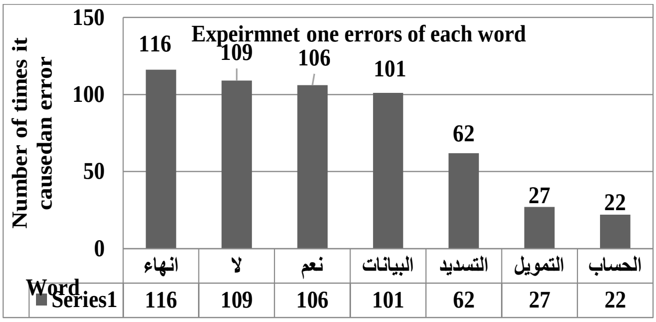


Figure (3) Experiment one errors of each word

The second experiment yielded an accuracy of 83.45%. Figure (4) depicts the number of times each testing word caused the system to misidentify the speaker.

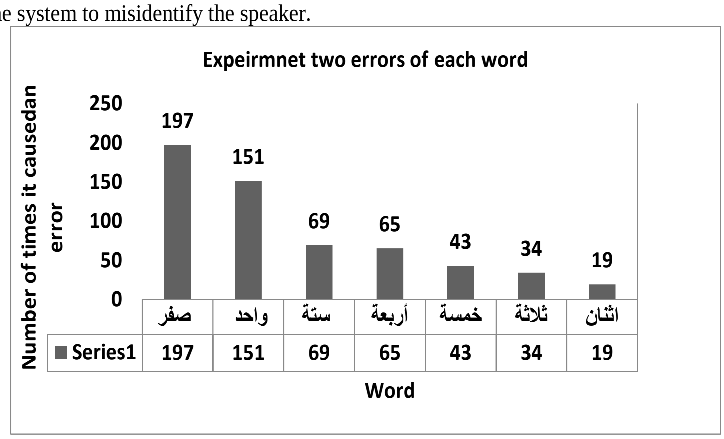


Figure (4) Experiment two errors of each word

The third experiment yielded an accuracy 94.59%. Figure (5) depicts the number of times each testing word caused the system to misidentify the speaker

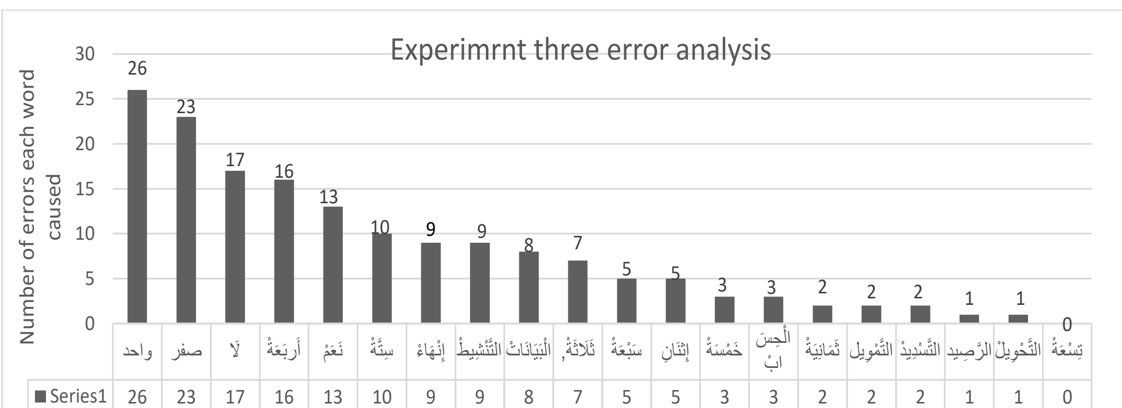


Figure (5) Experiment three errors of each word

8 DISCUSSION AND CONCLUSION

The thee experiments yielded good results regarding the task of speaker identification. Of course, the text independent speaker identification task is more challenging than the text dependent speaker one. This can be seen in the drastic ten percent drop of the accuracy from 94.59% in the text dependent task to 84% which represents the average accuracy of the text independent two tasks.

Changing the training data in the text independent two experiments had a small impact on the accuracy as the accuracy dropped by 1.03% in the second experiment.

Running an error analysis for all the experiments, it was found that, words that were short in length are the reason why the model behaved poorly. Table 2 illustrates the average length of each word in milliseconds and the average length of the sonorant sounds in each word in milliseconds. As shown the short words "**لَا**", "**إِنْهَاءْ**","**نَعَمْ**" and "**الْحِسَابْ**" are the ones that caused the highest errors in experiment one. The word "**التَّسْدِيدْ**" has a very short sonorant sounds length which made the model performance poor. The two words "**التَّمْوِيل**" and "**الْبَيَانَاتْ**" are both long in length and have the longest sonorant sounds yet they caused a high error rate.

In experiment two, the two words "**صِفْرْ**", "**سِتَّةْ**", "**خَمْسَةْ**"and "**وَاحِدْ**" caused the highest errors as they have short sonorant sounds. The least two words the caused errors are the longest two words in the testing set which are "**ثَلَاثَةْ**" and "**إِثْنَانِ**". The word "**أَربَعَةْ**" has a long length and long sonorant sounds yet it caused a high error rate.

In experiment three, the short words are the ones that caused most of the errors except for "**تِسْعَةْ**" and "**خَمْسَةْ**". On the other hand, the long words caused the least errors except for "**الْبَيَانَاتْ**" and "**التَّنْشِيطْ**".

Table 3: the average length of each word and the average length of the sonorant sounds in each word in milliseconds

Word	Average Length in MSEC	Average Sonorant sounds length in MSEC
لَا	27	27
إِنْهَاءْ	28	28
تِسْعَةْ	35	20
خَمْسَةْ	40	20
ثَلَاثَةْ	42	32
صِفْرْ	45	20
أَربَعَةْ	46	46
سَبْعَةْ	46	30
سِتَّةْ	47	22
وَاحِدْ	54	38
إِثْنَانِ	54	40
ثَمَانِيَةْ	55	45
نَعَمْ	65	65
الْحِسَابْ	65	40
التَّمْوِيل	66	65
التَّسْدِيدْ	71	30
التَّحْوِيلْ	72	35
التَّنْشِيطْ	73	35
الرَّصِيد	73	60

البَيَانَاتْ	78	78

To conclude, the three experiments show promising results for speaker identification. The researchers suggest that the error rate could be decreased if both the training data and the testing data are longer in length.

REFERENCES

[1] Joseph P. Campbell, JR. Speaker Recognition: A Tutorial. PROCEEDINGS OF THE IEEE, VOL. 85, NO. 9, SEPTEMBER 1997.
[2] A Gaussian Mixture Model Spectral Representation for Speech Recognition Matthew Nicholas Stuttle Hughes Hall and Cambridge University Engineering Department. 2003.
[3] Speaker- Identification over Call Records. Atul Anand, Bharti Parmar, Ankit Kumar. ITM university, India.
[4] Tolba, H. (2011). A high-performance text-independent speaker identification of Arabic speakers using a CHMM based approach. *Alexandria Engineering Journal*, 43–47.
[5] Djemili, R., Bedda, M. & Bourouba, H. (2007). A Hubrid GMM/SVM System for Text Independnent Speaker Identification. International Journal of Electronic, Computer, Enteritic, Electronic and Communication Engineering, 713-719.
[6] Al-Ani, 1. S.,Mohammed, T.S., & Aljebory, k. M. (2007). Speakers Identification: A Hybrid Approach Using Neural Networks and Wavelet Transform. *Journal of Computer Science*, 304-309.
[7] Naseem, I., & Deriche, M. (2006). A new Algorithm for Speaker Idenification Using the Dempster-Shafer Theory Of Evidence. King Fahd University of Petroleum and Minerals. Saudi Arabia.
[8] Aldhaheri, R. W., & Al-Saadi, a. F. (2004). Text-Independent Speaker Identification in Noisy Environment Using Singular Value Decomposition. King Abdulaziz University. Saudi Arabi.
[9] The Handbook of Computational Linguistics and Natural Language Processing. Alexander Clark, Chris Fox, and

Shalom Lappin

[10] International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-2 Number-4 Issue-7 December-2012 119. Automatic Speaker Recognition System. Parul1, R. B. Dubey2
ECE, Hindu College of Engineering, Sonepat, India1,2
[11]Computing Science and Mathematics, Faculty of Natural Sciences, University of Stirling, Scotland

BIOGRAPHY

Amna Helmy: Assistant teacher at Phonetics and Linguistics department, Faculty of Arts, Alexandria University .She got a Bachelor of Arts in 2013 (Phonetics and Linguistics department – Faculty of Arts - Alexandria university – Alexandria - Egypt).She has graduated from Information technology institute (ITI) in 2014 intake 34.She worked as a demonstrator at Phonetics and Linguistics department , Faculty of Arts ,Alexandria university form 2013 till 2019 .She has discussed her Master thesis in 2019 about semantic role labeling. She had participated in three published papers.

Aya Nabil: Assistant teacher at Phonetics and Linguistics department, Faculty of Arts, Alexandria University She got a Bachelor of Arts in 2012 (Phonetics and Linguistics department – Faculty of Arts - Alexandria university – Alexandria - Egypt).She has graduated from Information technology institute (ITI) in 2013 intake 33.She works as a demonstrator at Phonetics and Linguistics department , Faculty of Arts ,Alexandria university form 2013 till now .She has discussed her Master thesis in 13-10-2018.

Rehab Arafat: Phonetics Laboratory' specialist at Phonetics and Linguistics department, Faculty of Arts, Alexandria University. She got a Bachelor of Arts in 2008 (Phonetics and Linguistics department – Faculty of Arts - Alexandria university – Alexandria - Egypt). She worked in UNL Center in Bibliotheca Alexandrina from 2008-2012. She has discussed her Master thesis in 18-3-2015.

Mervat Fashal: Professor of phonetics science, and in psycho-acoustic field, Department of phonetics and linguistics, Faculty of arts, Alexandria University. She had studied Ph.D. in Germany. She has articles in the filed of acoustics, prosody, experimental phonetics, speech perception and speech recognition and speaker identification in the field of forensic phonetics.

TRANSLATED ABSTRACT

التعرف الآلى على المتحدث باستخدام نموذج جاوس المختلط

آمنة ضيف[1]، آية نبيل[2] ، رحاب عرفات[3] ، مرفت فشل[4]

قسم اللغويات و الصوتيات، كلية الأداب، جامعة الاسكندرية

[1]amnaadheif@gmail.com

[2]aya.nabil253@gmail.com

[3]rehab.arafat@gmail.com

[4]mervat.fashal@alexu.edu.eg

ملخص:

تقدم هذه المقالة تجربة للتعرف الآلي على متحدثي اللغة العربية باستخدام نموذج جاوس الإحصائي . .(GMM) تم توصيف الموجات الصوتية للمتحدث باستخدام معاملات ميل التردد (MFCC). أجريت التجربة على مدونة صوتية مفتوحة المصدر، تتكون من 20 كلمة يتحدث بها خمسون متحدثًا باللغة العربية. تم استخدام المتحدثين الخمسين في كل من التجارب المعتمدة على النصوص والتجارب غير المعتمدة عن النصوص. أسفرت التجربة الأولى عن نسبة دقة في التعرف الآلي بلغت 84.48 ٪. بينما أسفرت التجربة الثانية عن نسبة تعرف بلغت 83.45 ٪. والتجربة الثالثة بلغت نسبة التعرف فيها 94.59%.

الكلمات المفتاحية: التعرف الآلي على المتكلم (SR) ، التحديد الآلي للمتكلم، التحقق الآلي من المتكلم، معاملات ميل التردد، نموذج جاوس الإحصائي، التعرف الآلي المعتمد على النص، التعرف الآلي غير المعتمد على النص.

End-to-End Arabic Speech Recognition: A Review

Abdelaziz A. Abdelhamid[*, **1], Hamzah A. Alsayadi[*2], Islam Hegazy [*3], Zaki T. Fayed[*4]

[*] *Computer Science Department, Faculty of Computer and information Sciences, Ain Shams University, Cairo, Egypt*

[1]Abdelaziz@cis.asu.edu.eg

[2]hamzah.sayadi@cis.asu.edu.eg

[3]islheg@cis.asu.edu.eg

[4]zaki.taha@cis.asu.edu.eg

[**] *Computer Science Department, College of Computing and Information Technology, Shaqra University, Saudi Arabia*

Abstract: — **Automatic speech recognition (ASR) is a crucial field of science due to its massive applications that can be developed to help humans to improve their daily life tasks. Despite its long history, ASR remains an active and interesting research field in general and on Arabic language in particular. Arabic is one of the most widely spoken languages. However, current research is still limited on it due to its high variations and complex morphology. Therefore, this paper highlights the most recent techniques and key milestones of Arabic speech recognition to guide researchers who are interested in working on the Arabic language. There are many machine learning techniques applied in building ASR systems. For long time, hidden Markov models (HMMs)-Gaussian mixed models (GMMs) were standing as the best frameworks for ASR. However, in last decade, hybrid HMM-deep neural network (DNN) models and end-to-end deep learning models have been emerged as a breakthrough in improving the performance of ASR. End-to-end deep learning is distinguished as the most recent methodology in the field and represents the main focus of this review. Therefore, the proposed review discusses the most recent achievements of research on Arabic speech from the end-to-end methodology perspective. In addition, the currently available services and toolkits necessary for building end-to-end models are explained.**

Keywords: **Automatic Speech Recognition (ASR), End-to-End; Deep learning; LSTM; CTC; RNN; Attention-Based; HMM.**

1 Introduction

Arabic language is one of the five most important languages in the world [1]. It is the official language about 25 countries [1]. Arabic is one of the most widely spoken languages around the world with an estimated number of over 313 million speakers and with 270 million as a second language speaker Arabic is ranked as the forth after Mandarin, Spanish and English, it contain about 12 million words [2]. Moreover, it is the language of the Islamic holy book "Quran" with 1.8 billion Muslims around the world in 2015 and projected to increase to 3 billion in 2060 [3]. There have been relatively little speech recognition researches on Arabic compared to other languages [4]. Arabic is a morphologically rich language. Prefixes and suffixes, affixes for short, augment word stems to form words [5].

Arabic language is one of the oldest languages in the world and is a Semitic language and has high variation [6]. There are three types in Arabic language: 1) Classical Arabic is the language of Quran and Hadith (the religion language) and language of the old Arabic poetry. 2) Modern Standard Arabic (MSA) is based on classical Arabic without some features like syntax structure and diacritics. MSA used in formal communication, news, modern books, newspapers, and modern books. 3) Dialectal Arabic has multiple regional forms and is used for daily spoken communication in non-formal settings. With the advent of social media, dialectal Arabic is also written. Each type of Arabic language has different grammatical, lexical, and morphological standard. This makes hardness of developing Arabic NLP applications to process data from different varieties [7].

There are several type of dialectal Arabic, each type differs from one country to another. This problem is even more pronounced for dialectal Arabic due to the following reasons:

1) Additional prefixes, and sometimes suffixes, are informally introduced during the everyday use of language. 2) The amount of text data available for dialectal Arabic is usually much smaller than that for

MSA, and hence it is not clear how to increase the vocabulary size to reduce out-of-vocabulary (OOV). 3) Even if vocabulary is increased using some means, the sparse text resources will lead to poor estimates of the language model probabilities, and hence may hurt performance on a different front [8].

Automatic speech recognition (ASR) is the automatic way to transcribe the speech into text. It is used to make machines understand the human voice. Over the last decades, ASR technologies plays an important role in many areas such as education, personal computers, robotics, mobile phones, dictation, military, health, security systems, …etc. ASR is important because speech is the simplest ways of communication among people. In addition, ASR systems are under active development and adopted by a wide range of applications due to their functionality and simplicity. For instance, it is used in customer care applications where users can interact with a voice-enabled service instead of human interactions. This helps with serving higher number of customers and reduce lengthy service queues. Another example is using ASR in smart homes to control heating, lighting, and other appliances. In this context, Automatic Digit/Command Recognition is considered as one of the most challenging domains in ASR.

The growing importance of Digit/Command recognition through the increasing use of applications that help human-machine interaction by natural languages such as command systems via pronounced digits [9, 10].

There are many systems presented to show the importance of ASR. The most popular systems are: Microsoft SAPI, Dragon Naturally Speaking, and IBM via voice. Open source speech recognition systems are available too, such as domain speech-to-text system and The SPHINX-II [11, 12]. Most researchers developed ASR system based on Hidden Markov Models (HMMs) [7]. HMM is a statistical model where the system being modeled is assumed to be a Markov process with unknown parameters, and the challenge is to determine the hidden parameters, from the observable parameters, based on this assumption. The extracted model parameters can then be used to perform further analysis, for example for pattern recognition applications. Its extension into foreign languages (English is the standard) represent a real research challenge area [13].

Arabic Automatic Speech Recognition (ASR) is a challenging task because of the morphology, data sparseness, and lexical variety of the language [14]. Although there are too many people who speak Arabic, there is little research in Arabic compared to other languages [15, 16, 17].

The first system on Arabic ASR is used to recognize the modern standard Arabic (MSA). The most difficult problems in building highly accurate ASRs system for Arabic are the morphological complexity, predominance of non diacritized text material, and the enormous dialectal variety [16].

D. Vergyri et al. examine the usage of morphology-based language model at different phase in a speech recognition system for conversational Arabic [16]. K. Kirchhoff et al. [17] examine the recognition of dialectal Arabic and study the discrepancies between formal and dialectal Arabic in the speech recognition point of view. D. Vergyri et al [15] examine the automatic diacritizing Arabic text for use in acoustic model training for ASR. Reducing the entry barrier to build robust Automatic Speech Recognition (ASR) for Arabic has been a research concern over the past decade [18]–[21].

Different studies have been investigated in the literature to propose recognition systems using different approaches [8, 22, 23]. However, compared to other languages such as English, the number of research papers in Arabic language is limited [7]. In this review, some studies concerning ASR systems for the Arabic language will be discussed.

Large-vocabulary automatic speech recognition (ASR) for conversational Arabic poses several challenges for the speech research community. Most acoustic training features for Arabic ASR is transcribed in the Arabic nondiacritized form, which does not include short vowels and other diacritics that reflect differences in pronunciation, such as the fattaha, kassra, etc. In particular, the Arabic diacritized text (standard script) form (e.g. broadcast news corpora) is much easier to recognize than the Arabic nondiacritized form.

The nondiacritized texts have constraint for features of recognizer training model. This constraint is cause problems for both language and acoustic modeling. First, if the identity and location of the short vowels in the signal is not known, hence, this makes it difficult to train accurate acoustic models. Second, the lack of diacritics leads to a great set of linguistic contexts for a particular word paradigm; language models trained the nondiacritized features may be less predictive than diacritized trained. Both of these factors may lead to a loss in recognition accuracy. The work in [24] shows the significant of both word error rate and language model perplexity is increased when the Arabic text does not contain vowel information.

There are software applications developed for the automatic diacritization of Arabic by some companies (Sakhr, Apptek, RDI) [25]. However, these products use to predict diacritics using only possible morphological analyses, text-based information, and such as the syntactic context of words. In the context of diacritization for speech recognition, by contrast, acoustic data is available that can be used as an additional knowledge source [25].

Traditional automatic speech recognition (ASR) systems used a modular design. In these systems, different model are trained for pronunciation lexicon, acoustic modeling, and language modeling separately. In contrast, in end-to-end (E2E) approach, all these models are trained to convert the features of acoustic to text transcriptions directly, potentially optimizing all parts for the end task [26]. The goal of End-to-end ASR is to make is to simplify training the above module-based components into a single-network architecture within a deep learning technique, in order to fix these issues. End-to-end ASR approach typically depends only on both acoustic and language data without linguistic knowledge, and train the model with a single algorithm [27].

Unfortunately, they are also less interpretable: identifying what different parts do and what properties they capture is less straightforward. It is a common problem in many neural network models besides E2E ASR. Therefore, a line of work is concerned with deciphering the information captured by learned representations in neural models that are trained on some downstream task [28].

Previous work analyzed different neural representations and various properties, such as evaluating how phonetic information is captured in neural acoustic models [29, 30]. However, E2E ASR models are still relatively under-explored.

Therefore, the end-to-end makes it easy to develop ASR systems without expert knowledge. The end-to-end ASR architecture have several types such as recurrent neural network (RNN) transducer [31], attention-based encoder decoder [32], connectionist temporal classification (CTC) [33], and their hybrid models [34, 35].

Recently, the use of external language models has shown significant improvement of accuracy in neural machine translation [36] and end-to-end ASR [37, 38]. This approach is called shallow fusion, where the decoder network is combined with an external language model in log probability domain for decoding [27].

This paper is organized as follows: In Section 2, we briefly give background of the end-to-end techniques. In Sections 3, we try to describe the related work to Arabic ASR using some techniques such as HMM-based, neural network, recurrent neural network, deep learning and etc. In Section 4 we over view the work in Arabic ASR that used the end-to-end approach. Section 5 mentioned the End-to-End Arabic ASR services like API services and toolkits. Finally Section 6 presents the conclusion and the future directions.

2 BACKGROUND

In this section we present detailed information about end-to-end approach and the techniques that are used in end-to-end approaches.

AA. End-to-end speech recognition approach

In recent years, with the advancement of deep learning, end-to-end solutions have emerged in many areas. A salient example is the wide application of deep convolutional neural networks (CNNs) to the classification task [39]. End-to-end represents a mapping between the sequence of input acoustic features and sequence of grapheme or words. In Conventional ASR, the language model, pronunciation, and trained acoustic components are trained separately as shown Fig. 6 [39].

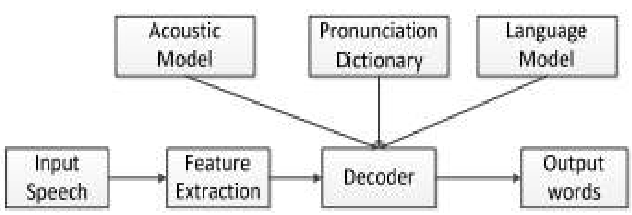


Figure 1: Conventional ASR Structure [39]

While end-to-end speech recognition greatly simplifies the complexity of traditional speech recognition, the models are not need to train separately, the pronunciation information or language can automatically learn as shown Fig. 2 [39].

Figure 2: End-to-end ASR Structure [39]

For more details, most end-to-end speech recognition models involve the following phases: 1) encoder, which realizes a mapping between the sequence of speech input and the sequence of feature; 2) aligner, which align the feature sequence to language; 3) decoder, which decodes the final identification result as in Fig. 3 [40].

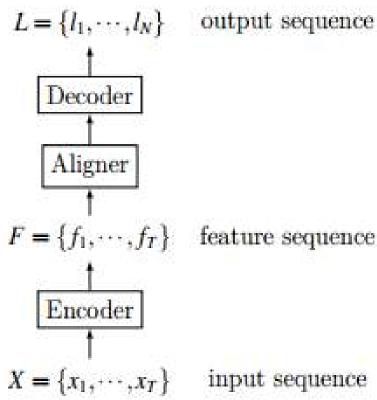


Figure 3: Function structure of end-to-end model [39]

End-to-end speech recognition is mainly based on deep recurrent neural network. A deep recurrent neural network alone is not enough for a speech recognition system. The first attempts used CTC, but it is incapable of learning the language and needs language model to clean up common mistakes. Instead of CTC, attention-based models were tested and they proved to be outperforming previous models, due to the ability of learning all components of a speech recognizer. Both methods still have some benefits over one another and recent works have started using a hybrid CTC/attention based architecture.

Using attention-based encoder–decoder models, the end-to-end idea can be naturally ported to speech recognition [38]. In principle, building an ASR system involves learning a mapping from speech feature vectors to a transcript (e.g., words, phones, characters, etc.), both of which that are in sequence, so no reordering is expected to take place. If we can learn such a mapping directly, all the components are optimized under a unified objective, which can enhance the final recognition performance obliterating the need for separate acoustic and language models [41].

BB. Deep Learning

Building intelligent machines has fascinated humanity for centuries [42]. The field of Artificial Intelligence (AI), however, started to develop relatively recently, when programmable digital computers were conceived. The rise of deep learning [43] has recently contributed to renew the interest in AI and has allowed current technology to achieve higher levels of artificial intelligence.

Deep learning is actually a very general machine-learning paradigm that follows a compositionality principle to represent the world around us efficiently. Current deep learning implementation exploits deep neural

networks, which are properly trained to progressively discover complex representations starting from simpler ones. This principle can be applied in several practical problems, including the problem of recognizing human speech [43].
The deep learning paradigm is currently implemented with Deep Neural Networks (DNNs), that are Artificial Neural Networks (ANNs) based on several hidden layers between input and output. Each layer learns higher-level features that are later processed by the following layer [42].

CC. Recurrent neural network (RNN)

The recurrent neural network was first developed in the 1980s [44]. RNN is a type of deep learning model that works best for handling sequential information. RNN assumes that all inputs, one or more hidden layers, and outputs are dependent on each other, unlike the traditional neural network. It keeps a memory of previous outputs and passes those as inputs from one-step of the network to the next as shown in Fig. 4. This way the network can have a deeper understanding of the application [45].

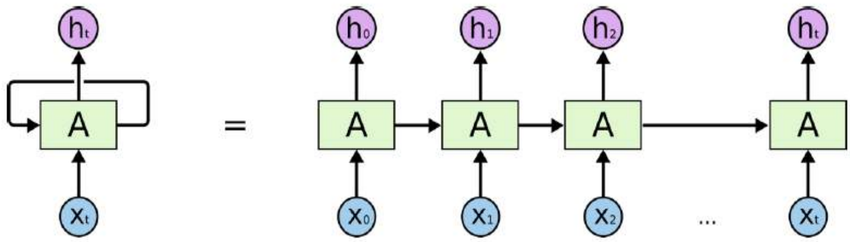


Figure 4: An unrolled RNN [44]

There are many applications of Recurrent Neural Networks such as Machine Translation, Robot control, Time series prediction, Speech recognition, Speech synthesis, Music composition, Handwriting recognition, Human action recognition, and etc. [44].
The above figure shows a chunk of neural network (A), that takes (Xt) and previous output as inputs and outputs a value (ht). The recurrence allows the network to pass information from one-step to the next [43]. This is the basic workflow of a RNN, but it is often used with bidirectional to get more accurate results.
In the training of RNN stage, the backpropagation algorithm used to calculate gradients and adjust weights in artificial neural network (ANN). It also applied in adjusting and modifying the weights after the updating of the feedback process [44]. The RNN uses the backpropagation through time (BPTT) method, this method utilizes working-backward manner to swap the weights of each unit according to the total output error [44].

DD. Long Short-Term Memory (LSTM)

Standard RNN architectures have a problem with multiple hidden layers. When passing information from one hidden layer to another, the information might get lost, if there are many layers. LSTMs are a special case of RNN, can able to use the long-term dependencies and save information long time as a default [46]. The LSTM model is organized as chain structure form. However, the repeating module has a different structure. The standard RNN has a single neural network, while LSTM uses four interacting layers with a unique communication link as shown in Fig. 5 [46].

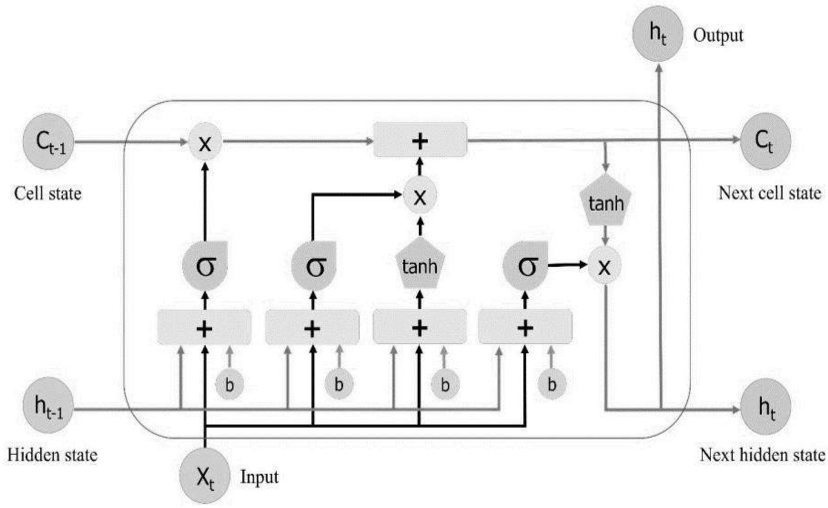


Figure 5: The structure of the Long Short-Term Memory (LSTM) neural network [44]

LSTM has many Applications like Robot control, Time series prediction, Speech recognition, Handwriting recognition, Human action recognition, Sign Language Translation, Protein Homology, Detection, and etc. LSTM handles this kind of situation and enables RNN to preserve memory throughout the whole learning process RNN by adding additional interactions per module (or cell) [44].

LSTM architecture consists of memory blocks, which are all recurrently connected to each other. Each memory block contains at least one self-connected memory cell [45]. Any next state receives two states, the cell state, which allows the data to flow forward without change these data and the hidden state. LSTMs include sigmoid gates, which used to add or remove data from the cell state, a gate contains weights like the layer or series of matrix operations [44]. It also uses gates to control the memorizing process as the following: first identify and remove the information is not required from the cell, then storing information from the new input in the cell state as well as to update the cell state. Finally, the output values is based on the output cell state [44].

EE. Connectionist temporal classification (CTC)

People talk with very different rates of speed which makes training an ASR system a lot more difficult. That is why the alignment between characters in the transcript and audio is always unknown [40]. One way of solving this problem is to manually align all characters to their location in the audio. The major downside is that it's very time consuming when dealing with large datasets. Another option is to use connectionist temporal classification (CTC) which has become a very popular among RNNs [47].

Graves et al [48] proposed CTC to allow for training an acoustic model without the need for frame-level alignments between the acoustics and the transcripts. The acoustic model training using CTC as the loss function is an end-to-end training, which does not need to align the data in advance, but only needs an input sequence and an output sequence to be trained. Structure as shown in Fig. 6 [39].

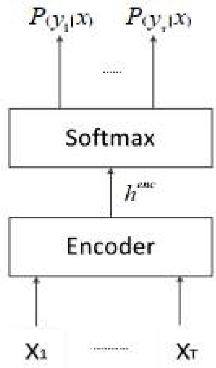


Figure 6: CTC structure [39]

CTC has applications like speech, handwriting recognition, and Scene Text Recognition. CTC is a type of neural network output and associated scoring function. It is used with RNNs to handle sequential problems. CTC sums over the probability of all possible alignments between the input and the output [49]. In this way, we do not need to align the data one by one, the CTC represents the spike in the whole speech as blank (which has no predicted value in any frame) [39]. Assuming that an input has a length greater than the actual word's length, one option for solving the problem is to collapse all repeating characters as shown in Fig. 7.

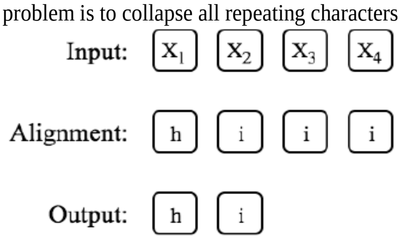


Figure 7: Merging repeated characters

FF. Attention model

Attention Identifies encoded frames that are relevant to producing current output. The first use of attention-based Encoder-Decoder Models in the context of neural machine translation. The purpose of this model was to solve the problem of RNN-based sequence to sequence (Seq2Seq) model. Seq2Seq is represent end-to-end machine translation that need to encode input texts into a fixed-length vector and decode the target result [39]. The attention-based Encoder-Decoder is used to encode the input data into a sequence of vectors, the decoder based the attention mechanism makes each vector in this sequence has different weights (different length). Then the output is the sum of sequence's weights and the output of previous sequences [39]. Because the signal in speech have a different length, the attention-based Encoder-Decoder is suitable for speech recognition tasks for the following reasons: 1) speech recognition, like translation task, need a sequence-to-sequence approach to recognize the output sequence from the input sequence, 2) the encoder–decoder mechanism based attention method can implicitly find the soft alignment between input and output sequences, which solves a signal length problem for speech recognition, 3) Encoding result is not depend on a single fixed-length vector, the model can still have a good effect on long input sequence, so it is also possible for model to deal with different length in speech input [40].

Attention-based end-to-end model includes three parts: encoder, aligner, and decoder. In particular, its aligner part uses attention mechanism as shown in Fig. 8.

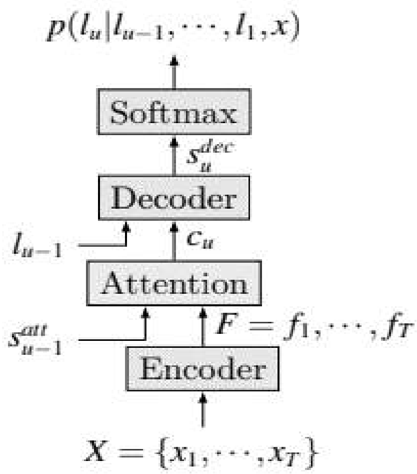


Figure 8: Attention model structure [39, 40]

3 RELATED WORK

The aim of speech recognition is to enable machines to accept sounds and act based on it. Automatic speech recognition is the ability for a machine to recognize "receive and interpret" the speech and convert it into readable form or text and performing an action based on the instructions defined by the human [50].

We will focus on Arabic ASR using HMM, neural network-based, deep learning, recurrent neural network, and hybrid approaches. We will discuss the literature in terms of Arabic ASR classification, stages, and techniques.

Yu and Deng [43] presented an architecture for Arabic ASR, consists of four stages: 1) preprocessing stage. 2) Feature extraction stage. 3) Pronunciation dictionary, Language model, and Acoustic model (decoding). 4) Post-processing results were the best hypothesis is produced. The mechanism of work as following: First, speech signals of utterance used as input in the preprocessing stage. Then, the output is processed speech signals used as input to feature extraction stage where the features are represent by vectors as output. After that, the vectors use as input in the next stage, the decoding stage. The decoding stage is working along with a pronunciation dictionary. Finally, the n-best hypothesis, the output of the pronunciation dictionary stage, is used as input to post-processing. As a result, the best hypothesis is produced from this work operation.

Turab, Khatatneh, and Odeh in [51] proposed system for the phoneme recognition as ASR system. Many techniques are used in that paper: firstly Gaussian Low Pass filtering algorithm along with the neural network in the pre-processing stage to have an enhancement the result. Then, catching a signal, sampling, quantization and setting energy is done in the phoneme recognition sage. After that, a neural network is used to improve the results is achieved. When applying the Gaussian Low Pass filter in voice signals, the enhanced impact in results hence, due to noise reduction. After that, in the training phase, the neural network has been used to train the system in order to recognize the speech signal.

Ahmed and Ghabayen in [4] suggested three approached to enhance the AASR. The first approach is the punctuation modeling, where they used a decision tree with variant pronunciation generation. Then, a hybrid approach is proposed to acclimate the native acoustic model with another native acoustic model. Finally, processed text was used to enhance and improve the language model. In the results, the word error rate (WER) is reduced by 1%, 1.2% and 1.9% for the pronunciation model, the acoustic modeling, and the language model respectively.

Emami and Mangu [52], scrutinized usage of the neural network to recognize the Arabic speech by a distributed word representation. Consequently, the neural network makes robust generalization able to fight the data sparseness problem in the best way. They were investigated many factors such as: n-gram order parameter experiment, output vocabulary, the method of normalization, model size and parameters. The evaluation used the Arabic news broadcast, and conversation broadcast. In the result, some enhancement has been achieved using optimized neural network model over 4-gram up to 3.8% relative WER.

Kirchhof, Bilmes and Stolcke in [53] proposed system for conversational Arabic speech recognition. They utilized a language model for morphology. This proposed system enhances the results for two different test sets, this result is 1.8% and 1.5% respectively.

The authors in [54] developed system for an Arabic broadcast news transcription system. The speaker-independent large vocabulary is used in this system. The developed system uses five-state HMM for triphone acoustic models, with 8 and 16 Gaussian mixture distributions. Arabic broadcast news corpus is used in experiments and performs 10.14 as WER.

Hyassat and Abu Zitar in [55] presented system for the holy Quran corpus using language model and lexicon and WER improved by 46.182%.

In [56] Elmahdy and Mohamed developed a dialectal Arabic speech recognition system using a new multilingual approach. This multilingual approach includes several acoustic models using HMM. The news broadcast speech corpus of modern standard Arabic speech and Egyptian colloquial Arabic are used for training and testing. The accuracy reached 99.34%.

Selouani, Sid Ahmed and Malika Boudraa in [57] used MSA continues speech corpus using Algerian Arabic Speech Database (ALGASD) and HHM. As a result the accuracy is achieved 91.65%.

The authors in [58] compared between two Arabic ASR using different techniques. These techniques are the traditional multi-layer perceptron and general regression neural network (GRNN) algorithm. The proposed system involves two components: first, pre-processing component which includes segmental normalization and feature extraction process, second, a classification component which uses neural networks based on nonparametric density estimation. From the results it is clear that the GRNN gives better results and faster performance than the feedforward backpropagation in the recognition rate using Arabic digits corpus.

The authors in [59] investigated the use of a recurrent neural network for Arabic digits related speech recognition. They developed system based on a multi-speaker mode and a speaker-independent mode with accuracy 99.5% and 94.5% correct digit respectively.

In [2] the authors proposed a novel approach for Arabic isolated speech recognition system. The system is implemented using modular recurrent Elman neural networks (MRENN). They conclude that the result almost reaches to the same result by the last traditional HMM-based speech recognition approaches.

The authors in [60] aimed to develop an Arabic ASR through the differences in the 29 letters of the Arabic alphabet. They proposed a system based on a fully-connected recurrent neural network with a backpropagation through time learning algorithm. The investigation has been done twice: first, to prove hidden layer makes the learning of complex classification tasks more efficient, then the comparison of the LPCCC and MFCC. The results overall the LPCCC outperform the MFCC performance by 0.7%.

The authors in [61] introduced three different system structures based on biologically inspired methods for Arabic ASR. The system consists of two parts: features extraction using Mel Frequency Cepstral Coefficients algorithm (MFCC), and adapt the dataset normalization to use for train and test the three different systems. The performances of these systems were 47.52%, 44.58% and 46.63% frame recognition for single MLP identification system, category-based phonemes recognition system and individual Phoneme classifier system respectively.

In [62], authors present the enhancement of the Arabic ASR performance in mobiles communication system. The proposed approach involves two modules: 1) Front-End represent the features extraction stage using MFCC-MT (Multitaper Frequency Cepstral Coefficients features) and Gabor features GF-MFCC. 2) Back-end is represent the recognizer stage, in this stage the authors were investigated different ASR techniques such as CHMM (Continues Hidden Markov Models), DNN (Deep Neural Network) and HMMDNN hybrid. They focused on HMMDNN and claimed that it can get consistently almost 8% of clean speech, 13% of AMR-NB coder and 8.5% of DSR coders.

El-Desoky et al. [63] proposed a novel approach for Egyptian Arabic ASR. They mix the best features of the morpheme-based LMs and feature-rich modeling with the DNN-LMs. On the other hand, make the mixture of words and morphemes along with their features. The result shows that the WER is reduced according to compared to the traditional word-based LMs.

An AASR system was introduced by AlHanai. Et al. [64] with a 1,200-h speech corpus. The developed system used Deep neural network (DNN), DNN structure involving many techniques: Feed-forward, Convolutional, TimeDelay, Recurrent Long Short-Term Memory (LSTM), Highway LSTM (H-LSTM) and Grid LSTM (GLSTM). The evaluation using corpus shows the best result with 18.3% WER using trained GLSTM models.

In [65] the authors presented description of comparison for some of the state-of-the-art speech recognition techniques details. The corpus that contain 50-h of transcription audio from a news channel "Al-jazeerah" is used to train all different approaches. The hybrid DNN/HMM approach with the MPE (Minimum Phone Error) criterion are trained by sequential DNN achieves the best result. The results were obtained 17.86%, 29.85%, and 25.6% WER for broadcast news reports, a broadcast conversations, and overall respectively.

The Arabic news speech recognition system is developed in [66], the authors of this system presented the KALDI recipe to build it. The broadcast news system using 200 hours GALE data used to trained model. The building of text normalization, vowelization, and language model details are described. The results using the building prototype broadcast news system are 15.81% WER on Broadcast Report (BR) and 32.21% WER on Broadcast Conversation (BC) with a combined WER of 26.95%.

This paper is description of ASR system in in the framework of the 2016 Multi-Genre Broadcast (MGB-2) Challenge in the Arabic language [67]. The main idea of this work was to take the GMM derived features for training a DNN and combined with the use of time-delay neural networks for acoustic models, the combined approach used to automatically phonetize Arabic words. The final system was a combination of a five systems where the result obtained succeeded the best single LIUM ASR system with a 9% of WER.

Ettaouil et al. [68] introduced the Arabic digit system using a hybrid model ANN/HMM. The core of this work is to determine the optimal codebook generated using Self Organizing Maps (SOM) and the optimal class using optimal neural network. From the results, the dictionary size affects the classification. The codebook vectors are with size 34, 36, and 48 they had a recognition rate 84%, 85%, and 86% respectively.

Wahyuni in [69] recognize spoken Arabic letters challenge issue. The proposed methods is built using Mel-Frequency Cepstral Coefficients (MFCC) based on feature extraction and ANN to distinguish between the pronounce of three different letters (sa, sya, and tsa). The accuracy for this work obtains with an average accuracy is 92.42%.

AbdAlmisreb et al. [70] presented the DNN with three hidden layers, 500 Maxout units with 2 neurons for the unit and used Mel-Frequency Cepstral Coefficients (MFCC) for feature extraction. This approach was trained and tested over a corpus which consisted of 20 Malay speakers of consonant Arabic phonemes recorded. The training set consisted of 5 waveforms and the tested set contained 15 waveforms. The result show that the Maxout based deep structure gave better performance with lowest error rate than other deep networks such as Restricted Boltzmann Machine (RBM), Deep Belief Network (DBN), Convolutional Neural Network (CNN), the conventional feedforward neural network (NN) and Convolutional Auto-Encoder (CAE) which had error rate between 2800 and 3000 (numbers).

4 END-TO-END ARABIC ASR

End-to-end approach has different techniques which can be applied on AASR in a simple way. In this paper we focus on the End-to-end methods for Arabic ASR.

Zerari, Naima, et al [71] present general end-to-end approach to recognize the Arabic spoken digit spelling of an isolated Arabic word. This approach used Mel Frequency Cepstral Coefficients (MFCC) for extraction the relevant features from the natural speech signal, these features are presented by a deep neural network able to deal with the non-uniformity of the sequences length of the speech utterances. It firstly used recurrent Long Short-Term Memory (LSTM) with three gates to decode sequences of these features as fixed vector, then a multilayer perceptron network will receive this vector to perform the classification.

Spoken Arabic Digit dataset is used for training and testing, this dataset consists of 8800 tokens by 88 individual (44 males and 44 females) Arabic native speakers. The performance of this approach is achieved 98.77% of global f-measure.

Authors in [72] have implemented and evaluated character-based modeling in a state-of-the-art speech recognition systems for Arabic using end-to-end ASR. This system used Kaldi toolkit to present an effective character-based ASR and evaluate the models based on words, characters, and statistical morphs. In this system, the word, morph, and character *n*-gram models are trained using a projection layer, an LSTM layer, and a highway layer. Recurrent Neural Network Language Models (RNNLMs) also is used for model training. An acoustic models utilize the MGB-2 corpus for training, which consists of 1,200 hours of broadcast data from multiple genres and even dialects.

Ahmed, Abdelrahman, et al. [73] presents the first end-to-end recipe for an Arabic speech-to-text transcription system based on BDRNNs. This system includes a character based decoder which is used for search to avoid using a word lexicon. The Connectionist Temporal Classification (CTC) is also used which, objective function is used to maximize the output character sequences given the acoustic features as an input. It consists of: 1-a BDRNNs acoustic model, 2- a language model and 3- a character based decoder. In addition, the training and decoding process are based on Arabic grapheme. The objective function used to train BDRNNs is CTC that removes the need for pre-segmented acoustic observations. The evaluation for the test set will be performed on both the word and character levels in order to validate the results with other word based models. The recipe was evaluated using 1200 hours corpus of the Aljazeera multi-Genre broadcast programs. On the development set, the WER is 12.03% for non-overlapped speech. The morph-based models yield the best recognition results for both well-resourced and lower-resourced tasks, but the character-based models are close to their performance in the lower-resource tasks, outperforming the word-based models. Character-based models are especially good at predicting novel word forms that were not seen in the training data. In the results, the word model outperforms the character-based model for the full dataset, while for the under-resourced scenario the character-based model improves over the word model by 6%.

In [74], the authors are developed framework for multilingual speech recognition on low-resource languages. This work used sequence-to-sequence attention-based models and a single transformer for recognition. Sub-words are utilized as the multilingual modeling unit without need for any pronunciation lexicon. The ASR transformer is chosen to be the basic architecture of sequence-to-sequence attention-based model. They employed the language-specific softmax layer instead of softmax layer to solve the problem of few training data on low-resource languages. . The CALLHOME datasets with 6 languages: Mandarin (MA), English (EN), Japanese (JA), Arabic (AR), German (GE) and Spanish (SP) are used in experiment. The result on this Arabic dataset was 13.5% average WER.

In [75] an end-to-end speech recognition model analyzed in terms of Arabic phonetic and graphemic representations as well as different articulatory features. In this work traditional ASR systems and separate components for phoneme or grapheme modeling were employed to run forced-alignment and get annotated data to be used by the classifier.

The lexicon also was used the phonetic lexicon for the phoneme system and the word list with 1:1 word-to-character mapping for grapheme system. Convolutional layers and 5 recurrent long short-term memory (LSTM) are used for encoding and trained with CTC. The system consists of two steps: 1) training the E2E model in the normal fashion, on pairs of utterances and transcriptions, 2) running the trained model on a dataset with frame-level annotations. The steps are used to analyze the representation quality in the end-to-end ASR model. MGB-2 corpus which comprises 1,200 hours of broadcast videos from the Aljazeera Arabic TV channel used for training and testing. Experiments on corpus yield 12:94% and 10:60% WER for Phonemes and Graphemes respectively.

Zerari, Naima, et al. [1] proposed to recognize a set of isolated Arabic utterances issued from two ASR applications, namely: a) TV spoken command recognition, b) spoken digit recognition. The proposed system involves, first, extracting pertinent features from the natural speech signal using Mel Frequency Cepstral Coefficients MFCC (static and dynamic features), the Filter Banks (FB) coefficients, and, next, the extracted features are padded in order to deal with the non-uniformity of the sequences length. Then, a deep architecture represented by a recurrent LSTM or GRU (Gated Recurrent Unit) architectures are used to encode the sequences of MFCC/FB features as a fixed size vector that will be introduced to a Multi-Layer Perceptron network (MLP) to perform the classification (recognition).

Based on recurrent neural networks to process sequences of variable lengths of (1) MFCCs, (2) FBs and (3) delta-delta features of the different spoken digits/commands was presented. The extracted features using the different techniques are, first, encoded as a fixed size vector by a recurrent LSTM/GRU neural network, next, a standard Multi-Layer Perceptron is used to classify the spoken digits/commands with the obtained vector as input. Two datasets are used in this work: the first dataset is the spoken Arabic digit A number of 88 (44 males and 44 females) Arabic native speakers were asked to utter all digits ten times. Accordingly, the database consists of 8800 tokens for spoken digit recognition, the second dataset, speaker-independent mode is considered, where one hundred of Arabic native speakers were participated (50 males comprising 37 adults and 13 kids whereas 50 females including 31 adults and 19 kids) for TV spoken command recognition. The accuracies achieved are **98.77% and 96% for** spoken digit recognition and TV spoken command recognition respectively.

5 END-TO-END ARABIC ASR SERVICES

A. API Services

Google developed Cloud Speech-to-Text service application [35] used to convert Arabic speech or audio file to text using a deep-learning neural network algorithm. Cloud Speech-to-Text service allows for its translator system to directly accept the spoken word to be converted to text then translated. The service offers an API for developers with multiple recognition features. Microsoft Speech API [36] is developed by Microsoft. It used deep neural networks to build speech recognition systems. IBM cloud provide Watson service API for speech to text recognition [37] support modern standard Arabic language until now there is not any work use this API with Arabic.

B. Toolkits

Kaldi [76]. It is a toolkit for speech recognition using deep neural network and support Arabic language. Ali et al. [66] shows the usage of Kaldi to build Arabic broadcast news speech recognition system. They use all Kaldi conventional models.

Miao et al. [77] present the Eesen toolkit for end-to-end ASR, this tool is open source and used sequence-to-sequence learning. The deep recurrent neural networks (RNNs) are used as the acoustic models, and the Long Short-Term Memory (LSTM) units the RNN building blocks. The weighted finite-state transducers (WFSTs) are used for decoding. The CTC labels, lexicons and language models components are encoded into WFSTs.

ESPnet [78] is an open source toolkit is provide a neural end-to-end platform for ASR and other speech processing. It emploes Chainer, PyTorch and dynamic neural network as a main deep learning engine. For data processing, feature extraction/format, and recipes this tool used Kaldi ASR toolkit style.

ESPRESSO [79] is a novel neural end-to-end system for ASR ESPRESSO is built upon the popular NMT framework FAIRSEQ2 , and the flexible deep learning framework PyTorch. By extending FAIRSEQ, ESPRESSO inherits its excellent extensibility: new modules can easily be plugged into the system by extending standard PyTorch interfaces. Additionally, we gain ability to perform distributed training over large data sets for ASR. They also present the first fully parallelized decoder for end-to-end ASR, with look-ahead word-based language model fusion, tightly integrated with the existing sets of optimized inference algorithms (e.g. beam search) inherited from FAIRSEQ and tailored to the scenario of speech recognition.

6 CONCLUSIONS

This paper introduced a review on Arabic speech recognition using end-to-end technology. It mentioned the importance of Arabic speech recognition and the limitation of Arabic language. The literature covered the newest papers in Arabic speech recognition based serval techniques and end-to-end approach include Modern Standard Arabic (MSA) and Dialectal Arabic.

Furthermore, the end-to-end model is an important research direction of speech recognition. It uses the deep learning technique and include two parts: attention model and CTC to solve the data alignment size of signal vector issues. Attention method also used to encoding and decoding.

REFERENCES

[1] Zerari, Naima, et al. "Bidirectional deep architecture for Arabic speech recognition." *Open Computer Science* 9.1 (2019): 92-102.

[2] El Choubassi, M.M., El Khoury, H.E., Alagha, C.E.J., Skaf, J.A., Al-Alaoui, M.A.: Arabic speech recognition using recurrent neural networks. In: Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology (IEEE Cat. No. 03EX795), Darmstadt, Germany, pp. 543–547 (2004)

[3] Lipka, M., Hackett, C.: Why Muslims are the world's fastest-growing religious group. Pew Research Center (2017). http://www.pewresearch.org/fact-tank/2017/04/06/why-muslimsare-the-worlds-fastest-growing-religious-group/. Accessed 14 Nov 2018

[4] Ahmed, Basem HA, and Ayman S. Ghabayen. "Arabic automatic speech recognition enhancement." *2017 Palestinian International Conference on Information and Communication Technology (PICICT)*. IEEE, 2017.

[5] Afify, Mohamed, et al. "On the use of morphological analysis for dialectal Arabic speech recognition." *Ninth International Conference on Spoken Language Processing*. 2006.

[6] Alsayadi, Hamzah A., and Abeer M. ElKorany. "Integrating semantic features for enhancing arabic named entity recognition." *Int. J. Adv. Comput. Sci. Appl.(IJACSA)* 7.3 (2016): 2016.

[7] Algihab, Wajdan, et al. "Arabic Speech Recognition with Deep Learning: A Review." *International Conference on Human-Computer Interaction*. Springer, Cham, 2019.

[8] Lippmann, Richard P. "Review of neural networks for speech recognition." *Neural computation* 1.1 (1989): 1-38.

[9] Rabiner L. R., Juang B. H., Fundamentals of speech recognition, PTR Prentice Hall Englewood Cliffs, 1993

[10] Jelinek, Frederick. *Statistical methods for speech recognition*. MIT press, 1997.

[11] Ordowski, Mark, et al. "A public domain speech-to-text system." *Sixth European Conference on Speech Communication and Technology*. 1999.

[12] Huang X., Alleva F., Hon W., Hwang M., and Rosenfeld R., "The SPHINX-II Speech Recognition System: An Overview," Computer Journal of Computer Speech and Language, vol. 7, no. 2, pp. 137-148, 1993.

[13] Satori, Hassan, et al. "Investigation Arabic Speech Recognition Using CMU Sphinx System." *International Arab Journal of Information Technology (IAJIT)* 6.2 (2009).

[14] Ali, Ahmed, Hamdy Mubarak, and Stephan Vogel. "Advances in dialectal arabic speech recognition: A study using twitter to improve egyptian asr." *International Workshop on Spoken Language Translation (IWSLT 2014)*. 2014.

[15] Huang X., Acero A., and Hon H., Spoken Language Processing: A Guide to Theory, Algorithm and System Design, Prentice Hall, 2001.

[16] Hiyassat H., Nedhal Y., and Asem S., "Automatic Speech Recognition System Requirement Using Z Notation," in Proceedings of of AMSE' 05, France, pp. 514-523, 2005.

[17] Huang D., Automatic Speech Recognition: The Development of the SPHINX System, Kluwer Academic Publishers, 1989.

[18] F. Diehl, M. J. F. Gales, M. Tomalin, and P. C. Woodland, "Morphological decomposition in Arabic ASR systems," *Comput. Speech Lang.*, vol. 26, no. 4, pp. 229–243, Aug. 2012.

[19] D. Rybach, S. Hahn, C. Gollan, R. Schluter, and H. Ney, "Advances in Arabic broadcast news transcription at RWTH," in *2007 IEEE Workshop on Automatic Speech Recognition & Understanding (ASRU)*, 2007, pp. 449–454.

[20] L. Mangu, H.-K. Kuo, S. Chu, B. Kingsbury, G. Saon, H. Soltau, and F. Biadsy, "The IBM 2011 GALE Arabic speech transcription system," in *2011 IEEE Workshop on Automatic Speech Recognition & Understanding*, 2011, pp. 272–277.

[21] B. Kingsbury, H. Soltau, G. Saon, S. Chu, H.-K. Kuo, L. Mangu, S. Ravuri, N. Morgan, and A. Janin, "The IBM 2009 GALE Arabic speech transcription system," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 4672–4675.

[22] Juang, Biing-Hwang, and Lawrence R. Rabiner. "Automatic speech recognition–a brief history of the technology development." *Georgia Institute of Technology. Atlanta Rutgers University and the University of California. Santa Barbara* 1 (2005): 67.

[23] Anusuya M. A., Katti S. K., Speech recognition by machine, a review, arXiv preprint arXiv:1001.2267, 2010.

[24] K. Kirchhoff, J. Bilmes, J. Henderson, R. Schwartz, M. Noamany, P. Schone, G. Ji, S. Das, M. Egan, F. He, D. Vergyri, D. Liu, and N. Duta. 2002. Novel approaches to Arabic speech recognition - final report from the JHU summer workshop 2002. Technical report, Johns Hopkins University.

[25] Vergyri, Dimitra, and Katrin Kirchhoff. "Automatic diacritization of Arabic for acoustic modeling in speech recognition." *Proceedings of the workshop on computational approaches to Arabic script-based languages*. Association for Computational Linguistics, 2004.

[26] Belinkov, Yonatan, Ahmed Ali, and James Glass. "Analyzing Phonetic and Graphemic Representations in End-to-End Automatic Speech Recognition." *arXiv preprint arXiv:1907.04224* (2019).

[27] Hori, Takaaki, Jaejin Cho, and Shinji Watanabe. "End-to-end speech recognition with word-based RNN language models." *2018 IEEE Spoken Language Technology Workshop (SLT)*. IEEE, 2018.

[28] Y. Belinkov and J. Glass, "Analysis methods in neural language processing: A survey," *Transactions of the Association for Computational Linguistics (TACL)*, 2019.

[29] T. Nagamine, M. L. Seltzer, and N. Mesgarani, "Exploring how deep neural networks form phonemic categories," in *Interspeech*, 2015.

[30] Nagamine, Tasha, Michael L. Seltzer, and Nima Mesgarani. "On the Role of Nonlinear Transformations in Deep Neural Network Acoustic Models." *Interspeech*. 2016.

[31] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton, "Speech recognition with deep recurrent neural networks," in *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on*. IEEE, 2013, pp. 6645–6649.

[32] Jan K Chorowski, Dzmitry Bahdanau, Dmitriy Serdyuk, Kyunghyun Cho, and Yoshua Bengio, "Attention-based models for speech recognition," in *Advances in Neural Information Processing Systems (NIPS)*, 2015, pp. 577–585.

[33] Alex Graves and Navdeep Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in *International Conference on Machine Learning (ICML)*, 2014, pp. 1764– 1772.

[34] Suyoun Kim, Takaaki Hori, and Shinji Watanabe, "Joint CTC attention based end-to-end speech recognition using multitask learning," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 4835–4839.

[35] Takaaki Hori, Shinji Watanabe, and John R. Hershey, "Joint CTC/attention decoding for end-to-end speech recognition," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL): Human Language Technologies: long papers*, 2017.

[36] Caglar Gulcehre, Orhan Firat, Kelvin Xu, Kyunghyun Cho, Loic Barrault, Huei-Chi Lin, Fethi Bougares, Holger Schwenk, and Yoshua Bengio, "On using monolingual corpora in neural machine translation," *arXiv preprint arXiv:1503.03535*, 2015.

[37] Takaaki Hori, Shinji Watanabe, Yu Zhang, and William Chan, "Advances in joint CTC-attention based end-to-end speech recognition with a deep CNN encoder and RNN-LM," in *INTERSPEECH*, 2017.

[38] Anjuli Kannan, Yonghui Wu, Patrick Nguyen, Tara N Sainath, Zhifeng Chen, and Rohit Prabhavalkar, "An analysis of incorporating an external language model into a sequence-to-sequence model," *arXiv preprint arXiv:1712.01996*, 2017.

[39] Wang, Song, and Guanyu Li. "Overview of end-to-end speech recognition." *Journal of Physics: Conference Series*. Vol. 1187. No. 5. IOP Publishing, 2019.

[40] Wang, Dong, Xiaodong Wang, and Shaohe Lv. "An Overview of End-to-End Automatic Speech Recognition." *Symmetry* 11.8 (2019): 1018.

[41] Miao, Y., & Metze, F. (2017). End-to-End Architectures for Speech Recognition. In *New Era for Robust Speech Recognition* (pp. 299-323). Springer, Cham.

[42] N. Bostrom. Superintelligence: Paths, Dangers, Strategies. Oxford University Press, 1st edition, 2014.

[43] Yu, D., Deng, L.: Automatic Speech Recognition: A Deep Learning Approach, pp. 13–21. Springer, London (2015). https://doi.org/10.1007/978-1-4471-5779-3.

[44] Le, Xuan-Hien, et al. "Application of long short-term memory (LSTM) neural network for flood forecasting." *Water* 11.7 (2019): 1387.

[45] D. Britz, "WILDML," 17 September 2015. [Online]. Available: http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1- introduction-to-rnns/. [Accessed 21 March 2018].

[46] Olah, C. Understanding LSTM Networks. Available online: http://colah.github.io/posts/2015-08-Understanding-LSTMs/ (accessed on 20 June 2020).

[47] A. Hannun, "Sequence Modeling with CTC," *Distill*, 2017.

[48] Graves et al. "Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks."(ICML 2006).

[49] A. Graves, Supervised Sequence Labelling with Recurrent Neural Networks, Berlin: Springer-Verlag Berlin Heidelberg, 2012.

[50] Nasereddin, H.H.O., Omari, A.A.R.: Classification techniques for automatic speech recognition (ASR) algorithms used with real time speech translation. In: 2017 Computing Conference, London, pp. 200–207 (2017)

[51] Turab, N., Khatatneh, K., Odeh, A.: A novel Arabic Speech Recognition method using neural networks and Gaussian Filtering. (IJEECS) Int. J. Electr. Electron. Comput. Syst. 19 (01) (2014)

[52] Emami, A., Mangu, L.: Empirical study of neural network language models for Arabic speech recognition. In: 2007 IEEE Workshop on Automatic Speech Recognition & Understanding (ASRU), The Westin Miyako Kyoto, pp. 147–152 (2007)

[53] Kirchhoff, K., Vergyri, D., Bilmes, J., Duh, K., Stolcke, A.: Morphology-based language modeling for conversational Arabic speech recognition. Comput. Speech Lang. 20(4), 589–608 (2006)

[54] Alghamdi, M., Elshafei, M., Al-Muhtaseb, H.: Arabic broadcast news transcription system. Int. J. Speech Technol. 10(4), 183–195 (2007)

[55] Hyassat, H., Abu Zitar, R.: Arabic speech recognition using SPHINX engine. Int. J. Speech Technol. 9(3–4), 133–150 (2006)

[56] Elmahdy, M., et al.: Modern standard Arabic based multilingual approach for dialectal Arabic speech recognition. In: Eighth International Symposium on Natural Language Processing, SNLP 2009. IEEE (2009)

[57] Selouani, S.A., Boudraa, M.: Algerian Arabic speech database (ALGASD): corpus design and automatic speech recognition application. Arab. J. Sci. Eng. 35(2C), 15 (2010)

[58] Amrouche, A., Rouvaen, J.M.: Arabic isolated word recognition using general regression neural network. In: 2003 46th Midwest Symposium on Circuits and Systems, Cairo, Egypt, vol. 2, pp. 689–692 (2003)

[59] Alotaibi, Y.A.: Spoken Arabic digits recognizer using recurrent neural networks. In: Proceedings of the Fourth IEEE International Symposium on Signal Processing and Information Technology, Rome, Italy, pp. 195–199 (2004)

[60] Ahmad, A.M., Ismail, S., Samaon, D.F.: Recurrent neural network with backpropagation through time for speech recognition. In: IEEE International Symposium on Communications and Information Technology, ISCIT 2004, Sapporo, Japan, vol. 1, pp. 98–102 (2004)

[61] Hmad, N., Allen, T.: Biologically inspired continuous Arabic speech recognition. In: Bramer, M., Petridis, M. (eds.) SGAI 2012, pp. 245–258. Springer, London (2012). https://doi.org/10.1007/978-1-4471-4739-8_20

[62] Bouchakour, L., Debyeche, M.: Improving continuous Arabic speech recognition over mobile networks DSR and NSR using MFCCs features transformed, 12, 8 (2018)

[63] El-Desoky Mousa, A., Kuo, H.-K.J., Mangu, L., Soltau, H.: Morpheme-based feature-rich language models using deep neural networks for LVCSR of Egyptian Arabic. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, pp. 8435–8439 (2013)

[64] AlHanai, T., Hsu, W.-N., Glass, J.: Development of the MIT ASR system for the 2016 Arabic multi-genre broadcast challenge. In: 2016 IEEE Spoken Language Technology Workshop (SLT), San Diego, CA, pp. 299–304 (2016)

[65] Cardinal, P., et al.: Recent advances in ASR applied to an Arabic transcription system for AlJazeera, p. 5

[66] Ali, Ahmed, et al. "A complete KALDI recipe for building Arabic speech recognition systems." *2014 IEEE spoken language technology workshop (SLT)*. IEEE, 2014.

[67] Tomashenko, N., Vythelingum, K., Rousseau, A., Esteve, Y.: LIUM ASR systems for the 2016 multi-genre broadcast Arabic challenge. In: 2016 IEEE Spoken Language Technology Workshop (SLT), San Diego, CA, pp. 285–291 (2016)

[68] Ettaouil, M., Lazaar, M., En-Naimani, Z.: A hybrid ANN/HMM models for arabic speech recognition using optimal codebook. In: 2013 8th International Conference on Intelligent Systems: Theories and Applications (SITA), Rabat, Morocco, pp. 1–5 (2013)

[69] Wahyuni, E.S.: Arabic speech recognition using MFCC feature extraction and ANN classification. In: 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, pp. 22–25 (2017)

[70] AbdAlmisreb, A., Abidin, A.F., Tahir, N.: Maxout based deep neural networks for Arabic phonemes recognition, p. 6 (2015)

[71] Zerari, Naima, et al. "Bi-directional recurrent end-to-end neural network classifier for spoken Arab digit recognition." *2018 2nd International Conference on Natural Language and Speech Processing (ICNLSP)*. IEEE, 2018.

[72] Smit, Peter, et al. "Character-based units for unlimited vocabulary continuous speech recognition." *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. IEEE, 2017.

[73] Ahmed, Abdelrahman, et al. "End-to-End Lexicon Free Arabic Speech Recognition Using Recurrent Neural Networks." *Computational Linguistics, Speech And Image Processing For Arabic Language* 4 (2018): 231.

[74] Zhou, Shiyu, Shuang Xu, and Bo Xu. "Multilingual end-to-end speech recognition with a single transformer on low-resource languages." *arXiv preprint arXiv:1806.05059* (2018).

[75] Belinkov, Yonatan, Ahmed Ali, and James Glass. "Analyzing Phonetic and Graphemic Representations in End-to-End Automatic Speech Recognition." *arXiv preprint arXiv:1907.04224* (2019).

[76] KALDI. http://kaldi-asr.org/. Accessed 18 Feb 2019

[77] Miao, Yajie, Mohammad Gowayyed, and Florian Metze. "EESEN: End-to-end speech recognition using deep RNN models and WFST-based decoding." *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*. IEEE, 2015.

[78] Watanabe, Shinji, et al. "Espnet: End-to-end speech processing toolkit." *arXiv preprint arXiv:1804.00015* (2018).

[79] Wang, Yiming, et al. "Espresso: A Fast End-to-end Neural Speech Recognition Toolkit." *arXiv preprint arXiv:1909.08723* (2019).

BIOGRAPHY

Dr. Abdelaziz A. Abdelhamid is assistant professor in Faculty of Computer and Information Sciences. He got his MSc degree from Faculty of computer and information sciences in 2009, and got his PhD degree from Faculty of Engineering, The university of Auckland in 2014. He has more than 25 publications in international conferences and journals. His research interests are speech recognition, speech synthesis, image processing, and deep learning.

Hamzah A. Alsayadi is PhD student in Faculty of Computer and Information Sciences. He got his BSc degree from Faculty of sciences in 2007, and held an MSc degree from Faculty of Computer and Information Technology, Cairo university in 2016. His research interests are speech recognition, speech synthesis, Neutral language processing, and deep learning.

Dr. Islam Hegazy is an Assistant Professor in the Faculty of Computer and Information Sciences, Department of Computer Science. Dr. Islam has more than 18 years experience in research and teaching in many fields in the fields of Computer Science. He holds a PhD degree from the University of Calgary, AB, Canada. He did his M.Sc. and B.Sc. degrees in the Faculty of Computer and Information Sciences, Ain Shams University. His research interests focus on network, security, and cloud computing. Dr. Islam acquired several managerial skills as the Director of the Scientific Computing Center, Ain Shams University, and the coordinator of the Software Engineering Credit Hours program, Faculty of Computer and Information Sciences.

Prof. Zaki Taha is a Professor in the Faculty of Computer and Information Sciences, Department of Computer Science. Prof. Zaki has more than 30 years experience in research and teaching in many fields in the fields of Computer Science. He holds a PhD from Communication Department, Facility of Engineering, Ain Shams University 1997. He did his M.Sc. in Computer Science (Speech Processing), Communication and computer department, Faculty of Engineering, Ain Shams University, 1992 and B.Sc. degrees in in Electronic Engineering (Communication Engineering), Military Technical College, 1976. His research interests focus on Speech Processing and Computer Security. Prof. zaki held several positions in Military IS Center, Ministry of Defense, Military Academy, Armed Forces and Faculty of Computer and Information Sciences.

TRANSLATED ABSTRACT

أستعراض الطريقة الشاملة للتعرف على الكلام باللغة العربية

عبدالعزيز عبدالمنعم *،** ، حمزة احمد الصيادي * ، إسلام حجازي * ، زكي طه فايد

*قسم علوم الحاسوب – كلية الحاسبات و المعلومات – جامعة عين شمس – مصر

[1]Abdelaziz@cis.asu.edu.eg

[2]hamzah.sayadi@cis.asu.edu.eg

[3]islheg@cis.asu.edu.eg

[4]zaki.taha@cis.asu.edu.eg

**قسم علوم الحاسوب – كلية الحاسبات و تقنية المعلومات – جامعة شقراء – المملكة العربية السعودية

ملخص:

يعد التعرف التلقائي على الكلام مجالًا مهمًا بسبب تطبيقاته الكثيرة التي يمكن تطويرها لمساعدة البشر على تحسين مهام حياتهم اليومية. على الرغم من تاريخ التعرف على الكلام الطويل ، إلا أنه لا يزال مجالًا بحثيًا مهما وممتعًا بشكل عام وفي اللغة العربية بشكل خاص. اللغة العربية هي واحدة من أكثر اللغات انتشارًا. ومع ذلك ، لا تزال الاعمال البحثية الموجودة حاليا محدودة بسبب الاختلافات الكثيرة في تراكيب الجمل و الحركات الاعرابية المعقدة. لذلك ، هذا العمل يسلط الضوء على أحدث التقنيات المستخدمة في التعرف على الكلام باللغة العربية لمساعدة الباحثين المهتمين بالبحث في التعرف على الكلام باللغة العربية. هناك العديد من تقنيات التعلم الآلي المطبقة في بناء أنظمة التعرف على الكلام. أستمرت تقنيات (HMMs) والنماذج المختلطة الجاوسية (GMMs) لفترة طويلة أفضل طرق للتعامل مع انظمة التعرف على الكلام. ومع ذلك ، في العقد الماضي ، برزت نماذج الشبكة العصبية الهجينة (DNN) المختلطة مع (HMM) وطرق التعلم العميق و الطريقة الشاملة (end-to-end) كتقنبات جديدة لتحسين أداء هذه الانظمة. تتميز الطريقة الشاملة (end-to-end) بأنها أحدث منهجية في هذا المجال ويمثل المحور الرئيسي لهذا العمل. لذلك ، ينافش هذا العمل أحدث إنجازات البحث في مجال التعرف على الكلام باللغة العربية من منظور المنهجية الشاملة (end-to-end). بالإضافة إلى ذلك ، يتم شرح بعض الخدمات وعدد من المكتبات البرمجية المتوفرة حاليًا اللازمة لبناء نماذج شاملة (end-to-end) للتعرف على الكلام.

الكلمات المفتاحية: التعرف الآلى على الكلام، التعلم العميق، **LSTM; CTC; RNN; HMM** .

19ESOLEC

The Egyptian Society Of Language Engineering Conference

The Nineteenth Conference of Language Engineering

Proceedings of the Conference

26-29 September, 2020
Alexandria, Egypt

Sponsors

Preface

The proceedings of the nineteenth Conference on Language Engineering contain nineteen papers reported to the conference. Two papers are written in Arabic, two papers are written in French and the rest are written in English language. The papers in the present document are classified in ten sessions corresponding to ten scopes of those mentioned in the call for papers.

The proceedings contain three invited papers. The first one is presented by Prof. Mohsen Rashwan, Professor in the Department of Electronics and Electrical Communications, faculty of Engineering, Cairo University, Egypt. He is Managing Director of RDI Corporation. The paper is about New Trends in Developing the Human Language Technologies and it will be presented in session one. The second invited paper is presented by Prof. Wafaa Kamel, Professor in Department of Arabic Language, Faculty of Arts, Cairo University, Egypt. The invited paper is about Lexicon and automatic processing systems and will be presented in session two. The third Invited paper is presented by Prof. Mervat Fashal, Professor in Department of Phonetics and Linguistics, Faculty of Arts, Alexandria University, Alexandria, Egypt. The third invited paper is about Human perception vs machine interface: an overview. It will be presented in session six.

The third session contains two papers about Computational Linguistics (I). The fourth session contains three paper dealing with Artificial Intelligence and NLP, while the fifth session contains four paper deals with Corpus based NLP. The invited paper will be in session six. Session seven contains two papers about Speech and Speaker Recognition.

Session eight includes two papers in the area of Speech perception. Session nine includes two papers about computational linguistics (II), while the four papers of session ten are concerned with speech analysis.

Conference Chairman

Prof. Dr. Mohamed Adeeb Riad Ghonaimy

Scope of the conference

Natural Language Processing has gained a lot of importance nowadays with many applications requiring real-time performance. In order to achieve the real-time requirements, the components of a Natural Language Processing (NLP) system should be made more efficient.

NLP overlaps to a large degree with computational linguistics (CL), especially when both are applied to standard Arabic or spoken varieties.

The Egyptian Society of Language Engineering (ESOLE) is a leading institution that interested in language engineering and computational linguistics especially for Arabic. Over the past 18 years, the ESOLE, in its conferences, has brought together researchers from across the field of natural language processing and computational linguistics and provided a wide-scope forum for discussing natural language processing researches as well as the best practices in its applications.

The 19th Annual Conference of Language Engineering (ESOLEC'19) will be held in Bibliotheca Alexandrina, Alexandria from 26 September to 29 September 2020. It continues this tradition and thus welcomes papers on all topics related to both natural language processing and computational linguistics, with the expectation that papers may include linguistic insight.

The relevant topics for the conference include, but are not limited to, the following topics:

Syntax, Semantics, Grammar, and the Lexicon.
Lexical Semantics and Ontology.
Phonology/Morphology, Word Segmentation, Tagging.
Text Mining, Paraphrasing and Summarization.
Speech Processing, Recognition and Synthesis.
Computational Linguistics.
Natural Language Processing for Information Retrieval.
Word Sense Disambiguation.
Automatic Character Recognition.
Semantic Role Labeling.
Sentiment Analysis and Opinion Mining.
Corpus-Based Modeling of Language.
Machine Translation and Translation Aids.
Multilingual Processing.
Statistical and Machine Learning Methods.
Social Networks and Contents Development Challenges.

Computational Forensic Phonetics and Linguistics.

Two workshops will be held in conjunction with ESOLEC'19 to encourage the exchange of ideas and to discuss challenging research issues in NLP. They will be held pre the main conference on Saturday September 26, 2020 and post the main conference on Tuesday, September 29, 2020.

Organization

Conference Chairman:

Prof. M. A. R. Ghonaimy
Conference Co-chair:
Prof. Sameh Alansary
Prof. Salwa Elramly
Prof. Magdy Nagy
Prof. Ayman Bahaa

Conference Organizing Committee

- Prof .M. Nagi
- Prof .A. Fahmy
- Prof. H. Barhamtoshy
- Prof. Y. Elhamalawy
- Prof. S. Alansary
- Prof. H. Korashy
- Prof. P. Elkafrawy
- M .Z. Abdelmageed

Local Organizing Committee

Maha Mousa
Marwa Saber
Amany Fashwan
Nahla Mohamed
Mayar Saeed
Israa Elhosiny

Technical Program Committee:

Prof. Taghrid Anber, **Egypt**
Prof. I. Abdel Ghaffar, **Egypt**
Prof. W. Fakhre, **Egypt**
Prof. M. Z. Abdel Mageed, **Egypt**
Prof. Khalid Choukri, ELDA, **France**
Prof. Nadia Hegazy, **Egypt**
Prof. Mona T. Diab, **USA**
Prof. Neamat Abdelkader, **Egypt**
Prof. Afaf AbdelFattah, **Egypt**
Prof. Mostafa Aref**, Egypt**

Prof. Allain Ramsey, **UK**
Prof. M. Elhamalawy**, Egypt**
Prof. S. Abdou**, Egypt**
Prof. A. A. Fahmy**, Egypt**
Prof. I. Farag, **Egypt**
Prof. Wafa Kamel, **Egypt**
Prof. S. Krauwer, **Netherlands**
Prof. Bente Maegaard, CST, **Denmark**
Prof. S. Elramly**, Egypt**
Prof. M. Nagy**, Egypt**
Prof. A. Rafae, **Egypt**
Prof. K. Refaat, **Egypt**
Prof. M. Rashwan**, Egypt**
Prof. K. Shalaan, **UAE**
Prof. H.M. AL-Barhamtoshy**, Saudi Arabia**
Prof. M. F. Tolba, **Egypt**
Prof. S. Elkareh, **Egypt**.

Invited Speakers

Prof. Mohsen A. A. Rashwan
Prof. Wafaa Kamel Fayed
Prof. Mervat Fashal

The Nineteenth Conference on Language Engineering
Main Conference program

Sunday, 27/9/2020			
8:30	**-**	**10:00**	Registration
10:00	**-**	**11:00**	Opening Session
11:00	**-**	**12:00**	**Session 1:** Chairman: Prof. Aly Aly Fahmi **Invited Paper1: New Trends in Developing the Human Language Technologies** *Prof. Mohsen Rashwan* *Electronics& Communications Engineering Department, Faculty of Engineering, Cairo University.*
12:00	**-**	**12:30**	**Coffee Break**
12:30	**-**	**13:15**	**Session 2**: Chairman: Prof. Salwa Al-Ramly **Invited Paper 2: Lexicon and automatic processing systems.** *Prof. Wafaa Kamel Fayed* *Department of Arabic Language, Faculty of Arts, Cairo University, Cairo, Egypt.*
13:15	**-**	**14:15**	**Session 3: Computational Linguistics (I)** Chairman: Prof. Wafaa Kamel **1. Bibalex Arabic Linguistic Resources and Tools for Language Engineering** [*†]SamehAlansary, [**†]Magdy Nagi [†]*Bibliotheca Alexandrina, Alexandria, Egypt* [*]*Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt* [**]*Computer and System Engineering Department, Faculty of Engineering, Alexandria University, Alexandria, Egypt* **2. Shallow Parsing for Automatic Arabic Text Summarization** Sameh Alansary *Bibliotheca Alexandrina, Alexandria, Egypt* *Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt*
14:15	**-**	**15:15**	**Lunch Break**

15:15	**-**	**16:45**	**Session 4**:**Artificial Intelligence and NLP** Chairman: Prof. Mohsen Rashwan **1. Chatbot System Architecture** Moataz Mohammed[*] and Mostafa M. Aref[**] [*]*Computer Science Department, Faculty of Computer and Information Sciences, Ain-shams University, Cairo, Egypt.* **2. Arabic Optical Character Recognition using Sequence to Sequence Models** Mohamed Sobhi[*], Yasser Hifny[**], Saleh Mesbah[*] [*]*Arab Academy for Science, Technology and Maritime Transport, Alexandria, Egypt.* [**]*University of Helwan, Cairo, Egypt.* **3. The Methodology and Uses of Whale Swarm Algorithm** Amr M Sauber[*], Passent M. El-Kafrawy[*,**], Amr. F. Shawish[*] [*]*Faculty of Science, Menoufia University, Egypt* [**]*School of Information Technology and Computer Science, Nile University, Egypt*
16:45	**-**	**18:45**	**Session 5**: **Corpus based NLP** Chairman: Prof M. Younis Elhamalawy **1. A Critical Review of Language Resources and Tools for Arabic Sentiment Analysis.** Miramar Etman*, SamehAlansary* **Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt.* **2. A Pilot Study of Biber's Model for Language Variation Detection: A Language Engineering Approach** MaramElsaadany[*], SamehAlansary[**] [*]*Pharous University, Alexandria University, Alexandria, Egypt* [**]*Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt.* **3. Semantic role labeling system for modern standard arabic: a rule based approach** AmenaDeif,,SamehAlansary *Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt.* **٤. دور القواعد اللغوية في التمييزالآلي بين معاني الحروف** **خالد مصطفى أبو شبانة*، د. أحمد عبد الغني**** *قسم اللغة العربية، كلية الآداب، جامعة الإسكندرية، مصر **قسم الصوتيات، كلية الآداب، جامعة الإسكندرية، مصر
18:45	-	**22:00**	Banquet

Monday, 28/9/2020

10:00	**-**	**10:45**	**Session 6: Artificial Intelligence** Chairman: Prof. Sameh Alansary **Invited Paper 3: Human perception vs machine interface: an overview** Prof. Mervat Fashal *Professor in Department of Phonetics and Linguistics, Faculty of Arts, Alexandria University, Alexandria, Egypt*
10:45	**-**	**11:45**	**Session 7: Speech and Speaker Recognition** Chairman: Prof Waleed Fakhre **1. Automatic Arabic Speaker Recognition Using Gaussian Mixture Model** Mervat Fashal*, Amna Dheif*, Aya Nabil*, Rehab Arafat* **Phonetics and linguistics Department, Faculty of Arts, Alexandria University Alshatby, Alexandria, Egypt* **2. End-to-End Arabic Speech Recognition: A Review** Abdelaziz A. Abdelhamid[*], Hamzah A. Alsayadi[**], IslamHegazy[*], Zaki T. Fayed[*] [*]*Computer Science Department, Faculty of Computer and information Sciences, Ain Shams University, Cairo, Egypt* [**]*Computer Science Department, College of Computing and Information Technology, Shaqra University, Saudi Arabia*
11:45	**-**	**12:15**	**Coffee break**
12:15	**-**	**13:15**	**Session 8: Speech perception** Chairman: Prof Mervat Fashal **1. The Perception of Arabic Vowel Length by Native and Non-native Listeners: An Experimental Investigation** Eman Kassem* and Lamyaa Tawfik* **Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt* **2. An Investigation of the Correlation between Perceived Pauses and Syntactic Structures** Israa Elhosiny*, Mervat Fashal* **Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt*
13:15	**-**	**14:15**	**Session 9: Computational Linguistics (II)** Chairman: Prof. Seham El-Qareh **1. L'Application du Formalisme des FonctionsLexicales sur la Langue Arabe.** Racha Mohammad Salem *Département de Langue et de LittératureFrançaises, Faculté des Lettres, Universitéd'Alexandrie, Alexandrie, Egypt.*

			2. QuellesContraintes pour Traduire la Morphologie et la Syntaxe? AsmaaGaafar Abdel-Rassoul *Faculte de Lettres, Universite de Menoufia*
14:15	**-**	**15:15**	**Lunch Break**
15:15	**-**	**17:15**	**Session 10: Speech Analysis** Chairman: Amr Gody **١. قياس انفعالات الممثل الصوتية باستخدام تقنيات هندسة اللغة** د. صديقة لاشين قسم الدراسات المسرحية-كلية الآداب-جامعة الإسكندرية **2. The Acoustic Characteristics of Read and Spontaneous Colloquial Arabic Speech Corpora: A Pilot Study.** Rudyna Ahmed *Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt* **3. Syllables Classification of ASR using Hybrid Visual Features in Fixed HMM** Doaa A. Lehabik *, Amr M. Gody *, Mohamed H. Merzban*, Sameh F. Saad ** **Electrical Engineering Department, Fayoum University, Fayoum, EGYPT* ***Modern Sciences and Arts University, 6 October City, Giza, Egypt* **4. Creating and Implementing ArSL Corpus for Deaf Drivers** Samah A. Abbas[*], Hassanin M. Al-Barhamtoshy[**], Fahad M. Al-Otaibi[**] [*]*Management Information Systems Department, Faculty of Economics and Administration, King Abdulaziz University* *Jeddah, Saudi Arabia* [**]*Information Technology Department, Faculty of Computing & Information Technology, King Abdulaziz University* *Jeddah, Saudi Arabia*
17:15	**-**	**18:00**	**Closing session**

Computational Linguistics

Bibalex Arabic Linguistic Resources and Tools for Language Engineering

*†1Sameh Alansary, *‡ 2Magdy Nagi

1sameh.alansary@bibalex.org

2Magdy.nagi@bibalex.org

Bibliotheca Alexandrina, Alexandria, Egypt

†*Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt*

‡ *Computer and System Engineering Department, Faculty of Engineering, Alexandria University, Alexandria, Egypt*

Abstract— **The need for linguistic resources and tools for natural language processing (NLP) is constantly increasing due to the information revolution and technological development. Researchers in the field of language engineering have suffered from little or no presence of these resources. Bibliotheca Alexandrina (BA) has adopted a center for building linguistic resources and tools needed for natural language processing tasks and applications in order to contribute in building computational applications to keep pace with the huge technological development. This paper review the efforts of the ICT sector in BA over the last 15th year.**

Keywords: **Arabic linguistic resources, Arabic corpus, computational lexicon, Arabic automatic diacritization, Arabic automatic summarization, semantic annotation, tools for NLP.**

1 INTRODUCTION

There is no doubt that the technological development that we are witnessing in our live has begun to penetrate into our public and private worlds and Interject itself into our most personal matters with our desire or against our will. This technology was not only limited to man, but also included all aspects of life, and entered the worlds of languages, literatures, science, information, mathematics and philosophy, etc.[1]. And to keep up with this phenomenal development and deal with it the computer science has emerged, which constituted a specific leap in the field of science and technology, and this science has imposed itself urgently in all fields of knowledge, and this computer science or computer engineering has become the axis around which all aspects of human life revolve. This led to the emergence of language engineering or natural language processing. Natural language processing (NLP), including Information Retrieval, Machine Translation and other Natural Language-related disciplines, is showing more interest in the Arabic language in recent years.

Interest began in the engineering of the Arabic language, more than two decades ago, Suitable resources for Arabic are becoming a vital necessity for the progress of this research, but it was individual efforts and not organized and institutional work as the researcher and his team work, write and conduct his experiments and present important and effective results, and the one who comes after does not complete what he begins with then he starts from scratch, so the efforts Scattered, refined and construction horizontally, not vertically [1]. This resulting in the lack of linguistic resources and tools that can be adopted for building intelligent applications. For example, and not limited to, corpora are an important resource but Arabic lacks sufficient resources in this field, so a research projects need to compile a corpus, which represents the state of the Arabic language at the present time and the needs of end-users. Therefore many trials have been conducted to build Arabic corpora but some of them were unsuccessful trials and others were for commercial purposes. Another problem is the share of poorly resourced languages like Arabic in contributing to the field of computational linguistics and lexical resources is much less than the share of more well-resourced languages like English. Therefore, concerted efforts must be made and institutions and bodies supporting the field strive to establish research centers that work to build resources and tools.

In this regard, in 2005, Bibliotheca Alexandrina adopted the idea of building language tools and resources to serve the Arabic language in general and the field of Arabic natural language processing in particular. This is one of its objectives "A leading institution of the digital age" as it is a center of excellence in the production and dissemination of knowledge. ICT sector in Bibliotheca Alexandrina built a trained and dedicated team of linguists and engineers as a human resources and the infrastructure in order build the linguistic resources and tools needed for Arabic NLP tasks and applications. In this paper we will review the efforts of the ICT sector over the last 15th year. Section 2 discusses the linguistic resources that have been built, section 3 shows the tools that have been developed. Section 4 is a conclusion.

2 LINGUISTIC RESOURCES

A. International Corpus of Arabic (ICA)

Bibliotheca Alexandrina (BA) has initiated a big project to build the "International Corpus of Arabic (ICA)", a real trial to build a representative Arabic corpus as being used all over the Arab world to support research on Arabic. The International Corpus of Arabic is planned to contain 100 million words. It is planned to be analyzed morphologically, syntactically and semantically. The collection of samples is limited to written Modern Standard Arabic selected from a wide range of sources designed to represent a wide cross-section of Arabic; it is stimulating the first systematic investigation of the national variety as being used all over the Arab world [2].

It is designed to include 11 genres, namely; Strategic Sciences, Social Sciences, Sports, Religion, Literature, Humanities, Natural Sciences, Applied Sciences, Art, Biography and Miscellaneous which are in turn further classified into 24 sub-genres, namely; Politics, Law, Economy, Sociology, Islamic, Pros etc. Moreover, there are 4 sub-sub-genres, namely; Novels, Short Stories, Child Stories and plays

Planning of ICA data collection is based on some criteria related to corpus design such as representativeness, diversity, balance and size were taken into the consideration. In collecting a corpus that represents the Arabic Language, the focus was to cover the same genres from different sources and from all around the Arab nations. Almost all publications of the Arab nations have been covered and other publications from outside the Arab nations as al-Hayat magazine which is published in London.

ICA data is composed of Modern Standard Arabic (MSA) written texts. There are different resources for compiling the data. It is composed of four sources, namely; (1) Press source that is divided into three sub-sources, namely; (a) Newspapers, (b) Magazines and (c) Electronic. (2) Net articles, (3) Books and (4) Academics.

Corpus analysis is both qualitative and quantitative. One of the advantages of corpora is that they can readily provide quantitative data that intuitions cannot provide reliably. The use of quantification in corpus linguistics typically goes well beyond simple counting. Table 1 shows some of the numbers of ICA data coverage. It must be noted that total number of "Tokens" refers to all word forms except numbers, foreign words and punctuations to reflect the real size of the used word forms before the analysis stage.

TABLE I

QUALITATIVE LINGUISTIC ANALYSIS FOR ICA STATISTICS

Statistics	Total Number
No. of texts	70,022
No. of words	79,569,384
No. of Tokens	76,199,414
No. of Types	1,272,766
No. of ICA sources	4
No. of sub sources	3
No. of genres	11
No. of sub genres	24
No. of sub sub-genres	4
No. of countries	20
No. of covered years	22

No. of writers	1021

B. MASAR: A Morphologically Annotated Gold Standard Arabic Resource

The first stage of linguistic analysis of the International corpus of Arabic is to analyze the 100 million words of the ICA corpus morphologically that began at Bibliotheca Alexandrina (BA) in 2007. Before beginning the morphological analysis process, each text is preprocessed and marked up with some structural markup such as beginning and end of document, title, paragraph or question. It was preferred to develop our own morphologically annotated gold standard analyzed resource to be used while analyzing the whole ICA data, since it contains more information and details than any other annotated data. It has two releases; the first one consists of about 500 thousand manually annotated words and the second one consists of about 1.5 million automatically annotated words using BASMA [3] that are verified for quality assurance. In this paper, the first release is our concern. The following is the description of the selected data of the first release and the issues that were faced during the analysis process:

The stem-based approach (concatenative approach) has been adopted as the linguistic approach to analyze the ICA. The second version of Buckwalter Morphological Arabic Morphological Analyzer (BAMA 2.0) [4] has been selected since it is a well-known analyzer in the literature and has even been considered as the "most respected lexical resource of its kind". Although Buckwalter has many advantages including its ability to provide a lot of information such as Lemma, Vocalization, Part of Speech (POS), Gloss, Prefix(s), Stem, Word Class, Suffix(s), Number, Gender, Definiteness and Case or Mood, it does not always provide all the information the ICA requires, and in some cases, the analyses provided would need some modification. Its results may give the right solution for the Arabic input word, provide more than one result that needs to be disambiguated to reach the best solution, provide many solutions, but none of them is right, segment the input words wrongly without taking the segmentation rules in consideration or provide no solutions. Consequently, solutions enhancement is needed in these situations. For more details about BAMA problems and how these problems have been handled see [5].

As a results of BAMA's problems, Number, gender and definiteness need to be modified according to their morph-syntactic properties. Some tags had been added to BAMA's lexicons, some lemmas and glossaries had been modified and others had been added. In addition, three new features had been used while developing MASAR; Name Entity, Root and Stem Pattern.

In the first release of MASAR 1,111 text documents are selected from ICA corpus for texts that were published in 2006-2007. It contains 570,137 tokens of which 69,937 are punctuations, numbers, and Latin strings, and 500,200 are Arabic word tokens (81,487 word types). These texts are selected from different sources in ICA; Press, Net Articles and Books. Moreover, these selected texts covered more than one genre as Table 1 shows. In Press Source, the texts are selected from newspapers, magazines and electronic press covering different countries.

The data of MASAR have been morphologically annotated by ten well-trained linguistic annotators using their linguistic information behind the knowledge of traditional MSA grammar. For a quality control comparison of annotators, nine files with total of 9,153 words (and varying number of POS choices per word) were each tagged independently. Out of 9,153 words, 449 words show some disagreement. All three agreed on 89% of the words; the pairwise agreement is at least 94.8%.

Once the annotation process is done, the annotated files are saved in a database in way where each feature is saved separately in order to ease the next stages of syntactic and semantic analysis processes as shown in figure 1.

word	lemmai	voc	gloss	pr1	stem	suf1	gen	num	def	casee	root	Stem_Pattern
في	fiy	fiy	in		fiy/PREP						NONE	NONE
أثناء	vanaY	>avonA'i	during		>avonA'/I		FEM	PL_BR	DEF (EI	i/GEN	vny	>afoEaAl
توجههم	tawaj~uh	tawaj~uhi	attitude		tawaj~uh/	him/PO	MASC	SG	DEF (EI	i/GEN	wjh	tafaE~ul
بسيارته	say~Arap	bisay~Ara	by/with	bi/PRE	say~Ar/N	at/NSU	FEM	SG	DEF (EI	i/GEN	syr	faE~aAl
إلى	<ilaY	<ilaY	to/tow		<ilaY/PRE						NONE	NONE
مدرستهم	madorasa	madorasa	school +		madoras/	at/NSU	FEM	SG	DEF (EI	i/GEN	drs	mafoEal
في	fiy	fiy	in		fiy/PREP						NONE	NONE
شارع	$AriE	$AriEi	street		$AriE/NO		MASC	SG	DEF (EI	i/GEN	$rE	faAEil
المدارس	madorasa	AlmadAri	the + sc	Al/DET	madAris/		FEM	PL_BR	DEF	i/GEN	drs	mafaAEil
بحي	Hay~	biHay~i	by/with	bi/PRE	Hay~/NO		MASC	SG	DEF (EI	i/GEN	Hyy	faEol
الرمال	ramol	Alr~imAli	the + sa	Al/DET	rimAl/NO		FEM	PL_BR	DEF	i/GEN	rml	fiEaAl
المكتظ	mukotaZ~	Almukota	the + ov	Al/DET	mukotaZ~		MASC	SG	DEF	i/GEN	kZZ	mufotaEal/mufotaEil
بالمدارس	madorasa	biAlmadA	with/by	bi/PRE	madAris/		FEM	PL_BR	DEF	i/GEN	drs	mafaAEil
الابتدائية	{ibotidA}i	Al{ibotid	the + el	Al/DET	{ibotidA}i	ap/NSU	FEM	SG	DEF	i/GEN	bd'	{ifotiEaAliy~
غرب	garob	garoba	west/W		garob/NC		MASC	SG	DEF (EI	a/ACC	grb	faEol
غزة	gaz~ap	gaz~ap	Gaza		gaz~ap/N		FEM	SG	DEF		NONE	NONE
.	Punc	Punc	Punc	Punc	Punc	Punc	Punc	Punc	Punc	Punc	Punc	Punc
P/	EOF_Prg	EOF_Prg	EOF_Pr	EOF_P	EOF_Prg	EOF_Pr	EOF_Prg	EOF_P	EOF_Pr	EOF_Prg	EOF_Prg	EOF_Prg
/P	BOF_Prg	BOF_Prg	BOF_Pr	BOF_P	BOF_Prg	BOF_Pr	BOF_Prg	BOF_P	BOF_P	BOF_Prg	BOF_Prg	BOF_Prg
وذكرت	*akar-u	wa*akara	and + m	wa/CC	*akar/PV	at/PVSU					*kr	faEal
مصادر	maSodar	maSAdiru	sources		maSAdir/		FEM	PL_BR	INDEF	u/NOM	Sdr	mafaAEil
أمنية	>amoniy~	>amoniy~	security		>amoniy~	ap/NSU	FEM	SG	INDEF	N/NOM	'mn	faEoliy~
فلسطينية	filasoTiyn	filasoTiyn	Palestin		filasoTiyn	ap/NSU	FEM	SG	INDEF	N/NOM	NONE	NONE
أن	>an~a	>an~a	that		>an~a/SU						NONE	NONE
مسلحين	musal~aH	musal~aH	armed/		musal~aH	iyna/NS	MASC	PL	INDEF	ACC	slH	mufaE~al
ملثمين	mulav~an	mulav~an	masked		mulav~an	iyna/NS	MASC	PL	INDEF	ACC	lvm	mufaE~al
يستقلون	{isotaqal~	yasotaqil~	they (p	ya/IV3	sotaqil~/I	uwna/IV				MOOD:I	qll	sotafoEil

Figure 1: Sample of ICA Gold Standard Resource

C. LESAN: Lexical Semantic Annotated Resource

The second stage of linguistic analysis of the International corpus of Arabic is to analyze the 100 million words of the ICA corpus on the lexical semantic level which is repented in LESAN; LExical Semantic ANnotated Resource. It is built during the process of developing the International Corpus of Arabic (ICA) and benefited from its morphological analysis stage using a built semantic lexicon. The used semantic lexicon is a lemma-based lexicon and each word is assigned the suitable lexical semantic meaning according to its context and its selected lemma/tag in the morphological analysis stage.

To detect the different meanings of the same word form, we need first to detect its different morphological analyses according to the context in which it occur as table II shows:

TABLE III
Morphological analyses of the word form ‘عين’ with some of its meaning

Morphological Analysis	Tag	Sense	Context
عَيَّنَ	Past Verb	حَدَّدَ	عَيَّنَ الْمُدِيرُ أَسَاسَ الْمُشْكِلَةِ
		وَظَّفَ – اِخْتَار - قَلَّدَ	عَيَّنَ رَئِيسُ الْعَمِلِ الشَّخْصَ الْمُنَاسِبَ لِلْوَظِيفَةِ
		خَصَّصَ	عَيَّنَ الْمَلِكُ حِصَّةً مِنْ الْمَالِ لِلْفُقَرَاءِ
عَيْن	NOUN	عُضْوُ الْإِبْصَارِ لِلْإِنْسَانِ وَغَيْرِهِ مِنْ الْكَائِنَاتِ	عَيْنُ الذُّبَابَةِ مُرَكَّبَةٌ
		جَاسُوس	وَجَعَلَهُ عَيْنَ مِصْرَ عَلَى إِسْرَائِيلَ
		نَبْع	عَيْنٌ جَارِيَةٌ

There are two competing models of lexical processing in the literature. The first proposes that we rely on mental lexicons. The second claims there are no mental lexicons; we identify certain items as words based on semantic knowledge. Thus, the former approach – the multiple-systems view – posits that lexical and semantic processing are sub-served by separate systems, whereas the latter approach – the single-system view – holds that the two are interdependent [4].

We have developed a lexicon in which the possible senses for a word are defined. Various lexicographic sources have been used as references to build the inventory of senses for each word, mainly the electronic lexicon "قاموس المعاني"[1] . The developed lexicon is lemma based; it contains 44358 lemmas. The lexicon entries cover all content words in ICA[2].

Regarding Modern Standard Arabic, some senses need to be added in the semantic lexicon and have been dealt with in the manual annotation process (section 4) due to two main reasons:

1. The word is newly used in Arabic and is not included in classical lexicons.
2. The senses of certain words are found in the classical Arabic lexicons, but the modern usage of these words require new senses to be added. For example, the word "فاجأ" "fAja>" has a new sense "أَدْهَشَ" ">adoha$" "surprise" as in example [1].

   [1] تَمَسُّكُ التُّونِسِيِّينَ بِالْإِسْلَامِ فَاجَأَ الْغَرْبَ

   tamas~uku Alt~unisiy~iyna biAl<isolAmi fAja>a Algaroba

   The Tunisians' adherence to Islam surprised the West

Once the annotation process is done, the annotated files are saved in a database in a way where the suitable sense of each word depending on the context in which it occurs is saved as shown in figure 2:

word	lemmaid	tags	Lexical_Semantic_Annotation
/P		BOF_Prg	
(		Punc	
ساحة	sAHap	NOUN	ساحة: مكان واسع
كبيرة	kabiyr	ADJ	كبير : هائل عظيم / ذو مرتبة عالية
في	fiy	PREP	
قرية	qaroyap	NOUN	قَرْيَة: عَدَدٌ قَلِيلٌ مِنَ الدُّورِ فِي بُقْعَةٍ مِنَ الأَرْضِ فِي السَّهْلِ أَوِ الجَبَلِ .
.		Punc	
نوافذ	nAfi*ap	NOUN	نَافِذَة : شُبَّاك
وأبواب	bAb	NOUN	باب: مدخل
تطل	>aTal~	IV	أطلّ: أشرف
على	EalaY	PREP	
الساحة	sAHap	NOUN	ساحة: مكان واسع
-		Punc	
بعض	baEoD	NOUN	بَعْض: جزء ، نوع من ، طائفة
الأشجار	Sajarap	NOUN	شَجَرةُ : نباتٌ يقوم على ساقٍ صُلْبة وقد يُطْلَقُ على كلِّ نبات غير قائم
الذابلة	*Abil	ADJ	ذَابِل : يابس ، أصفر ، ميت
،		Punc	
وبعض	baEoD	NOUN	بَعْض: جزء ، نوع من ، طائفة
الأشجار	Sajarap	NOUN	شَجَرةُ : نباتٌ يقوم على ساقٍ صُلْبة وقد يُطْلَقُ على كلِّ نبات غير قائم
المقطوعة	maqoTuwE	ADJ	مقطوع : مفصول بعضه عن بعض
أغصانها	guSon	NOUN	غصن: فرع ، ساق
،		Punc	
أو	>aw	CONJ	
جذوعها	ji*oE	NOUN	جذع : ساق النخلة ونحوها
،		Punc	
تفاجئ	fAja>	IV	فاجأ : جاء في وقت غير متوقع، باغت
عين	Eayon	NOUN	عَيْنٌ : عضو الإبصار للإنسان وغيره من الحيوان
الناظر	nAZir	NOUN	ناظر : باصر بعينيه
فتصدمها	Sadam-i	IV	صدَم: أفزع
.		Punc	

Figure 2: Semantically Annotated Sample

Twenty well-trained linguistic annotators have semantically annotated the data of LESAN. In order to make sure that the annotators follow the same guidelines and of almost the same level of professionalism, nineteen files with total of about 19,225 words (and varying numbers of senses choices per word) were annotated independently by each annotator and they were compared together. Out of 19, 225 words, only

---

[1] https://www.almaany.com/ [Accessed 6-2-2020]

[2] https://www.bibalex.org/ica/ar/default.aspx [Accessed 6-2-2020]

2884 words show some disagreement. All twenty annotators agreed on 85% of the words; the pairwise agreement is at least 92.3%.

D. MUHIT

There is a clear need for dictionaries translating between a large number of languages. The creation of a dictionary of good quality takes a lot of time, and given the fact that 5000-6000 languages yield 25-30 million pairs of languages, it is important to have a database that provides the possibility to translate directly between pairs of languages. A well-known problem is that words are often hard to match across languages i.e. different words from different languages do not have the same range of meanings, not all words from one language have an equivalent in the other, etc. Moreover, a multilingual lexical database should meet a number of requirements [6].

Over the last few decades, a large amount of new lexical resources have arisen: machine readable dictionaries, lexical databases, full-form lexicons, morphological databases, semantic networks, dictionary databases, etc.

The purposes of usage of a lexical database are different from those of a dictionary. MUHIT database differs from the design of dictionary databases in a number of things. Since it does not list only lemmas, but complete inflected forms as well as the amount and type of information stored for each lemma is different [7].

In this section, we will be presenting MUHIT (Multilingual Harmonized database). MUHIT database differs from the design of dictionary databases in a number of things. Since it does not list only lemmas, but complete inflected forms as well as the amount and type of information stored for each lemma is different [7].

1) What is MUHIT?:

"MUHIT" is an abbreviation for (MUltilingual Harmonized dIcTionary) but it is not just an abbreviation, it constitutes a meaningful word. The name "MUHIT" has been inspired by the Arabic word “المحيط” (al-Muhit), which means "Ocean" and "comprehensive". Moreover, it is part of one of the most celebrated Arabic dictionaries (al-Qamus al-Muhit), compiled by al-Firuzabadi (1329–1414) that has been widely used for centuries [8].

MUHIT is a multilingual electronic lexical database which has been developed within the universal networking language (UNL) framework [9], [10], [11] and it is one of the UNDL Foundation [12] products in cooperation with Bibliotheca Alexandrina. MUHIT is available on the UNLLab (http://www.unlweb.net/lab/) where entries have been interlinked by sense, and natural language word forms have been associated to a uniform concept identifier (UWs), the words of UNL [10], [13]. In 2013 MUHIT contained about 10,000,000 word forms collected from more than 40 languages, by 2019 updated and the number of words reached 20,000,000 words collected from more than 140 languages.

MUHIT was developed mainly for cross-language word search. This means that MUHIT can help users in finding and using information in their native or non–native languages. This is clear from the design of MUHIT. The methodology adopted in developing the computational lexicon “MUHIT” depends on the combination of WordNet approach and corpus based approach through many projects (BRUNO, MIR, LPP, LIS and LEWIS & SHORT) which have been developed through the UNLarium environment.

2) MUHIT linguistic infrastructure

It is important to explain the linguistic infrastructure of MUHIT. Linguistic knowledge that appears in MUHIT has been assigned to all words through UNL^{arium} encompassing different linguistic levels: morphological information, morpho-syntactic information, syntactic information and semantic information. UNL uses a standard and universal list of features (Tagset) to describe all types of the linguistic information concerning every natural language word in MUHIT. The linguistic infrastructure of MUHIT discussed in details in [14].

3) How toUse the System?

MUHIT is available on the UNL^{lab} and it is accessible through the websitehttp://www.unlweb.net/lab/. Being an online application provides the user with a number of advantages; data is stored remotely hence requiring no disk space from the part of the user, no installation or updating is required, and most importantly providing an easy access through the internet. Moreover, the user is not required to have an account in the UNL to access this application. MUHIT is a free and open source application. MUHIT provides a variety of search options for more comprehensive results. A distinct advantage that the system provides is that there is no need to choose a specific language ahead, either as a source language or as a target language, rather the system searches for the string in all existing dictionaries belonging to the different participating languages. The search does not only include base forms, which are the typical headwords of most dictionaries, but all existing inflected forms as well as. MUHIT also has adopted the regular expression system to provide a concise and flexible means for matching strings of text, such as particular characters, words, or patterns of characters. Figure 3 shows the results of searching the word "cate in MUHIT in different languages.

Figure 3: Searching the word "cate in MUHIT

The result appearing for the search word is accompanied with either three or four icons on its left. Each icon is responsible for providing the user with certain information. The information displayed are the author of the entry, The features of the entry, such as part of speech, number, gender etc., The inflections of the entry, if any. For each sense of the entry, it is also provided: The set of synonyms in the same language. The set of synonyms in different languages.

Figure 4: The lexical description of the entry

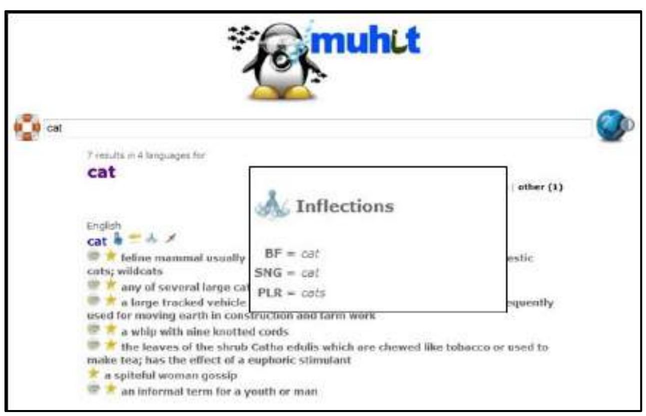


Figure 5: The inflected forms of the search word

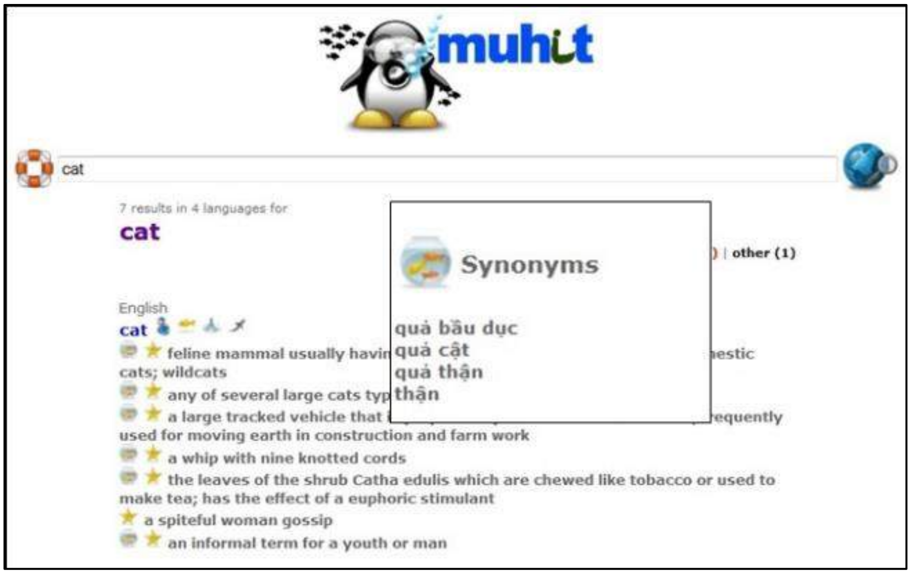


Figure 5: The synonyms of the search word

Figure 6: The available translations of the search word

The options mentioned above have combined both the advantages of a monolingual dictionary and a multilingual one. On the one hand, it provides a detailed linguistic description, that is usually found in learners dictionary only, for each result. On the other hand, it provides a translation for each single sense. MUHIT is an open initiative. Hence, it is designed in a way that makes the participation easily accessible. The participation is done on different levels. All details about MUHIT are discussed in [14].

E. Arabic Computational Lexicon

The quality of a NLP system depends to a great extent on the quality of the linguistic resources it uses, with dictionaries being the single most important variety of lexical resources as they contain an enormous amount of linguistic knowledge; lexical, semantic, morphological and syntactic [15].

Bibliotheca Alexandrina (BA) built an Arabic computational lexicon using the UNL framework. In this attempt, we sought to overcome the problems of accessibility and usability. Since most of the available lexical resources are not really available, as they are usually exclusively used within the software they are embedded in, we tried to offer a solution by making our dictionary open source thus making it a universally available lexical resource. Lexical entries in the Arabic computational lexicon are composed of an Arabic headword that is accompanied by a list of linguistic attributes. After a process of experimentation, it has been decided that the most suitable base form for our Arabic computational lexicon is the lexeme. This result has been reached after proving that a lexeme-based Arabic headword can be more efficiently and economically transformed into the required word form; plural, dual, feminine, past tense...etc. by using the linguistic information assigned to the headword within the entry. The use of a base form has eliminated the need to include all the possible word forms of a single lexical item in the dictionary; and has thus cut down the redundancy to a great degree. Moreover, this has rendered the dictionary more efficient and robust.

1) Linguistic description

This subsection presents how Arabic words are described in the Arabic computational lexicon using a list of features extracted from the UNDL Foundation tagset[3]. The Arabic computational lexicon utilizes a semantic ontology. This ontology classifies the entities existing in the natural world in a semantic hierarchy. This hierarchy points out the particular type of each concept and the kind of relation it holds with other concepts in the ontology. Each entry in this hierarchy carries a set of features and attributes and all subclasses of this concept inherit the properties of that class [16].

The linguistic information contained in the lexicon is divided into three types. 1) a list of simple features describing the lexical structure of words such as part of speech, gender, number, voice, transitivity and etc. 2) inflection paradigms to describe the morphological behavior of the Arabic words, these are a kind of morphological rules that are responsible for generating the different word forms out of the stored Arabic word as shown in figure 8.

(1) MCL&SNG:=0>"";	→	مدرس
(2) FEM&SNG:=0>"ة";	→	مدرسة
(3) MCL&DUA&NOM:=0>"ان";	→	مدرسان
(4) FEM&DUA&NOM:=0>"تان";	→	مدرستان
(5) MCL&DUA&ACC:=0>"ين";	→	مدرسين
(6) FEM&DUA&ACC:=0>"تين";	→	مدرستين
(7) MCL&DUA&GNT:=0>"ين";	→	مدرسين
(8) FEM&DUA&GNT:=0>"تين";	→	مدرستين
(9) MCL&PLR&ACC:=0>"ين";	→	مدرسين
(10) MCL&PLR&GNT:=0>"ون";	→	مدرسون
(11) FEM&PLR:=0>"ات";	→	مدرسات

Figure 7: The rules used to generate the different word forms out of the lexeme "مدرس" paradigm M532

---

[3] http://www.unlweb.net/wiki/Tagset

(1) MCL,SNG,3PS,PAS,ACV	قال
(2) MCL,SNG,3PS,PRS,ACV	يقول
(3) FEM,SNG,3PS,PAS,ACV	قالت
(4) FEM,SNG,3PS,PRS,ACV	تقول
(5) MCL,PLR,3PP,PAS,ACV	قالوا
(6) FEM,PLR,3PP,PAS,ACV	قلن
(7) MCL,SNG,2PS,PAS,ACV	قلت
(8) MCL,SNG,2PS,PRS,ACV	تقل
(9) MCL,DUA,2PP,PRS,ACV	قلتما
(10) MCL ,PLR,2PP,PAS,ACV	قلتم
(11) SNG,1PS,PRS,ACV	أقول
(12) PLR,1PS,PAS,ACV	قلنا
(13) MCL,SNG,3PS,PAS,PSV	قيل
(14) MCL,SNG,3PS,PRS,PSV	يقال
(15) FEM,SNG,3PS,PAS,PSV	قيلت
(16) FEM,SNG,3PS,PRS,PSV	تقال

Figure 8: The rules used to generate the different word forms out of the lexeme "قال" paradigm M103

2) Subcategorization rules to describe the syntactic behaviour of the words. These are the rules that determine the number and types of the necessary syntactic arguments (specifiers, complements and adjuncts) of the verb. For example in the sentence "نقل الإسكندر علوم فاس وبابل إلى أثينا"

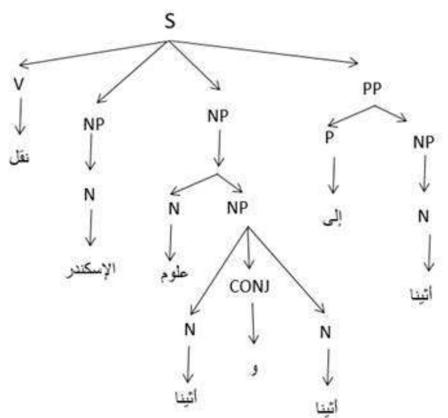


Figure 9: The necessary syntactic arguments of the verb "نقل" Subcategorization Frame Y521

All of the above mentioned features enable us to build a comprehensive Arabic lexical resource especially designed to suit the different applications of NLP. Currently, the BA Arabic computational lexicon contains about 1,500,000 Arabic headwords. Figure 11 shows a sample of the dictionary entries in their final form as they are ready to be exported.

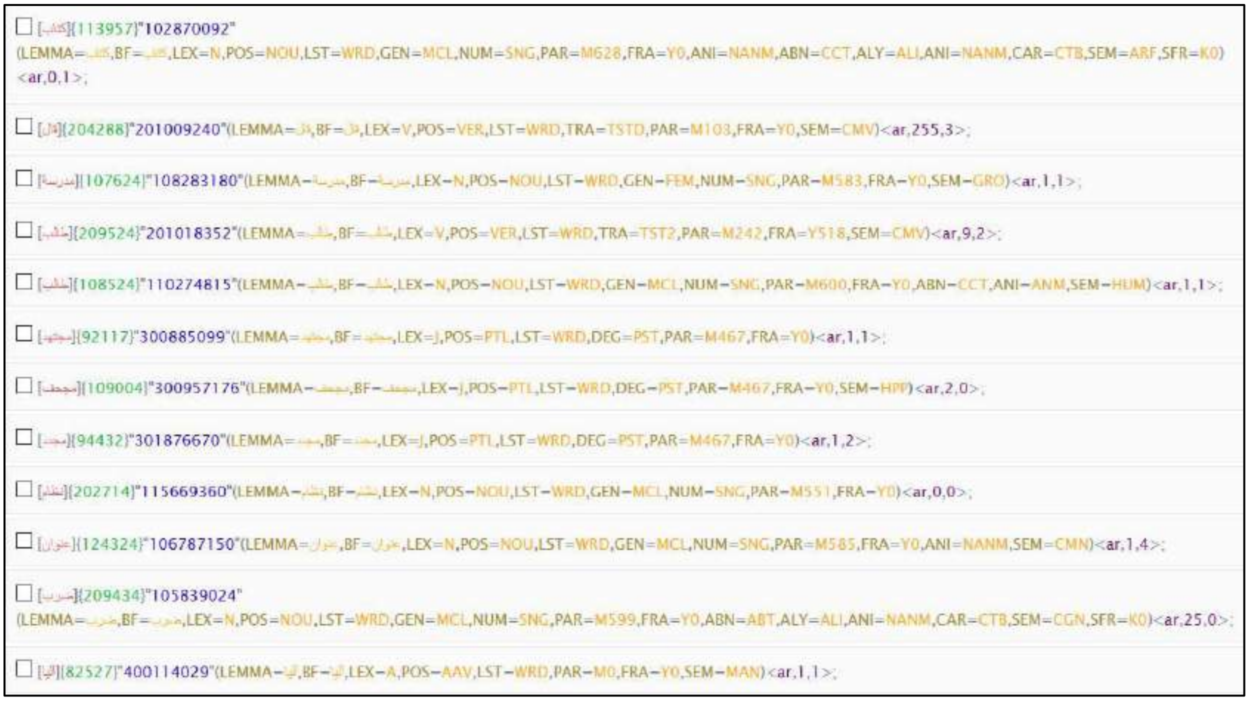

[كتاب](113957)"102870092"
(LEMMA=كتاب,BF=كتاب,LEX=N,POS=NOU,LST=WRD,GEN=MCL,NUM=SNG,PAR=M628,FRA=Y0,ANI=NANM,ABN=CCT,ALY=ALI,ANI=NANM,CAR=CTB,SEM=ARF,SFR=K0)
<ar,0,1>;
[قال](204288)"201009240"(LEMMA=قال,BF=قال,LEX=V,POS=VER,LST=WRD,TRA=TSTD,PAR=M103,FRA=Y0,SEM=CMV)<ar,255,3>;
[مدرسة](107624)"108283180"(LEMMA=مدرسة,BF=مدرسة,LEX=N,POS=NOU,LST=WRD,GEN=FEM,NUM=SNG,PAR=M583,FRA=Y0,SEM=GRO)<ar,1,1>;
[خلف](209524)"201018352"(LEMMA=خلف,BF=خلف,LEX=V,POS=VER,LST=WRD,TRA=TST2,PAR=M242,FRA=Y518,SEM=CMV)<ar,9,2>;
[خلف](108524)"110274815"(LEMMA=خلف,BF=خلف,LEX=N,POS=NOU,LST=WRD,GEN=MCL,NUM=SNG,PAR=M600,FRA=Y0,ABN=CCT,ANI=ANM,SEM=HUM)<ar,1,1>;
[[illegible]](92117)"300885099"(LEMMA=[illegible],BF=[illegible],LEX=J,POS=PTL,LST=WRD,DEG=PST,PAR=M467,FRA=Y0)<ar,1,1>;
[[illegible]](109004)"300957176"(LEMMA=[illegible],BF=[illegible],LEX=J,POS=PTL,LST=WRD,DEG=PST,PAR=M467,FRA=Y0,SEM=HPP)<ar,2,0>;
[[illegible]](94432)"301876670"(LEMMA=[illegible],BF=[illegible],LEX=J,POS=PTL,LST=WRD,DEG=PST,PAR=M467,FRA=Y0)<ar,1,2>;
[نظام](202714)"115669360"(LEMMA=نظام,BF=نظام,LEX=N,POS=NOU,LST=WRD,GEN=MCL,NUM=SNG,PAR=M551,FRA=Y0)<ar,0,0>;
[عنوان](124324)"106787150"(LEMMA=عنوان,BF=عنوان,LEX=N,POS=NOU,LST=WRD,GEN=MCL,NUM=SNG,PAR=M585,FRA=Y0,ANI=NANM,SEM=CMN)<ar,1,4>;
[ضرب](209434)"105839024"
(LEMMA=ضرب,BF=ضرب,LEX=N,POS=NOU,LST=WRD,GEN=MCL,NUM=SNG,PAR=M599,FRA=Y0,ABN=ABT,ALY=ALI,ANI=NANM,CAR=CTB,SEM=CGN,SFR=K0)<ar,25,0>;
[آليا](82527)"400114029"(LEMMA=آليا,BF=آليا,LEX=A,POS=AAV,LST=WRD,PAR=M0,FRA=Y0,SEM=MAN)<ar,1,1>;

Figure 10: The Arabic UNL dictionary

The entry in the Arabic computational lexicon contains the Arabic Headword, a number string representing the Headword meaning or rather its place in the hierarchy of concepts, a list of linguistic features including the list of simple features and the number of the inflectional paradigm that describes the morphological behavior of the headword as well as the number of the Subcategorization rule that describes it syntactic behavior, finally, the entry contains frequency and priority numbers that are used to indicate the frequency of usage of the headword; this piece of information is necessary to determine the most used senses of a particular Arabic word and can, thus, aid the process of word-sense disambiguation. All details about the description of the Arabic computational lexicon are discussed in [17].

3 TOOLS

A. *Arabic Diacritization System (Alserag)*

In Modern Standard Arabic, texts are typically written without diacritical markings. The diacritics are important to clarify the sense and meaning of words. The process of automatically restoring diacritical marks is called diacritization. Diacritization helps the reader in disambiguating the text or simply in articulating it correctly. As Arabic is a language where the intended pronunciation of a written word cannot be completely determined by its standard orthographic representation; it rather depends on a set of special diacritics. The absence of these diacritics in Arabic text increases lexical and morphological ambiguity, because one written form can have several vocalizations, each vocalization may have different meaning(s) [18, 19]. However, these diacritics are generally left out in most genres of written Arabic, which results in widespread ambiguities in vocalizations and meaning Diacritizing [20]. Arabic written text is crucial for many NLP tasks, translation can be enumerated among a longer list of applications that vitally benefit from automatic diacritization [21, 22, and 23].

Much work has been done on Arabic diacritization. The actually implemented systems can be divided into two categories [24]: Systems implemented by individuals as part of their academic activities and systems implemented by commercial organizations for realizing market applications. There are also other available systems as Mishkal Arabic diacritizer, and Harakat Arabic diacritizer; they are free Arabic diacritizers, which are available online. Finally, on March Google has launched an innovative new Google Labs Arabic tool called Tashkeel, a tool that adds the missing diacritics to Arabic text. Unfortunately, the tool is not available now [20].

The developed system (Alserag) is a rule based system. Alserag has been developed in 2016, it depends on two resources: the Arabic diacritized dictionary and a set of linguistics rules. The Arabic diacritized dictionary is a dictionary where Arabic natural language words exist with their diacritics, along with the corresponding linguistic features, which describe the Arabic word morphologically, syntactically and semantically. While the linguistic rules work through three modules in order to provide fully diacritized Arabic words namely, morphological analysis module, syntactic analysis module and morph-phonological processing module [20, 25]. These modules are achieved through 7 main phases: (i) Preprocessing which is responsible for auto-correcting the raw text and segmenting the Arabic text into sentences. (ii) Tokenization which is the process of splitting the natural language input into lexical items. (iii) Disambiguation which is a process of choosing the right internal diacritization for the word from the dictionary. (iv) Name entity recognition. (v) Syntactic shallow parsing which is an analysis of a sentence by identifying its constituents (NPs, JPs---etc.). (vi) Case ending module which is responsible for predicting the arguments of the predicate and assigning the diacritical marks that are attached to the ends of words to indicate their grammatical function. (vii) Morph-phonological module which is a series of rules that focus on the sound changes that take place in morphemes (minimal meaningful units) when they are combined to form words [20, 25].

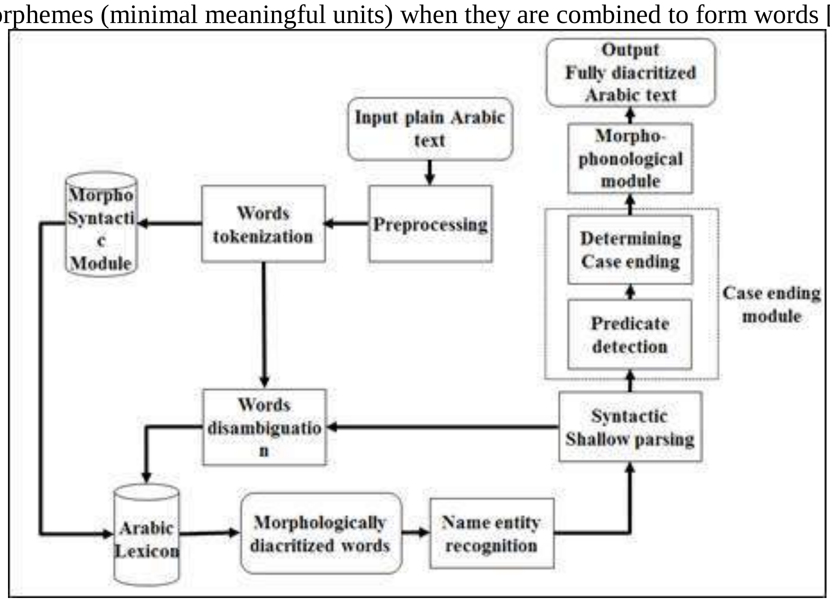


Figure 11: Architecture of ALSERAG

There are two engines that are used in Alserag, the first is Interactive ANalyzer (IAN), which is used in the analysis process, and includes a grammar for natural language analysis. The syntactic processing is done automatically through the natural language analysis grammar, the second is dEep-to-sUrface natural language GENErator engine (EUGENE) which is used in the generation process, and receives the analyzed input and provides a diacritized output without any human intervention [26].

Although the system was able to overcome many challenges, it still has some limitations. The system has been improved in this phase and still there is more potential for further improvements. One of the most difficult problems that faces a parser is structural ambiguity, since it leads to problems in determining the boundaries between constituents. Despite these limitations, the system is proved to be promising when tested and demonstrated [25].

The corpus has been selected from the International Corpus of Arabic (ICA). The selected corpus size is 400,000 Modern Standard Arabic words; they are divided into 300,000 words as training data and 100,000 words as testing data. The selected texts are from different sources; Newspapers, Net Articles and Books representing the following genres; politics: 148,211, miscellaneous: 100,253, child stories: 57,174, economy: 34,930, society: 32,955 and sports: 26,477 [20]. The results of the system were evaluated for accuracy against the reference using two metrics; diacritization error rate (DER) and word error rate (WER). In addition to calculating DER and WER, the evaluation system calculates internal diacritics and case ending separately [25]. Table IV shows the evaluation history of the automatic diacritization system (Alserag) over the last three years from 2017 to 2019 and figure 13 shows the latest evaluation results of the 100.000 words.

TABLE V

EVALUATION HISTORY OVER LAST THREE YEARS (2017-2019)

Year	Morphology	Case Ending	DER	WER
2017	2.56%	13.51%	4.86%	15.79%
2018	1.77%	9.39%	3.37%	11.7%
2019	1.53%	8.44%	2.98%	10.50%

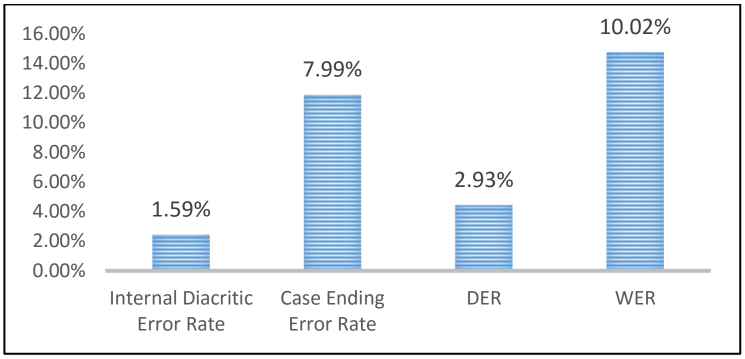


Figure 12: Evaluation of the whole data of Alserag

Alserag system, which is a rule-based system, is benchmarked against three known diacritization systems: Harakat, Mishkal, and Aldoaly, which are statistical based systems as shown in table VI [25].

TABLE VII

BENCHMARKING OF THE WHOLE DATA OF ALSERAG AMONG THE OTHER THREE SYSTEMS

	Alserag	Harakat	Mishkal	Aldoaly
Internal Diacritic Error Rate	1.59%	43.30%	32.53%	80.92%
Case ending Diacritic Error Rate	7.99%	16.23%	31.15%	89.72%
Diacritic Error Rate (DER)	2.93%	37.63%	32.24%	82.76%
Word Error Rate (WER)	10.02%	43.49%	65.00%	97.87%

According to the recent results obtained by the benchmarking process, our system scored the least error rate followed by Harakat and Mishkal and finally Aldoaly, which scored over 80% error rate.

B. Arabic Summarization System

Due to the daily increasee of the electronic documents on the Internet, everyone should benefit from this revolution of information. The important way to access these documents and get the core content of them and utilize from the information existing in these documents became an urgent need. Hence the need for tools to facilitate this appeared, which is called Automatic Text Summarization. Radev et al. (2002) defines a summary as “a text that is produced from one or more texts, that conveys important information in the original text(s), and that is no longer than half of the original text(s) and usually significantly less than that”. In order to generate a summary, we have to identify the most significant pieces of information exist in the text, omitting the redundant information and reducing details.

Text summarization strongly appeared in many applications that will help in facilitating life, for example we do not need to consume time in reading news with all its details, so we can summarize news to SMSs or to news mobile applications that will save a lot time. In addition, businessmen need fast summaries for their reports and documents, etc. Therefore, when we extract this summarized version from a certain document by means of a computer, automatically, we call this Automatic Text Summarization.

There are two main approaches of summarization, extraction approach, where all sentences are first rated according to their importance, and then a summary is generated after choosing a number of top scoring sentences. On the other hand, the abstraction approach in which documents have to be recognized and interpreted, and then the summary is generated.

Arabic language is more sophisticated and needs much time compared with English and other European languages. There are different existing systems of automatic text summarization, which classified into two categories: Systems implemented on Arabic languages, such as AQBTSS, ACBTSS, Sakhr Arabic

Summarizer, Lakhas, Aramedia, Ikhtasir. Moreover, other systems implemented on other different languages, such as Copernic Summarizer, Pertinence summarizer, Kify Text Summarizer, SweSum, MEAD.

In a series of developments, an Automatic Arabic Text Summarization System is proposed, this system uses different stages in order to summarize Arabic documents. The first stage is responsible for extracting the most informative sentences of the document. There are many factors should be considered while selecting the important sentences such as: the type of the document, the sentence length, word frequency, and the number of topic words the sentence contain, they are all helping factors in determining the most important sentences. The second stage is responsible for excluding the non-informative constituents after identifying which constituents are incident and could be omitted, and which are principal and vital. The system represents the constituents syntactic structure of each sentence using X-bar theory in order to categorize them into incident and principal. Our proposed automatic summarizer adopts the X-bar theory, because of its compatibility with Arabic language, and because it supplies a systematic description of Standard Arabic sentence formation.

There are two engines that are used in our proposed Arabic automatic summarization system during the linguistic processing stage, the first is Interactive ANalyzer (IAN) which represents the analysis process, the second is dEep-to-sUrface natural language GENErator engine (EUGENE) which represents the generation process.

The system depends on three linguistics resources, namely; the word-net, UNL encycolopedia, and EDGES. EDGES: is the Entity Discovery and Graph Exploration System, a user-friendly visualization tool used for exploring semantic networks by enabling concept (words of the document) expansion, collapsing and navigation.

UNL Encyclopedia: It is also known as UNL Example Base; it contains semantic relations between UWs along with a degree of probability. It comprises information that is related to the probability of occurrence rather than the possibility of occurrence.

There is an inventory study has been done on a number of documents in order to measure the accuracy of our proposed system and the established summarization grammar. A sample of this test was taken from a random document whose total number is 115 words as shown in figure 14, while when automatically summarized it becomes 43 words as in figure 15. The summarization grammar could detect that some constituents, which have specific grammatical functions, could be omitted from a document during the summarization process, such as the highlighted constituents in the document in figure 14 which might be Appositions, phrases starts with specific keywords, Relative clauses. All of those do not add additional information to the sentence`s meaning.

أكد المعلق الرياضي أحمد الطيب، أن محمد أبو تريكة لاعب الأهلي المعتزل، رفض تقاضي أي أموال نظير مشاركته مع نجوم العالم في احتفال دولة الكويت بافتتاح ملعب جابر الدولي.

وقال الطيب لمصراوي: "أبو تريكة اعتذر عن تقاضي مبلغ 50 ألف يورو نظير مشاركته في الاحتفال، وقال إنه لشرف أن يشارك في يوم عظيم وحدث كبير بدولة الكويت".

وضمت تشكيلة نجوم العالم كل من: "الحارس جيمس ليمان، والحارس ديفيد جميز، وديفيد بيكهام، ميشيل سلجادو، اليساندرو نيستا، وكارلوس بويول، وزامبروتا، وريرتو كارلوس، ولويس فيجو، وروبيرتو بيرس، رونالدينهو، واليساندرو ديل بيرو، وداوت يورك، أندريه شفشينكو، ومحمد أبو تريكة".

وطلب الشيخ صباح الأحمد الجابر الصباح مصافحة أبو تريكة بين شوطي المباراة، التي انتهت بفوز نجوم الكويت برباعية مقابل هدفين.

Figure 13: Original Document

أكد أحمد الطيب، أن محمد أبو تريكة لاعب الأهلي المعتزل، رفض تقاضي أي أموال نظير مشاركته مع نجوم العالم في احتفال دولة الكويت بافتتاح ملعب جابر الدولي.

وضمت تشكيلة نجوم العالم

وطلب الشيخ صباح الأحمد الجابر الصباح مصافحة أبو تريكة بين شوطي المباراة

Figure 14: The automatic summarized document

There are many tools and measures that could be used in the evaluation of the automatic summary; however, many researches admitted that there are two main points have to be considered in the evaluation, the Compression Ratio (CR) and the Omission Ratio (RP). Therefore, our proposed automatic summarizer follows this pace and the results are significantly accurate.

C. BASMA : BibAlex Standard Arabic Morphological Analyzer

The process of developing a morphological analyzer tool for ICA began in 2007 which is known as BibAex Arabic Morphological Analyzer Enhancer (BAMAE) but later known as BibAlex Standard Arabic Morphological Analyzer (BASMA). It is a system that has been built to morphologically analyze and disambiguate the Arabic texts depending on BAMA's enhanced output of ICA. It was preferred to use BAMA's enhanced output of ICA since it contains more information than any other systems of BAMA's enhanced output. This is the reason why the members of ICA team aimed to build their own morphological disambiguator (BAMAE), figure 16 shows the architecture of BASMA.

In order to reach the best solution for the input word, BAMAE preforms automatic disambiguation process carried on three levels, depends primarily on the basic POS information (Prefix(s), Stem, Tag and Suffixes) obtained from enhanced BAMA's output:

- Word level which avoids or eliminates the impossible solutions that Buckwalter provides due to the wrong concatenations of prefix(s), stem and suffix(s).
- Context level where some linguistic rules have been extracted from the training data to help in disambiguating words depending on their context.
- Memory based level which is not applicable in all cases; it is only applicable when all the previous levels failed to decide the best solution for the Arabic input word.

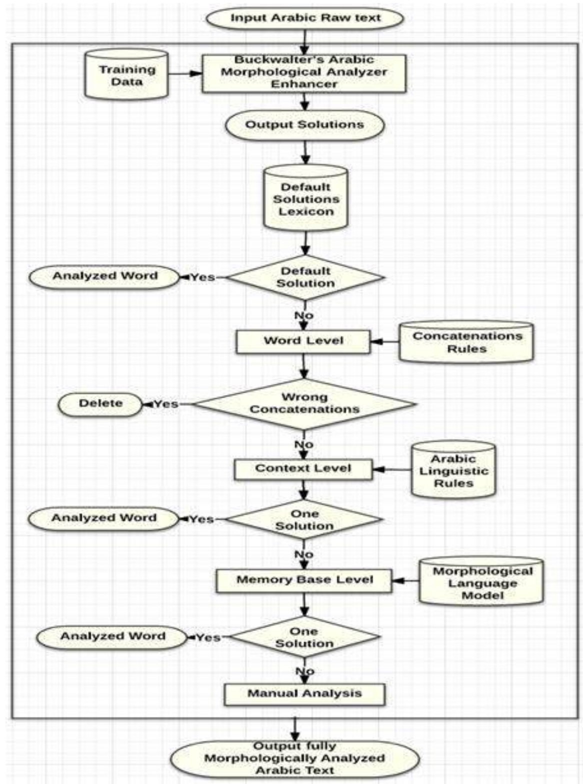


Figure 15: BAMAE Architecture

After selecting the best POS solution for each word, BAMAE detects the rest of information accordingly. It detects the lemmas, roots (depending primarily on the lemmas), stem patterns (depending on stems, roots and lemmas), number (depending on basic POS and stem patterns), gender (depending also on basic POS, stem patterns and sometimes depending on number), definiteness (depending on POS or their sequences), case (depending on definiteness and sequences of POS) and finally it detects the vocalization of each word. For more details see [3].

D. UNL Editor (An Annotation tool for Semantic Analysis)

Semantic Annotation has become an increasingly important research topic being a fundamental element of many Natural Language Processing applications like information retrieval, query answering and information extraction. Semantic annotation is additional information in a document that identifies or defines the semantics of a part of that document [27].

In the context of the UNL (The Universal Networking Language), a semantically based interlingua to break language barriers between human languages, the UNDL Foundation in cooperation with Bibliotheca Alexandrina has started an initiative for building a tool for semantic annotation called the UNL Editor; a visual editor designed with the intention of providing full semantic annotation, thus analyzing natural language texts and, generating UNL documents. This tool is based upon a comprehensive visualization of the entire process of the annotation. It is uniquely designed on linguistic background; adopting certain linguistic theories closely related to computational linguistics in terms of using unified super sets of semantic relations [28] thus overcoming the problem of conflicting and confusing names [29], and making use of renowned lexical recourses; WordNet [30]. Moreover, it provides a powerful visual interface for working with UNL data both in a textual and graphical mode with friendly interface creating an appropriate environment for navigating through the needed steps of providing the analysis; it offers a visualization of the analysis through graphs which aids the representation of the semantic network created with every sentence analyzed.

The UNL Editor provides a means enabling the analysis of the underlying semantic relations composing the Natural Language sentences. It is designed on linguistic bases. On a semantic assumption or rather on semantic theory stating that a deep semantic analysis for a natural language text requires two levels of semantics; lexical semantics and grammatical semantics and it is discussed in details in S. Alansary, M. Nagi and N. Adly, 2011 [31]. Lexical semantics in UNL Editor is expressed through creating the nodes, a process in which every word or rather every concept in the sentence to be analyzed is matched with its corresponding ID, meaning that a single node may contain more than one lexical item; a compound word, as long as it is representing a single concept. Grammatical Semantics in the UNL Editor is expressed in terms of a range of semantic relations, and a list of attributes. UNL Editor has proposed a unified super set of the semantic relations. These relations are highly standardized as each relation is clearly defined in the UNL framework. The tool includes 45 semantic relations and they are a closed set of relations. Relations are used to describe the objectivity information of sentences. In the UNL, relations are normally regarded as representations of semantic cases or thematic roles (such as agent, object, instrument, etc.) between concepts. They are used in form of arcs connecting a node to another node in a UNL graph. They correspond to two-place semantic predicates holding between two concepts. Relations are always used to describe semantic dependencies between syntactic constituents. For more information about the semantic relation within the UNL frame work see [32].

Other additional information are being presented through attributes, representing information conveyed by natural language grammatical categories (such as tense, mood, aspect, number, etc) [33]. In opposition to relations, attributes correspond to one-place predicates; attributes are intended to be used as annotations made to nodes or hypernodes of a U

NL hypergraph. Moreover, they are also a closed set. Attributes modify concepts or semantic networks to indicate subjectivity information such as about how the speaker views these states-of-affairs and his attitudes toward them and to indicate the property of the concepts, for more information about the attributes within the UNL frame work see [32].

1) How to Use the UNL Editor?

In order to use the tool, the user will have to sign in the UNL web then access the UNL Editor via UNL dev application (The UNL Integrated Development Environment). Figure 17 describes the steps for reaching the semantic graphic representation. Within the UNL Editor Frame work, the process of decision making is completely human: the user uploads the text to be analyzed; selects the corresponding IDs; relate nodes through creating semantic relations; and assigns attributes to nodes. The first step will be the text input and text segmentation followed by concepts selection to create the nodes and adding the appropriate attributes to each node then the final step in order to reach the semantic graph will be linking the created nodes by semantic relations [34].

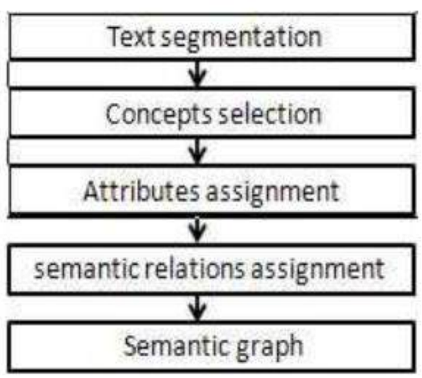
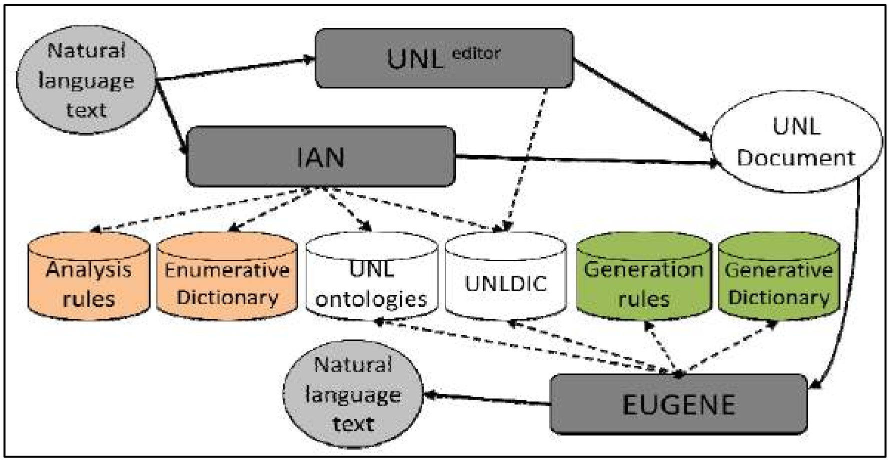


Figure 16: Steps for reaching the semantic graph

The detailed description of how the UNL Editor works is illustrated with examples and figures in [35].



4 CONCLUSIONS

The paper shed light on the achievements and contributions of Bibliotheca Alexandrina in the field of language engineering over the last 15th years. Where Bibliotheca Alexandrina made efforts to build infrastructure and human resources to work on building linguistic resources and tools to be available for researchers and specialists in the field of language engineering. The paper briefly discussed these resources and tools.

REFERENCES

[1] N. Ahmed "هندسة اللغة العربية: مطلب قومي وهدف استراتيجي" Arabiyât: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban, 4, (1), 2017, 88-101.

[2] S. Alansary, M. Nagi and N. Adly, *Building an International Corpus of Arabic (ICA): progress of compilation stage.* In proceedings of the 7th International Conference on Language Engineering. Cairo, Egypt, 5–6 December 2007.

[3] S. Alansary, *BASMA: BibAlex Standard Arabic Morphological Analyzer*. In the proceedings of (ESOLE). Egypt, Cairo, 9-10 December.

[4] T. Buckwalter, *Buckwalter Arabic Morphological Analyzer Version 2.0. Linguistic Data Consortium*, University of Pennsylvania, 2004. LDC Catalog No.: LDC2004L02.

[5] S. Alansary, *BAMAE: Buckwalter Arabic Morphological Analyzer Enhancer*. In proceedings of Arabic Language Processing Conference. Rebate, Morocco, 2-3 May: Mohamed Vth University.

[6] M. Janssen, "*SIMuLLDA*: *a Multilingual Lexical Database Application using a Structured Interlingua*", Doctoral dissertation, Utrecht University, June, 2002.

[7] M. Janssen. *Lexical vs. Dictionary Databases: design choices of the MorDebe system. Papers in Computational Lexicography* - COMPLEX, Budapest, Hungary, 2005.

[8] The UNLweb website: http://www.unlweb.net/muhit/index.php?muhit=help, (accessed in October 2013).

[9] H. Uchida, M. Zhu, T. G. Della Senta, "A Gift for a Millennium", November 1999.

[10] S. Alansary, M.Nagi, N.Adly, UNL+3: The Gateway to a Fully Operational UNL System. In Proceedings of 10th International Conference on Language Engineering, Cairo, Egypt, 2010.

[11] J. Cardeñosa, A. Gelbukh, E. Tovar (eds.): Universal Networking Language: advances in theory and applications.(Research on Computer Science, 12). Mexico City: National Polytechnic Institute. 443pp, 2005.

[12] The UNDL Foundation website: www.undl.org

[13] R. Martins, V. Avetisyan, "Generative and Enumerative Lexicons in the UNL Framework," in Proc. Of Seventh International Conference on Computer Science and Information Technologies (CSIT 2009), 28 September - 2 October, 2009, Yerevan, Armenia Proceedings of CSIT 2009.

[14] S. Alansary, *MUHIT: A Multilingual Lexical Database*, 13th International Conference on Language Engineering, Ain Shams University, Cairo, Egypt, December 11 - 12 2013.

[15] P. Bjakman and V. Raskin. *What Linguists Might Contribute to Dictionary Making If They Could Get Their Act Together*. In The Real World Linguistics. Ed. Ablex, Norwood, NJ, 1986.

[16] M. Obitko, *Ontologies description and applications*, Research Report No. 126/01Czech Technical University, Praue, 2001.

[17] S. Alansary, *A UNL Based Approach for Building an Arabic Computational Lexicon* , INFOS2012, Cairo University - Egypt, May 16 - 17 2012.

[18] Bouamor. H., Zaghouani.W., Diab.M., Obeid. O., Oflazer. K., Ghoneim.M., and Hawwari. A.: A Pilot Study on Arabic Multi-Genre Corpus Diacritization Annotation. Proceedings of the Second Workshop on Arabic Natural Language Processing, pages 80–88, Beijing, China, c2014 Association for Computational Linguistics (2015).

[19] EL-Desoky. A., Fayz. M. and Samir, D.*: A smart Dictionary for the Arabic Full-Form Words*. (IJSCE). ISSN: 2231-2307, Volume-2, Issue-5 (2012).

[20] S. Alansary, "*Alserag: An Automatic Diacritization System for Arabic*". The 2nd International Conference on Advanced Intelligent Systems and Informatics (AISI'16), Cairo, Egypt, 2016.

[21] Smr, O.: Yet Another Intro to Arabic NLP,http://ufal.mff.cuni.cz/~smrz/ANLP/anlp-lecture-notes.pdf (2005).

[22] Rashwan., M., Abdou.S., Rafea., A.: *Stochastic Arabic Hybrid Diacritizer*, IEEE trans. Natural Language Processing and Knowledge Engineering, pp.1-8, 24-27 (2009).
[23] Attia, M., Mohsen A. A. Rashwan, Mohamed A.S. A. A. Al-Badrashiny.: Fassieh®, *a Semi-Automatic Visual Interactive Tool for Morphological*, PoS-Tags, Phonetic, and Semantic Annotation of Arabic Text Corpora, IEEE trans. AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 17, NO. 5, pp.916-925 (2009).
[24] Al Badrashiny, M.: *Automatic Diacritizer for Arabic Text*. A Thesis Submitted to the Faculty of Engineering, Cairo University in Partial Fulfillment of the Requirements for the Degree of master of science in electronics & electrical communication (2009).
[25] S. Alansary, (2016, December). Improving Alserag Arabic Diacritization Grammar through Syntactic Analysis. In 16th international conference on language engineering, Cairo, Egypt.2016.
[26] Alansary. S.: A Suite of Tools for Arabic Natural Language Processing: A UNL Approach, the special session on Arabic Natural Language Processing: Algorithms, Resources, Tools, Techniques and Applications, (ICCSPA'13), Sharjah, UAE (2013).
[27] H. Bunt and Ch. Overbeeke," *A note on the definition of semantic annotation Languages*" , Proceedings of the 8th International Conference on Computational Semantics, pages 268–271, Tilburg, January 2009. c 2009 International Conference on Computational Semantics 2009.
[28] Ch. Johnson and Ch. J. Fillmore: "T*he FrameNet tagset for frame-semantic and syntactic coding of predicate-argument structure*". In the Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics (ANLP-NAACL 2000), Seattle WA, pp. 56-62, April 29- May 4, 2000.
[29] D.R Dowty, *Thematic Proto-Roles and Argument Selection*, Linguistic Society of America, 1991.
[30] C. Fellbaum, *WORDNET: An Electronic Lexical Database*, the MIT Press, 1998.
[31] N. Chomsky, *Studies on Semantics in Generative Grammar*, Mouton publisher, 1972.
[32]Sameh Alansary, Magdy Nagi, Noha Adly , *UNL+3: The Gateway to a Fully Operational UNL System* , 10th International Conference on Language Engineering, Ain Shams University, Cairo, Egypt, December 22 - 23 2010.
[33] D. Pisoni and R. Remez, *The Handbook of Speech Perception*, Blackwell Publishing Inc., 2004.

[34] M.Zhu and H.Uchida, "UNL annotation", UNL Center, UNDL Foundation specifications and manuals, 2003.

[35] S. Alansary, M. Nagi, N. Adly *, UNL Editor: An Annotation tool for Semantic Analysis* , 11th International Conference on Language Engineering, Ain Shams University, Cairo, Egypt, December 14 - 15 2011.

BIOGRAPHY

Dr. Sameh Alansary: *Director of Arabic Computational Linguistic Center at Bibliotheca Alexandrina, Alexandria, Egypt.*

He is professor of computational linguistics in the Department of Phonetics and Linguistics and the head of Phonetics and Linguistics Department, Faculty of Arts, Alexandria University. He obtained his MA in Building Arabic Lexical Databases in 1996, and his PhD from Nijmegen University, the Netherlands in building a formal grammar for parsing Arabic structures in 2002. His main areas of interest are concerned with corpus work, morphological analysis and generation, and building formal grammars.

He is also the head of Arabic Computational Linguistics Center in Bibliotheca Alexandrina. He is supervising and managing the Universal Networking Language project in Library of Alexandria since 1-6-2005 till now.

Dr. Alansary is the co-founder of the Arabic Language Technology Center (ALTEC), an NGO aims at providing Arabic Language resources and building a road map for Arabic Language Technology in Egypt and in the Middle East. He has many scientific works in Arabic Natural Language Processing published in international conferences and periodicals, and a member in many scientific organizations: (1) Egyptian Society of Language Engineering, Cairo, (2) Arabic Linguistic Society - USA, (3) Association of Computational Linguistics - USA – Europe, (4) Universal Networking Language foundation, United Nations, Geneva, Switzerland.

Dr. Magdy H. Nagi: *Senior Consultant, ICT Sector at Bibliotheca Alexandrina*

Dr. Nagi is a Professor in the Computer and Systems Engineering department, Faculty of Engineering, Alexandria University. He obtained his Ph.D. from the University of Karlsruhe, in 1974, where he served as Lecturer for two years and as a Consultant to its Computer Center from 1974-1990. During this period, he also served as Consultant to many companies in Germany such as Dr. Otker, Bayer, SYDAT AG, and BEC.

On the national level, he was a Consultant to many projects under the umbrella of either the University of Alexandria or the Faculty of Engineering for designing and/or implementing automation projects for governmental authorities or public sector companies, such as the Ministry of Interior, the Health Insurance Organization (HIO), the Social Insurance Organization (SIO), and the Customs Authorities.

Dr. Nagi has served, since 1995, as Consultant to the Bibliotheca Alexandrina. Among his activities were the design and installation of Bibliotheca Alexandrina's network and information system, namely a trilingual information system that offers full library automation.

In 2001, he got appointed as the Head of the Information and Communication Technology (ICT) Sector of the Bibliotheca Alexandrina and occupied that post till 2012. He currently serves as a senior Consultant to the ICT Sector and continues to oversee the various projects and partnerships established between the ICT Sector and many international institutions.

Dr. Nagi is a member of the ACM and the IEEE Computer Society as well as several other scientific organizations. His main research interests are in operating systems and database systems. He is author/co-author of more than 80 papers.

TRANSLATED ABSTRACT

المصادر والأدوات اللغوية العربية لمكتبة الإسكندرية في مجال هندسة اللغة

تتزايد الحاجة إلى الموارد والأدوات اللغوية لمعالجة اللغة الطبيعية (NLP) باستمرار بسبب ثورة المعلومات والتطوير التكنولوجي .لقد عانى الباحثون في مجال هندسة اللغة من قلة أو عدم وجود هذه الموارد. وقد تبنت مكتبة الإسكندرية (BA) مركزًا لبناء الموارد والأدوات اللغوية اللازمة لمهام وتطبيقات معالجة اللغة الطبيعية من أجل المساهمة في بناء التطبيقات الحاسوبية لمواكبة التطور التكنولوجي الضخم. تستعرض هذه الورقة جهود قطاع تكنولوجيا المعلومات والاتصالات في مكتبة الإسكندرية على مدار الخامسة عشر عاما الماضية منذ عام ٢٠٠٥ وحتى عام ٢٠٢٠.

Shallow Parsing for Automatic Arabic Text Summarization

Sameh Alansary

Bibliotheca Alexandrina, Alexandria, Egypt
Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt
sameh.alansary@bibalex.org

***Abstract*—in this paper, we propose a summarization method from the viewpoint of the shallow syntactic analysis involves different stages: sentences extraction stage, syntactic analysis stage, and generation stage. Finally, the evaluation summarization process will discussed.**

Keywords: **Text summarization (TS), text analysis, parts of speech, shallow parsing and abstractive summarization.**

1 INTRODUCTION

The automatic text processing is a research field that is currently extremely active. Automatic text summarization, aims to reduce the size of a text while preserving its information content. A summarizer is a system that produces a condensed representation of its input's for user consumption [1]. In order to generate a summary of a document or a sentence, we have to identify the most significant pieces of information exist in a document or a sentence, omitting the redundant information and reducing details. A summary can be employed in an indicative way – as a pointer to some parts of the original document, or in an informative way – to cover all relevant information of the text . In both cases, the most important advantage of using a summary is its reduced reading time [1]. Summary generation by an automatic procedure has also other advantages: (i) the size of the summary can be controlled; (ii) its content is determinist; and (iii) the link between a text element in the summary and its position in the original text can be easily established [1].
There different systems of automatic text summarization for different languages are listed and discussed in the previous work [2]. The actually implemented systems can be divided into two categories: Systems implemented on Arabic languages and other systems implemented on other languages rather than the Arabic:

There are number of Arabic text summarization systems as referred in the previous study [2]: the Arabic Query-Based Text Summarization System, the Arabic Concept-Based Text Summarization System, Sakhr summarizer, Lakhas Arabic summarizer, and the summarizer of Aramedia.

The Non-Arabic Text Summarization Systems: Copernic Summarizer, Kify Text Summarizer, SweSum, MEAD [2]. The systems are listed and each on is described in details in [2].

In the recent years, many approaches have appeared due to the huge amount of information overloaded on the Web, but there are two dominant approaches of summarization: the extraction approach and the abstraction approach.

Our proposed system depends on different important stages. The previous experiences and approaches were inspiring to build our proposed system. The first stage is the **Normalization**, which plays an important role in summarization process, in addition to, another two essential roles, Firstly, splitting long stream contexts into sentences or phrases to facilitate the processing. Secondly, modifying the spelling errors in the input texts. The second stage is to **disambiguate** the correct parts of speech of the word forms. After that comes the **shallow parsing stage**, how to parse, connect different constituents' boundaries and build the tree of each sentence. Then it comes to the **Summarization** stage (Also called **Deletion** stage) that decides which phrases are going to be remained or to be excluded. Finally, it is the role of **Generation** to convert the summarized trees to a well-formed generated text. Each stage will be described in details in section 3:

In this research, we mainly focus on introducing the effort that have been done in the area of the syntactic analysis of the documents. Our objective is to generate document summaries using the shallow parsing of the sentences document that is an analysis of a sentence, which first identifies constituent parts of sentences (nouns, verbs, adjectives, etc.) and then links them to higher order units that have discrete grammatical meanings. This paper is organized as follows: the following section describes some related works with respect to the use of the syntactic analysis of documents for summarization. Section 3 introduces our idea behind the selection of the shallow parsing of the sentences document that is to be summarized. Section 4 presents the summarization methodology and describes the experiment illustrated by examples, which illustrate the different stages of the analysis. Finally, Section 6 concludes this paper.

2 AUTOMATIC TEXT SUMMARIZATION DIFFERENT APPROACHES

In the recent years, many approaches have appeared due to the huge amount of information overloaded on the Web, but there are two dominant approaches of summarization as discussed before in our previous study [2]:

A. Extraction Approach

This approach makes use of different properties of the text to weight the sentences by using a combination of statistical heuristics or linguistic features. Each sentence in the text is assigned a score using a combination of statistical heuristics. The sentences are sorted in descending order according to their score values and an appropriate number of the highest scored sentences are selected from the text to form the summary according to the summarization ratio. The sentences that have the highest scores are considered very important and included in the summary [2].
Extraction is mainly concerned with judging the importance or the indicative power of each sentence in a given document. All sentences are first rated in terms of their importance, and then a summary is obtained by choosing a number of top scoring sentences [3].

The methods used for determining the weights of the sentences are: Cue method, location method and title method. One of the most shortage of using the extractive approach is that the important and relevant information is usually spread out throughout the document and the extractive techniques are unable to combine all of these unless increasing the size of the summery [3]. In addition, when the sentences are picked up as they are the pronouns often tend to lose their references thus creating a confusion to trace the meaning. If there is a confliction in the information, it may not be presented accurately. In order to overcome these problems abstractive summarization techniques can be used as in [3].

B. Abstraction Approach

The abstraction approach involves simplifying and condensing the text. When text summaries are created manually using the abstraction approach, humans read the text, reinterpret it, and rewrite it [3]. Producing abstracts is not a simple task because central topics have to be identified, topics have to be interpreted, and then the summary is generated. The abstraction approach is much more difficult to be programmed than extraction, therefore extraction is the more commonly used approach in automatic text summarization.

Abstractive summarization includes understanding the main concepts and relevant information of the main text, then expressing that information in short, and clear format. Abstractive summarization techniques be classified into two categories: 1) structured based and 2) semantic based methods. Each method will be classified in details in the following sub section.

1) Structured Based Abstractive Summarization Methods

Structured based approaches determines the most important information through documents by using templates, extraction rules and other structures such as tree, ontology etc. [3].The structured based abstractive category has three different methods are: the first is Rules based method, the second is ontology method and finally, Tree bank method. Each method will be explained in the following sub-subsection.

- Rule Based Method

This method includes identifying the categories of the documents to be summarized and form questions based on these categories such as What happened?, when did it happen?, who got affected ?, what were the consequences? Etc. Then, developing a set of rules base on these questions. verbs and nouns having similar meanings are determined and their positions are correctly identified. The context selection module selects the best candidate amongst these. Generation patterns are then used for the generation of summary sentences [3].

- Ontology Method

In this method, domain ontology is defined by the domain experts. Next phase is document processing phase. Meaningful terms from corpus are produces in this phase. The meaningful terms are classified by the classifier on basis of events of news. Membership degree associated with various events of domain ontology. Membership degree is generated by fuzzy inference [3].

Limitations of this approach are it is time consuming because domain ontology has to be defined by domain experts. Advantage of this approach is it handles uncertain data.

- Tree Based Method

In this approach, the preprocessing is done of similar sentences using shallow parser. After that those sentences mapped to the predicate-argument structure. Different algorithms can be used for selecting the common phrase from the sentences such as Theme algorithm. The phrase conveying the same meaning is selected and also some information added to it and arranged in a particular order. A language generator can be used for making the new summary sentences by combining and arranging the selected common phrase. Use of language generator increases the fluency of the language and reduces the grammatical mistakes. This feature is the main strength of this method [3]. The main problem with this method is that the context of the sentences does not get included while selection of common phrase and it is important part of the sentences even if it is not part of the common phrase.

2) Semantic Based Abstractive Summarization

- Multimodal Semantic Model

Multimodal semantic model captures the concepts and form the relation among these concepts. These selected concepts are expressed in the form of sentences. Multimodal consist of three phases (Semantic Modal, Rated Concepts, sentence Generation).

- Information item based method

In this method, instead of generating abstract from sentences of the input file, it is generated from abstract representation of the input file. The abstract representation is nothing but an information item which is the smallest element of information in a text. The modules of this framework are: Information item retrieval, sentence generation, sentence selection and summary generation [3].

From this method, a short, coherent, information rich and less redundant summary can be formed. In spite of so many advantages, this method has also many limitations. While making grammatical and meaningful sentences, many important information items get rejected. Due to which, linguistic quality of resultant summary gets reduced.

- Semantic Graph Based Methods

The semantic graph approach consists of three phases: The first phase represents input document using rich semantic graph (RSG). In RSG, the verbs and nouns of the input document are represented as graph nodes and the edges correspond to semantic and topological relations between them. The second phase reduces the original graph to a more reduced graph using heuristic rules. The third phase generates an abstractive summery.

The advantage of this method is that it produces less redundant and grammatically correct sentences. The disadvantage of this method is that it is limited to a single document and not multiple documents [3].

3 THE METHODOLOGY OF THE PROPOSED SYSTEM

Our proposed system depends on different important stages. The previous experiences and approaches were inspiring to build our proposed system. The first stage is the **Normalization**, which plays an important role in summarization process, in addition to, another two essential roles, Firstly, splitting long stream contexts into sentences or phrases to facilitate processing. Secondly, modifying the spelling errors in the input texts. The second stage is to **disambiguate** the correct parts of speech of the word forms. After that comes the **shallow parsing stage**, how to parse, connect different constituents' boundaries and build the tree of each sentence. Then it comes to the **Summarization** stage (Also called **Deletion** stage) that decides which phrases are going to be remained or to be excluded. Finally, it is the role of **Generation** to convert the summarized trees to a well-formed generated text. Each stage will be described in details in the following sub-sections:

A. Normalization

Text Normalization has become a common practice in the development of various applications. It helps in saving time, which might be consumed in later stages if not applied. This happens through Preprocessing Phase which is responsible for auto-correcting the raw text and segmenting the Arabic text into sentences [4, 5]. The normalization stage is useful for preparing the input text for later processing by transforming the text to a standard format without dispersed data; later operations are able to work with. Moreover, it plays a significant role in the summarization process in our proposed system. The more the input text is being normalized, the more the parser latterly is ready to parse accurately. Text Normalization stage has three essential steps that will be discussed in the following sub-sub-sections:

1) Extraction: Sometimes the Arabic Texts contains intuitive unnecessary constituents that do not add any addition to the whole meaning, so they are better to be deleted. There is an inventory study was done on a certain data to detect the common constituents that are categorized as unimportant constituents in whatever the surrounding context. This study yielded somehow realistic results when tested on various Arabic texts. The Normalizer extracts those constituents according to the applied rules, which remarked these constituents by punctuations such as comma, semi-colon, parenthesis and full stop. This step helps in the summarization process by excluding such sentences or phrases from the first beginning to get rid of the distraction. These extracted constituents might be words, phrases, or expressions. For example: Parenthetical expressions, such as phrases between brackets or hyphens and Phrases starts with specific Arabic keywords, such as "بالرغم من", "على حد قوله", "بسبب", "مثل", "التي", "ذلك" "وبعد".

2) Modification: Sometimes, the input Arabic texts contains some common spelling mistakes. In this step, the text normalization detects the spelling errors in the text and modify it according to a given rules. The errors can be either a missing character or a wrong character such as:

Missing character can a space or a letter as:

I. عبدالله → عبد الله

II. موسقى → موسيقى

Wrong character can be a wrong written letter such as:

III. الى → إلى

B. Disambiguation

Disambiguation is the process of choosing the right internal diacritization (token) for the word form from the dictionary [4, 5]. It prevents the wrong automatic lexical choices from the dictionary and obtains the right internally diacritized ones. Disambiguation rules decides which nouns appear together, which verbs comes with which nouns, which adjectives describe which nouns, which adverbs characterize which verbs and also which prepositions are correctly used; Therefore disambiguation is the stage of completing the word level which is called Morphology. The morphological analysis of the input text is responsible for analyzing Arabic words and assigning the correct POS and the internal diacritization of words, which is achieved through two processes; tokenization process and disambiguation process. Firstly, tokenization is the process of splitting the natural language input into lexical items; the tokenization process is mainly based on the dictionary, therefore the possibility of ambiguity increases with the increase in the number of entries in the dictionary. Lexical items are tokenized according to longest matched unless the possible longest match is blocked by the developed rules, this stage is explained in details in our previous study in [2]. Secondly, disambiguation is applied over the outcomes of the tokenization process; the disambiguation rules are used to reject the wrong lexical choices and re-obtain the right ones as referred in [4, 5].

As mentioned before disambiguation is concerned with preventing the wrong automatic lexical choices and obtaining the right internally diacritized words. Some linguistic indicators can help in solving the lexical ambiguity, which are morphological, adjacency and linguistic indicators as mentioned in [4, 5]. Morphological indicators: affixation has an important role as the first level of part of speech disambiguation, such as prefixes and suffixes. Adjacency indicators: many qualifiers such as number and gender qualifiers and functional word qualifier could control disambiguating the part of speech. Linguistic indicators: the co-occurrence of specific words with words with specific semantic features is used as an indicator [4, 5].
The disambiguation stage works through several modules; each one has its own important role. To start with, Segmentation or tokenization, which is the first step in disambiguation that segments longest match words, decides the outlines of all parts of speech and rejects some tokens, such as the word "بالمدرسة", this word could be segmented into "بال" verb + "مدرسة" noun or into "ب" preposition + "ال" article + "مدرسة" noun, all these entries exist in the dictionary. The first segmentation "بال" verb + "مدرسة" noun is rejected by the developed rules; because verbs are not adjacent to nouns without a middle space, another segmentation will be detected which is "ب" preposition + "ال" article + "مدرسة" noun, which is the correct one.

Then it comes to Collocation, which is essentially a lexical relation, and not subject to rules but to tendencies [6]. Mitchell (1965) defines collocation as an association of roots or potential lexical meanings rather than actual words; further "a linguistic item or class of items is meaningful not because of inherit properties of its own but because of the contrastive or differential relationships it develops with other items or classes [7]. For example, the sentence "السنة والشيعة", this sentence commonly shows that both words "سنة" sunna + "شيعة" 'shiaa' appear together with a specific meaning. The word "سنة" can be disambiguated as the noun "سَنة" year or as "سُنة" sunna, both tokens have the same part of speech, but the only difference between them is the internal diacritic marks that caused variation in meaning. The collocation rules are designed to express either the high possibility of occurrence or the low possibility of co-occurrence of specific words [4, 5]. Therefore, the word "سَنة" "year" will be rejected in this context, while the word "سُنة" "sunna" will be selected to match the meaning.
Moreover, there are rules, which are responsible for disambiguating prepositions, adjectives, nouns, etc., if they are wrongly tokenized, according to the surrounding context. Finally, the most important role in the disambiguation stage, after accurately completing all the previous modules, is detecting the agreement between nouns and adjectives and the agreement between verbs and nouns [4, 5]. By this final step, the sentence or the phrase has correctly been disambiguated and the input text is ready for being syntactically parsed.

C. Shallow Parsing

Partial or shallow parsing - the task of recovering only a limited amount of syntactic information from natural language sentences - has proved to be a useful technology for written and spoken language domains [8]. For example, within the Verbmobil project, shallow parsers were used to add robustness to a large speech-to-speech translation system [9]. Shallow parsers are also typically used to reduce the search space for full-blown, `deep' parsers [10]. Yet another application of shallow parsing is question-answering on the World Wide Web, where there is a need to efficiently process large quantities of (potentially) ill-formed documents [11, 12]. Moreover, more generally all text mining applications, e.g. in biology [13].

Abney (1991) is credited with being the first to argue for the relevance of shallow parsing, both from the point of view of psycholinguistic evidence and from the point of view of practical applications. His own approach used hand-crafted cascaded Finite State Transducers to get at a shallow parse [14]. Typical modules within a shallow parser architecture include the following [8]:

1) Part-of-Speech Tagging: given a word and its context, decide what the correct morphosyntactic class of that word is (noun, verb, etc.). POS tagging is a well-understood problem in NLP, to which machine learning approaches are routinely applied.

2) Chunking: given the words and their morphosyntactic class, decide which words can be grouped as chunks (noun phrases, verb phrases, complete clauses, etc.)

3) Relation detecting: given the chunks in a sentence, decide which relations they have with the main verb (subject, object, location, etc.).

Shallow parsing is a challenging domain for machine learning research. Note that shallow parsing does not refer to a single technique. Instead, it is better to consider it to refer to a family of related methods, all of which attempt to recover some syntactic information, at the possible expense of ignoring all other such information [8]. Building shallow parsers is therefore a labor-intensive task. Unsurprisingly, shallow parsers are usually automatically built, using techniques originating within the machine learning (or statistical) community [8].

Our Automated Shallow Parser adopts the "**X-bar theory**", the theory of syntactic category formation. The X-bar theory was chosen because of its compatibility with Arabic language. The goal of using X-bar theory will be to supply a systematic description of Standard Arabic sentence formation. There are many advantages of adopting X-bar format in lieu of the phrase structure rules. The latter suffer a severe failing. Their role seems largely redundant as they simply duplicate information included in the lexical entries of the lexical categories [15]. On the other hand, the consistency of X-bar theory towards all the phrasal categories is revealed. The X-bar format permits to bring out what is common to the different types of phrases. Another significant property of X-bar theory is that it throws light on the hierarchical organization of the phrase instead of the linear order of the constituents, which is intuitively felt to be wrong. Furthermore, the X-bar schema can be extended to embrace the constituents of the clause as a whole [15].

The proposed Parser starts to identify different types of constituents to fulfill the **sentence level**. In the beginning, **Adjectives** such as "جميلة" is projected to minimal projection "JB", and then if there are any complements exists, it will be combined within this "JB" to create a bigger "JB". Finally, it will be projected to maximal projection "JP" anyways, but after confirming the presence of a specifier such as the Determiner "ال" "DP", as shown in figure 1 for the phrase "الجميلة":

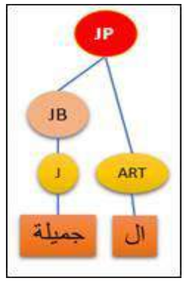


Fig.1. The tree diagram of Adjective phrase in X-bar representation.

For the **nouns**, all nouns is projected to minimal projection "NB", but if it is a **Proper Noun** such as "محمد", it will be projected directly to its maximal projection "NP" if it's not followed by any adjunct related to it as in figure 2. However, if it is a **Common Noun** such as "كتاب", it will be combined with any following *complements* if exists which is necessary to complete the meaning, in order to form "NB", such as "edafa", as in"كتاب محمد" as shown in figure 3, or "subcategorization frame" as in "الاعتماد على". After that nouns combine with *Adjuncts* -which is not syntactically required-, such as the adjective "جميل" in "كتاب جميل" as in figure 4, or "في الفصل" in "الكتاب في الفصل", therefore, "NB" will be formed. If there are no more surrounding complements or adjuncts, "NB" will be projected to maximal projection "NP" if specifiers are found such as the Determiner "ال" as in figure 5.

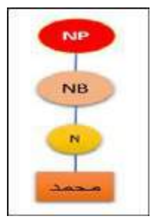


Fig.2. The tree diagram of Proper Noun in X-bar representation.

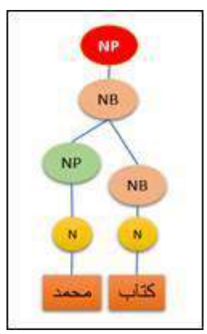


Fig.3. The tree diagram of Noun Phrase "edafa".

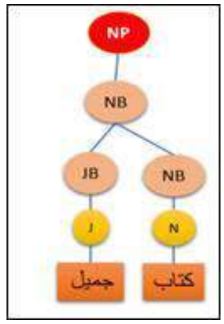


Fig.4. The tree diagram of Noun- Adjective Phrase.

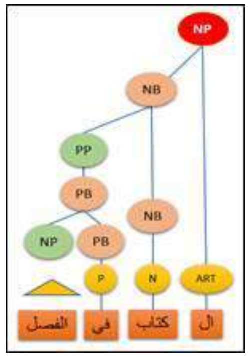


Fig.٥. The tree diagram of a final projected Noun Phrase "الكتاب في الفصل".

Then the parser identifies the **Prepositions** such as "إلى", "على", "في", etc…, the preposition is projected to be a preposition phrase bar "PB". Then it might combined with the complement such as "noun phrase" "المدرسة" in "إلى المدرسة" or a specifier such as "adverbial phrase" " قريبا فوق" in "الكتاب تقريبا فوق المكتب" to reach the maximal projection "PP" as in figure 6.

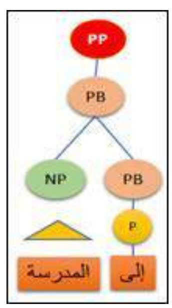


Fig.6. The tree diagram of maximal Prepositional Phrase in X-bar representation.

Moreover, **Adverbs** -either indicates time or manner- might have complements that completes the meaning such as "مقارنةً" which needs a prepositional phrase such as "بالماضي" to be projected to the maximal projection of the adverbial phrase "AP" as in figure 7. On the other hand, there are adverbs that completes the meaning directly on its own such as "اليوم", "أمس", "مبتسمًا"; therefore, they will be projected directly to the maximal projection "AP" as shown in figure 8.

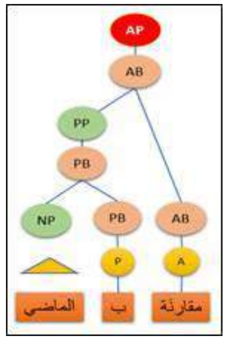


Fig.7. The tree diagram of maximal projection of Adverbial Phrase.

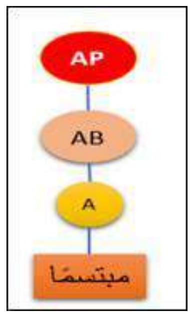


Fig.8. The tree diagram of directly projected Adverbial Phrase.

For the **VERBs**, the verb is projected to be a "VB". Complements of the verb differ according to the verb's transitivity; it may have no complements or have one or more complements. The parser plays an effective role in recognizing the verb's complements. For example, in the phrase "ذهب الولد", the "NP" "الولد" is predicted to be the subject of the verb as presented in figure 9. While the object of the verb can occur in the form of noun phrases, prepositional phrases or verbal phrases, for example, the complement occurred as "NP" "التفاحة" in the phrase "أكل محمد التفاحة" as presented in figure 9, the complement occurred as "PP" "إلى المدرسة" as in "ذهب إلى المدرسة" as presented in figure 10, and finally the complement occurred as "VP" in the phrase "بدأت الأم تصنع الطعام".Verbs can also have adjuncts, for example, the adverbial phrase "بسرعة" in "ذهب بسرعة".

During the analysis of the verb phrase, verb bars "VB" are directly formed, if the noun phrase "NP" precedes the verb the VB will be maximal projected to be "VP" immediately. In addition, there is another extra projection happens, only if the verb is an auxiliary verb, consequently, "VP" is projected to be an inflectional phrase "IP". Figure 11 shows the hierarchy of a fully analyzed verb phrase while connecting the tree diagram of the arguments of the verb phrase.

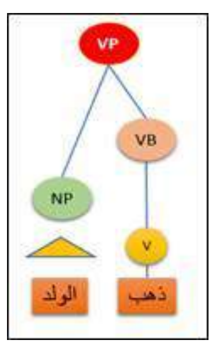


Fig.9. The tree diagram of verb phrase in the presence of the subject.

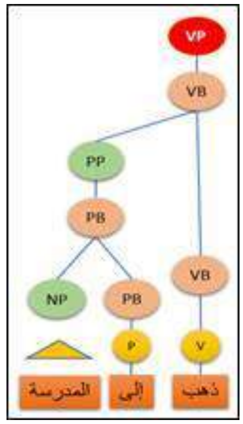


Fig.10. The tree diagram of verb phrase with complement.

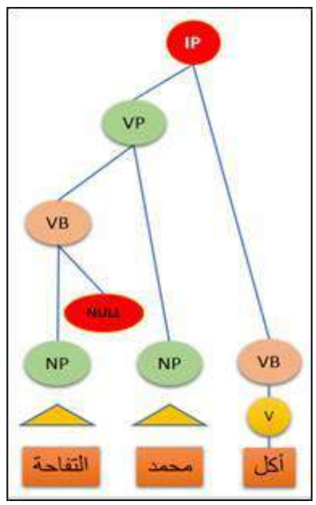


Fig.11. The tree diagram of a complete verbal phrase.

For more illustration, the sentence in (1) is an example to show how the parser works and figure 12 shows the tree diagram for it:

Sentence (1):

"واليوم انضم سيرف إلى وكالة ناسا الفضائية للعمل على مشروع إطلاق شبكة إنترنت في الفضاء الخارجي"

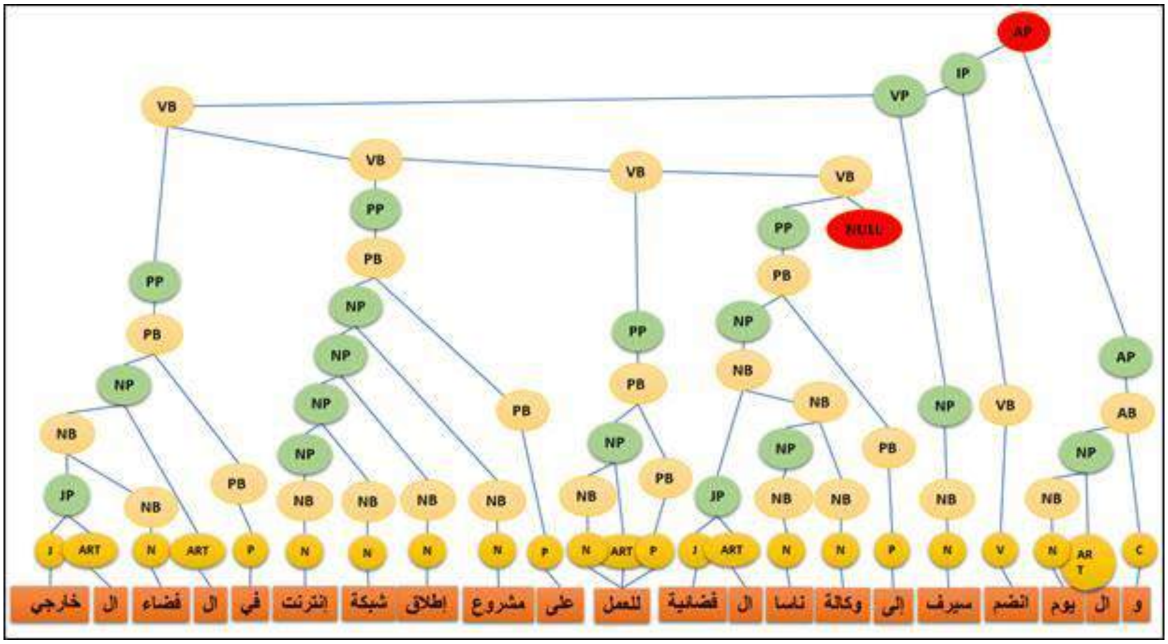


Fig.13. Tree diagram of sentence 1.

D. Deletion

After the parser completes forming the trees' diagram for each sentence in the document. Only one thing left now to start summarizing is the deletion of the insignificant constituents (does not share the meaning of the sentences). Automatic text summarization, aims to reduce the size of a text while preserving its information content. A summarizer is a system that produces a condensed representation of its input's for user consumption [3]. Deletion or omission stage happens by dispensing with certain un-important phrases or clauses in the inputted sentence or text. It is a crucial stage preparing the output to the generation stage.

As mentioned before there is an extraction has done in the normalization stage for some sentences before parsing stage. In addition, after parsing we investigate un-necessary phrases in order to be deleted. There are two types of phrases in the text, the **essential phrases** and the **optional phrases** [2]. The essential phrases are irreplaceable and convey the important information, while the optional phrases have no remarkable meaning and could be omitted such as prepositional phrases, adverb phrases, relative clauses and sometimes complementizer phrase CP.

For the PP, the deletion of it depends on their role. If the PP plays the role as an argument of a verb or an adjective or a noun (object or goal), in this case, the PP is very important and it will not be essential and could not deleted. Otherwise, it can be deleted if it is not add. Sometimes, adverb phrases have no impact in the meaning either indicating time or manner. For the relative clauses, it is better to be omitted because they are treated as the parenthetical expressions that are omitted in the normalization stage.

When Deletion was tested on the above mentioned sentence in (1), only essential phrases were kept during the summarization, while the optional ones were omitted as shown in sentence in (2). Figure (14) shows the omitted phrases and the kept ones for the sentence in (1) during the summarization process.

Sentence (2):

”~~واليوم~~ انضم سيرف إلى وكالة ناسا الفضائية ~~للعمل على مشروع إطلاق شبكة إنترنت في الفضاء الخارجي~~“

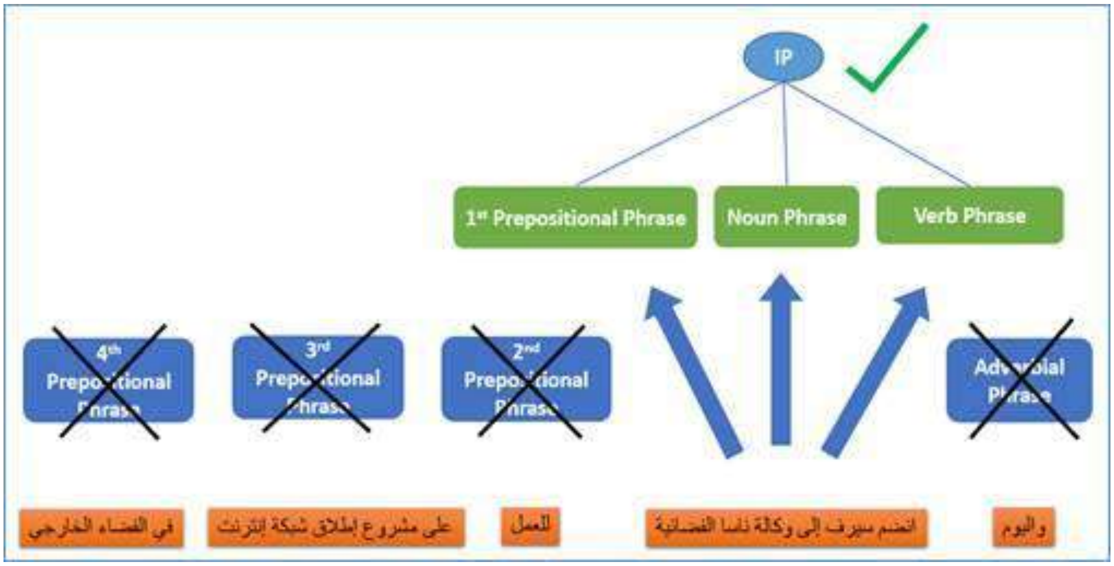


Fig.14. The deletion Process for sentence (1).

E. Generation

In the generation stage, there are a bund of settled rules built for handling the summarized trees after analysis in order to generate and list a well-formed Arabic summarized text. Sentence in (3) becomes the final generated condensed representation of the original sentence in (1). Sentence in (3) is the generated Arabic text of the sentence represented in figure 15:

Sentence (3):

"انضم سيرف إلى وكالة ناسا الفضائية"

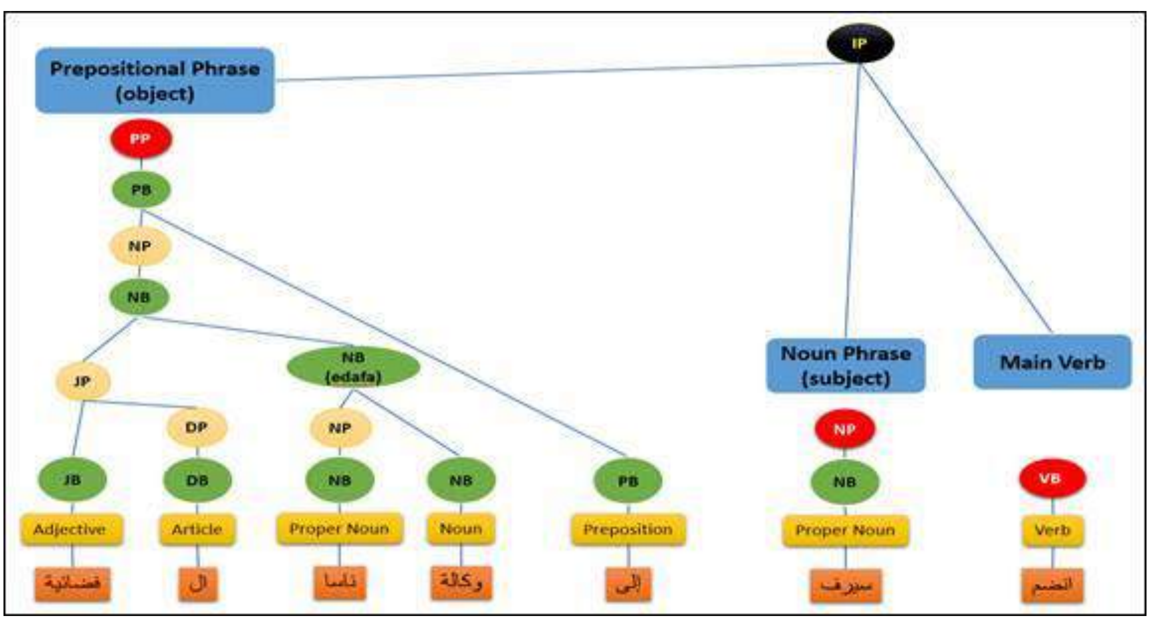


Fig.15. The tree diagram of sentence in (3).

One hundred new contexts were chosen to be automatically summarized. This data were classified into five stages (zero – one – two – three – four – five) according to their complexity. There are many factors that control this complexity, which represent in the type of the text, the sentence length, multiple boundaries, and the number of topic words the sentence may contain. The automatic summarizer in this stage works on the zero, one and two stages according to the established rules, the achieved studies and the results are surprising. The third and the fourth stages are not finished yet because both consist of somehow more sophisticated structures. The most advanced stage is the fifth stage, where all limitations of summarization such as the extra sentence length and the increasing multiple boundaries are encountered in. The proposed automatic system is still working on these difficult issues in order to handle the further work.

Moreover, an Arabic document with the title “الفضاء” was tested to be summarized as shown in figure 16. When applied to the summarizer, essential and optional constituents were detected. The developed rules were applied to segment and modify constituents, select adequate parts of speech, build trees, identifying boundaries and omitting the phrases that does not share meaning of the sentence. The most important stage in Summarization process is the Deletion as mentioned above.

بالنسبة لكوكب الأرض فإن الفضاء الخارجي هو المنطقة الواقعة على بعد 100 كيلومتر عاموديا على سطح الأرض، والتي
سيكون التنفس فيها صعبا جدا لقلة الأكسجين، حيث يكون الغلاف الجوي للأرض قد اختفى تقريبا . هنالك ثلاثة أقسام أساسية من
العلم تهتم بدراسة الفضاء أو السماء، وهي : علم الفلك، وعلم الفيزياء الفلكية، وعلم الكونيات، ولكل واحد منها مجاله المميز،
والذي يهتم بجانب معين من الكون . في هذا المقال سنتحدث عن بعض المعلومات العامة جدا عن الفضاء- وخاصة الأجسام
السماوية الأكثر شهرة - وبعض الأمور الغامضة في هذا الكون. النجوم هي كرات غازية مضيئة، وتتكون بشكل أساسي من
الهيدروجين والهيليوم) أخف عنصرين في الجدول الدوري(، وتحدث فيها اندماجات نووية تولد طاقة وعناصر جديدة مثل
الكربون، والأكسجين، والمغنيسيوم، ... ، وحتى الحديد، لكن الاندماجات النووية الحاصلة داخل النجوم لا تمتلك الطاقة الكافية
لدمج عنصر الحديد وتشكيل عناصر أثقل منها ، فالحديد هو أثقل عنصر يتم تشكيله داخل النجوم، لكن العناصر الأخرى الأثقل
منه يتم تشكيلها في انفجار المستعر الأعظم) بالإنجليزية(Super Nova : ،كالذهب مثلاً ، لذلك لو كان في يدكِ خاتمٌ من الذهب
فيجب أن تعرفي قيمة هذا الخاتم الحقيقية حيث تمّت صناعته في واحدٍ من أعظم مصانع الكون . ومن الأمثلةعلى النجوم وهو
أقرب نجم لنا الشمس . و تتكون الشمس من نفس العناصر الموجودة في الأرض ولكن بنسب مختلفة ؛ حيث إننا نعرف ذلك عن
طريق تحليلنا لطيف الشمس ، الأمر الذي أكدلنا أن الشمس، والأرض، وجميع كواكب المجموعة الشمسية قد تشكلوا من نفس
الغيمة السديمية، ونسب المواد الموجودة في الشمس هي كالآتي :71 %هيدروجين، 27 %هيليوم، 2 %عناصر أثقل.

Fig.16. Original Document.

The summarization of the document in figure (16) will be achieved after going through different stages: the deletion in normalization stage according to the surrounding punctuations as follows:

- Relative Clauses

The clause “والتي سيكون التنفس فيها صعب جدا لقلة الأكسجين” reveals an extra un-important meaning.

- Key-worded expressions

There are constituents that are remarked by some key words that are considered in the settled bund of applied rules; those constituents are deleted whenever they appear in the context. Phrases such as: “ وهي: علم الفلك، وعلم الفيزياء الفلكية، وعلم الكونيات، ولكل واحد منها مجاله المميز، والذي يهتم بجانب معين من الكون”, “ حيث إننا نعرف ذلك عن طريق تحليلنا لطيف الشمس”, “ لكن الاندماجات النووية الحاصلة داخل النجوم لا تمتلك الطاقة الكافية لدمج عنصر الحديد وتشكيل عناصر أثقل منها”, “ في هذا المقال سنتحدث عن بعض المعلومات العامة جدا عن الفضاء -وخاصة الأجسام السماوية الأكثر شهرة- وبعض الأمور الغامضة في هذا الكون” match this case of deletion.

- Parenthesis constituents

Constituents occur between parenthesis are deleted, such as “(أخف عنصرين في الجدول الدوري)”.

Then, after parsing each sentence in the document some deletion decision are made such as the following:

a) Adverbial phrases

The word “عموديا” has no effective meaning, so it will be omitted.

b) Prepositional phrases

The PP “على سطح الأرض”, “من العلم”, “لنا” and “داخل النجوم” are deleted because they are not considered as essential arguments.

Figure 17 shows the final generated output after deletion. In figure 16, the original document counts 240 words, while in figure 17 the automated summarized text counts 91 words.

بالنسبة لكوكب الأرض فإن الفضاء الخارجي هو المنطقة الواقعة على بعد 100 كيلومتر. هنالك ثلاثة أقسام أساسية تهتم بدراسة الفضاء أو السماء. النجوم هي كرات غازية مضيئة، وتتكون بشكل أساسي من الهيدروجين والهيليوم، وتحدث فيها اندماجات نووية تولد طاقة وعناصر جديدة، فالحديد هو أثقل عنصر يتم تشكيله. ومن الأمثلة وهو أقرب نجم الشمس. وتتكون الشمس من نفس العناصر الموجودة في الأرض، الأمر الذي أكد أن الشمس، والأرض، وجميع كواكب المجموعة الشمسية قد تشكلوا من نفس الغيمة السديمية، ونسب المواد الموجودة في الشمس هي كالآتي:
71% هيدروجين، 27% هيليوم، 2% عناصر أثقل.

Fig.17. Automated Summarized Document.

4 CONCLUSION

As there are no summarization systems that directly depend on syntactic analysis which helps in understanding the meaning of the sentences, we are trying to build a summarization system that attempts to understand and identify the essential arguments and the optional ones in the sentences. we have developed a shallow parser as a try to add a new method that helps in automatic summarization, this method is considered promising as it tries to overcome the limitation faced the existing summarization systems that are based on statistical methods. Some examples of the existing summarization systems have been mentioned. Our rule based summarization system is presented.

REFERENCES

[1] J Larocca Neto and A. Freitas and A. A. Kaestner “Automatic Text Summarization using a Machine Learning Approach”, Pontifical Catholic University of Parana (PUCPR) Rua Imaculada Conceicao, 1155.

[2] S Alansary. (2018) “Automatic Arabic Text Summarization: A pilot study”, Bibliotheca Alexandrina, Phonetics and Linguistics Department, Faculty of Arts, Alexandria University - Alexandria, Egypt.

[3] Kasture, N. R., Yargal, N., Singh, N. N., Kulkarni, N., & Mathur, V. (2014). A survey on methods of abstractive text summarization. Int. J. Res. Merg. Sci. Technol, 1(6), 53-57.

[4] S. Alansary, “Alserag: An Automatic Diacritization System for Arabic”. The 2nd International Conference on Advanced Intelligent Systems and Informatics (AISI’16), Cairo, Egypt, 2016.

[5] S. Alansary, (2016, December). Improving Alserag Arabic Diacritization Grammar through Syntactic Analysis. In 16th international conference on language engineering, Cairo, Egypt.2016.

[6] K. H. Nofal, “Collocations in English and Arabic: A comparative study”. English Language and Literature Studies, Department of English, Language Centre Philadelphia University, Jordan, 2012.

[7] Mitchell, T.F., 1965. Linguistic going–on: Collocations and other lexical matters arising on the syntagamatic record. In J. R. Firth, Eds. Principles of firthian linguistics. London: Longman.

[8] J hammerton, M Osborne, S Armstrong and W Daelemans “Introduction to Special Issue on Machine Learning Approaches to Shallow Parsing, journal of Machine Learning Research 2 (2002) 551-558.

[9] Wolfgang Wahlster, editor. Verbmobil: Foundations of Speech-to Speech Translation. Springer, 2000.

[10] Michael John Collins. A new statistical parser based on bigram lexical dependencies. In 34th Annual Meeting of the Association for Computational Linguistics. University of California, Santa Cruz, California, USA, June 1996.

[11] Sabine Buchholz and Walter Daelemans. Complex Answers: A Case Study using a WWW Question Answering System. Natural Language Engineering, 2001.

[12] R. Srihari and W. Li. Information extraction supported question answering. In Proceedings of TREC 8, 1999.

[13] T. Sekimizu, H. Park, and J. Tsujii. Identifying the interaction between genes and gene products based on frequently seen verbs in medline abstracts, 1998.

[14] S. Abney. Parsing by chunks. In Principle-Based Parsing, pages 257{278. Kluwer Academic Publishers, Dordrecht, 1991.

[15] Dalarna “X-bar Theory and Standard Arabic” Fall Term, 2006\7 Said Tamadla.

BIOGRAPHY

Dr. Sameh Alansary: *Director of Arabic Computational Linguistic Center at Bibliotheca Alexandrina, Alexandria, Egypt.*

He is professor of computational linguistics in the Department of Phonetics and Linguistics and the head of Phonetics and Linguistics Department, Faculty of Arts, Alexandria University. He obtained his MA in Building Arabic Lexical Databases in 1996, and his PhD from Nijmegen University, the Netherlands in building a formal grammar for parsing Arabic structures in 2002. His main areas of interest are concerned with corpus work, morphological analysis and generation, and building formal grammars.

He is also the head of Arabic Computational Linguistics Center in Bibliotheca Alexandrina. He is supervising and managing the Universal Networking Language project in Library of Alexandria since 1-6-2005 till now.

Dr. Alansary is the co-founder of the Arabic Language Technology Center (ALTEC), an NGO aims at providing Arabic Language resources and building a road map for Arabic Language Technology in Egypt and in the Middle East. He has many scientific works in Arabic Natural Language Processing published in international conferences and periodicals, and a member in many scientific organizations: (1) Egyptian Society of Language Engineering, Cairo, (2) Arabic Linguistic Society - USA, (3) Association of Computational Linguistics - USA – Europe, (4) Universal Networking Language foundation, United Nations, Geneva, Switzerland.

TRANSLATED ABSTRACT

التحليل النحوي السطحي من أجل التلخيص الآلي للنصوص العربية

سامح الأنصاري

مكتبة الإسكندرية، الشاطبي، الإسكندرية، مصر
قسم الصوتيات واللسانيات، كلية الآداب، جامعة الإسكندرية، الشاطبي، الإسكندرية، مصر
sameh.alansary@bibalex.org

ملخص:

في هذه الورقة، نقترح طريقة تلخيص من وجهة نظر التحليل النحوي السطحي ويتضمن نظام التلخيص مراحل مختلفة: مرحلة استخراج الجمل، مرحلة التحليل النحوي، ومرحلة التوليد .وأخيرا، ستناقش عملية تلخيص التقييم.

الكلمات المفتاحية: التلخيص الآلي (TS)، التحليل الآلي للنص، أجزاء الكلمات، التحليل السطحي والتلخيص التجريدي.

L'Application du Formalisme des Fonctions Lexicales sur la Langue Arabe

Racha MOHAMMAD SALEM
Département de Langue et de Littérature Françaises, Faculté des Lettres, Université d'Alexandrie
salem.racha@live.com

Résumé : ***Les recherches qui traitent de la formalisation de la langue arabe ont commencé dès les années quatre-vingt-dix du siècle précédent. Les linguistes arabes modernes cherchent à créer un système formel rigoureux qui soit inspiré des systèmes de formalisation des langues européennes et qui soit apte à gérer la langue arabe avec toutes ses richesses. Au cours des pages suivantes, nous proposerons une méthode de formalisation de l'arabe selon le système des fonctions lexicales (système d'encodage de la théorie Sens-Texte). Ce dernier sera appliqué sur trois phénomènes linguistiques : le cliché, l'hyperbole et le diminutif.***

Mots-clés : ***Fonctions lexicales, linguistique arabe, mathématisation de langue, théorie Sens-Texte, traitement automatique de langue naturelle.***

1 INTRODUCTION

Le mouvement de mathématisation des langues naturelles commençait dans les années trente avec la tendance de rendre la machine capable de gérer les langues humaines (1). « Mathématiser » consiste à représenter les règles linguistiques sous forme de règles mathématiques. Mais la question à laquelle les créateurs des systèmes de mathématisation devraient penser est la suivante : une formule mathématique est-elle censée refléter le signifiant du signe linguistique ou son signifié ? Autrement dit, est-ce qu'il s'agit de mathématiser la forme ou le fond ? Même si la réponse paraît claire et indiscutable, la réalité est que la plupart des langages de formalisation n'ont pas réussi à exprimer le contenu linguistique des signes dont une langue naturelle se compose. La preuve est que jusqu'à nos jours les applications du TALN (traduction automatique, analyse automatique de texte, etc.) ne sont pas très performantes. De nouveaux défis qui sont intraitables par les systèmes d'encodage surgissent continuellement.

En comparaison avec le mouvement de mathématisation des langues européennes, la mathématisation de la langue arabe s'est attardée vu la complexité de sa linguistique. Dans le présent article, nous présenterons un formalisme qui était appliqué aux langues anglaise et française et dont les créateurs ont pu démontrer sa capacité de déchiffrer les secrets des langues humaines. Il s'agit du formalisme des fonctions lexicales (FL) représentant le système d'encodage de la théorie Sens-Texte (TST). Son atout est qu'il garantit l'équilibre entre les deux niveaux : sémantique et formel. Dans la première partie de l'article, nous expliquerons en détail le formalisme des FL, puis nous l'appliquerons à trois phénomènes linguistiques arabes (le cliché, l'hyperbole et le diminutif) dans le but de démontrer sa capacité de comprendre les différents aspects de la langue arabe.

Cette recherche s'ajoute aux études qui ont été commencées au sein du laboratoire ATILF (Université de Lorraine, France) (2) en 2014 et qui concernent la création du Réseau Lexical de l'Arabe (RL-ar) à l'image du Réseau Lexical du Français (RL-fr) (3). Il est à noter que les réseaux lexicaux représentent la nouvelle génération des dictionnaires électroniques. La sémantique lexicale constitue le cadre linguistique selon lequel un réseau lexical est créé. En plus de la définition du lexème (**L**), l'article lexicographique d'un RL énumère les relations paradigmatiques et syntagmatiques liant (**L**) aux autres lexèmes ainsi que la représentation formelle de chacune de ces relations.

2 LA MATHÉMATISATION DES LANGUES NATURELLES

A. Deux Tendances Majeures

La mathématisation des langues naturelles a commencé avec le besoin de faire comprendre à la machine la langue humaine pour qu'elle puisse assumer les tâches exécutées par l'homme – surtout la traduction automatique – dans un laps de temps court et avec des prix réduits. Effectivement, il existe deux tendances ou deux méthodes de représentation formelle des langues humaines: la logico-mathématisation et l'automatisation-mathématique. La première qui est apparue dans les années trente se penche plutôt vers la mathématique que la linguistique. Cependant, la deuxième qui est surgie pendant les années cinquante afin de combler les lacunes de la première tendance, conserve l'équilibre entre les mathématiques et la linguistique.

En général, le processus d'encodage ne consiste pas seulement à la transformation des lexèmes en symboles ou en équations mathématiques, car cette méthode prive l'unité lexicale d'une partie importante de son sens : A l'encontre de la conception saussurienne, les théories contemporaines définissent le signe linguistique par le rapport liant entre le signifiant, le signifié et le syntactique [(4)]. Ce triplet souligne que l'accès au sens complet d'un lexème s'effectue sur deux niveaux : A partir de son signifié ou son sens propre et de son syntactique ou son sens contextuel [(5)]. Ce qui veut dire qu'un système de mathématisation fiable doit prendre en considération la combinatoire lexico-syntaxique du lexème afin de surmonter les problèmes linguistiques résultant de la pluralité de sens, à titre d'exemple. [Racha SALEM, 2017]

C'est pour cette raison que l'automatisation-mathématique paraît la plus logique pour les linguistes contemporains. Cette dernière prend son élan avec Noam Chomsky – le père de la grammaire générative – qui souligne que la formalisation d'une langue naturelle doit passer par l'analyse linguistique. Car il s'agit d'une formalisation de sens plutôt que de forme.

B. Une Grammaire Particulière

Pour aboutir à ses fins, Chomsky fonde la grammaire générative. A travers un nombre de règles dites « de réécriture », il analyse la structure syntaxique de la phrase pour dégager son sens profond, puis il formalise cette structure. La phrase « la fille mange les bonbons », par exemple, est représentée en :

1) S → NP – VP
2) NP → Det – N
3) VP → V – NP
4) NP → Det – N

La première ligne représente la structure générale de la phrase : le (NP) est le groupe nominal et le (VP) est le groupe verbal. La deuxième ligne détaille la structure syntaxique du groupe nominal alors que la troisième détaille celle du groupe verbal. La dernière ligne représente la structure syntaxique du groupe nominal qui est créé au sein du groupe verbal.

Donc, Chomsky lance la première tentative de formalisation des liens lexico-syntaxiques unissant les lexèmes à l'intérieur d'une phrase. Mais la pratique a montré que l'analyse syntaxique est insuffisante surtout en cas d'ambiguïté. C'est ce qui a poussé les linguistes à se recourir à la sémantique et à la déterminer comme étape introductive à l'étude syntaxique de la phrase. Le mathématicien et le linguiste russe Igor Mel'čuk et ses collègues ont travaillé sur l'importance de la sémantique dans le déchiffrement du sens contextuel.

3 UN MODÈLE D'ANALYSE LINGUISTIQUE MULTISTRATAL

Comme nous l'avons souligné au début, la TST diffère des autres théories de la linguistique informatique parce qu'elle garde l'équilibre entre l'étude linguistique de la langue naturelle et sa formalisation. A l'encontre de Chomsky, Mel'čuk précise qu'un système d'encodage fiable prend en considération les liens sémantico-syntaxiques liant le lexème aux différentes composantes de son réseau sémantique. C'est pourquoi le point de départ de la TST est l'analyse sémantique. Celle-ci est secondée par l'analyse syntaxique, puis morphologique et enfin phonologique du lexème. A l'issu de ce modèle, un aperçu complet de chaque lexème se tisse où toutes ses nuances de sens et ses dérivations syntaxiques seront déterminées. Par conséquent, le lexème ne sera pas codé par un seul symbole, mais chaque constituant de cet aperçu en procura un.

4 UN FORMALISME RIGOUREUX

Pour couvrir toutes les nuances de sens et les dérivations syntaxiques, le formalisme de la TST étudie les liens paradigmatiques [6] et syntagmatiques [7] qui font connecter les unités lexicales d'une langue naturelle. Les fondateurs de la théorie commencent leur projet par l'élaboration d'un inventaire de règles linguistiques qu'ils appellent « le système de paraphrasage » et qui groupe toutes les relations sémantiques et syntaxiques universelles. A travers une quarantaine de règles, le système détermine pour chaque lexème tous les équivalents ainsi que les dérivés qui pourraient le remplacer dans les différents contextes tout en conservant le sens principal. Ces règles qui sont formalisées par la suite sous forme de fonction mathématique : [**F** (x) = y] où (x) représente l'argument de la fonction et (y) est sa valeur.

Le formalisme des FL est réparti en deux rubriques principales dont la première groupe les FL paradigmatiques telles que la synonymie, l'antonymie, la conversion, la métaphore, la nominalisation, la verbalisation, le nom typique de l'actant et le nom du point d'origine.

Syn (أنار) = أضاء

Anti (قبِل) = رفض

Conv (باع ل) = اشترى من

Figur (دخان) = ~ ساتر من

S_0 (سافر) = سفر

V_0 (حديث) = تحدث

S_i (تحدث) = متحدث

Germ (غضب) = شرارة ال~

La deuxième rubrique est celle des fonctions syntagmatiques, nous citons à titre d'exemple : l'intensification, le laudatif et les verbes supports.

Magn (صورة) = ضخمة

Bon (اختيار) = موفق

Oper (أمر) = أعطى

Real (طائرة) = قاد

Parfois, la même FL est divisée en sous-catégories afin de gérer les micros nuances de sens au sein de la même relation linguistique. Prenons l'exemple de la synonymie qui est divisée en synonymie exacte (الترادف الكامل) et approximative (شبه الترادف).

Syn (حزن) ≡ غم (synonymie exacte)

Syn (أشرق) ≅ سطع (synonymie approximative)

A son tour, la synonymie approximative comporte trois sous-catégories comme les exemples suivants le montrent :

F (x) = y

- Synonyme moins spécifique : lorsque le sens de (y) est inclus dans (x)

Syn ⊏ (ديك) = طائر

- Synonyme plus spécifique : lorsque le sens de (y) inclut celui de (x)

Syn ⊃ (قتل) = اغتيال

- Synonyme à intersection : lorsque le sens de (y) et celui de (x) partagent certains traits lexicaux.

Syn ∩ (يلعب) = يمزح

Le formalisme des FL offre d'autres options pour couvrir les unités lexicales ayant un sens composé : Premièrement, la fonction lexicale composée. Elle comporte deux FL qui se combinent facilement sur le plan syntaxique, telles les fonctions **(Anti)** qui dénote « l'opposition » et **(Magn)** qui dénote « l'intensification ». La FL composée **(AntiMagn)** dénote « la diminution / l'atténuation ».

AntiMagn (حرارة) = منخفضة

Deuxièmement, la configuration de fonctions lexicales. Il s'agit d'une unité lexicale qui se compose de plusieurs sémantèmes dont chacun est modélisé par une FL. Ce groupe de FL est réuni dans une configuration. Prenons l'exemple de la collocation : « انفجر بالبكاء » ou « fondre en larmes ». Le verbe « انفجر » dénote à la fois « بدأ في البكاء » (commencer à) et « بكى كثيرا » (pleurer beaucoup). La collocation est donc modélisée par une chaîne de FL :

[**IncepReal** $_1$ (بكى) + **Magn** (بكى)] [8]

Le premier sens est formalisé par une FL verbale composée et le deuxième par une FL d'intensification. La fonction **(Incep)** exprime le sens de « بدأ / commencer » et la fonction **(Real)** exprime le sens de « أنجز / produire / exécuter ». Ainsi, **(IncepReal** $_1$**)** modélise le sens « بدأ في البكاء / commencer à pleurer ».

Troisièmement, la valeur fusionnée. C'est le cas d'un lexème qui exprime le même sens que celui résultant de la combinatoire de deux lexèmes. Prenons l'exemple du lexème « غفا » qui exprime « être au début du sommeil », c'est un sommeil léger. Donc, « غفا » reflète le sens exprimé par la collocation « بدأ في النوم / commencer à dormir ». La valeur fusionnée est marquée par le symbole « // ».

Incep (نام) = // غفا

Reste à mentionner que lors de la création du système des fonctions lexicales, Igor Mel'čuk et ses collègues l'ont voulu un système universel qui s'adapte à toutes les langues humaines. C'est pour cette raison qu'ils ont distingué les FL standard des FL non-standard. Les premières modélisent les relations linguistiques communes à toutes les langues alors que les deuxièmes formalisent les relations spéciales de chaque langue. Les FL standard sont invariables tandis que les FL non-standard sont variables.

5 TROIS PHÉNOMÈNES LINGUISTIQUES FACE AUX FONCTIONS LEXICALES

Dans la dernière partie du présent article, nous appliquons le formalisme des FL sur trois phénomènes linguistiques arabes. Ces derniers ont été sélectionnés précisément car ils diffèrent de leurs équivalents en langue française.

A. Les Clichés Linguistiques (التعبيرات المسكوكة)

Un cliché linguistique est un syntagme non libre de type pragmatique qui est dit dans une situation communicative déterminée. Dans les dictionnaires traditionnels, le cliché est cité sous l'entrée qui représente une de ses composantes (souvent celle portant le contenu sémantique principal du cliché). Examinons cet exemple : le cliché « بالرفاء و البنين » est cité sous l'entrée de la racine [9] « رفأ » qui exprime le sens « السكينة »

(10). Ce cliché se dit dans le contexte du mariage pour féliciter les nouveaux mariés et leur souhaiter une vie tranquille en espérant qu'ils auront d'enfants très bientôt.

Dans un réseau lexical, le même cliché est représenté de deux manières différentes. Il fait partie intégrante des champs sémantiques de ses composantes. Etant donné que chaque composante exprime une partie du sens du cliché, donc, il doit apparaître dans leurs champs sémantiques. De l'autre côté, en général, le sens d'un cliché est exprimé par un autre lexème, alors celui-là est mentionné dans le champ sémantique de celui-ci comme étant son synonyme. Reprenons le cliché « بالرفاء و البنين », il se compose de deux lexèmes : « رفاء » et « بنين ». De même, il est le synonyme du lexème « مبروك / félicitations». Ce qui veut dire que linguistiquement il sera mentionné trois fois. Formellement, il sera mentionné une seule fois sous « مبروك » qui est son synonyme.

Syn « مبروك » = ⌜بالرفاء و البنين⌝

Un autre cliché qui est fréquemment employé dans la presse sportive : « الساحرة المستديرة » ou « la charmante ronde » selon la traduction littérale. C'est une métaphore qui désigne l'amour excessif des gens pour le football. Le sens de ce cliché est composé vu qu'il est constitué du lexème « ساحرة » exprimant l'admiration de beaucoup de gens pour le football et le lexème « مستديرة » désignant la forme ronde du ballon. Dans le réseau lexical, le cliché
« الساحرة المستديرة » sera-t-il lié aux champs sémantiques des deux lexèmes dont il se compose ? En fait, ce cliché diffère du précédent, c'est un cas spécial puisqu'il reflète un sens figuré et non un sens propre. C'est alors qu'il sera plutôt mentionné dans le champ sémantique du lexème « football ». La FL qui est utilisée en cas des sens figurés est **(Figur)**.

Figur (كرة القدم) = الساحرة المستديرة

B. L'Hyperbole (صيغة المبالغة)

Parmi les formules d'intensification arabes, nous avons sélectionné celles qui sont utilisées pour caractériser les humains. La linguistique arabe précise cinq schèmes de ce type d'hyperbole :

مثال	الوزن
فتّاح / سفّاح	فعّال
مقدام	مفعال
شكور	فعول
سميع / عليم	فعيل
حذِر	فَعِل

Chaque schème exprime le fait de répéter une action plusieurs fois. Ci-dessous quelques exemples :

- هذا هو قطّاع الطرق.

Le lexème souligné est une exagération dérivée du lexème « قاطع » ou « bandit » en langue française. Il est utilisé pour montrer que ce bandit s'est habitué à voler et à commettre des crimes. Ainsi, « قطّاع » est paraphrasé par
« يقطع الطرق كثيرا ». Selon le formalisme des FL, **(Magn)** est la fonction qui modélise les adjectifs et les adverbes connotant l'intensification.

Magn (قاطع) = قطّاع

Prenons aussi l'exemple de « مقدام » qui est employé pour designer l'homme très courageux :

Magn (مُقدِم) = مقدام

C. Le Diminutif (صيغة التصغير)

Dernièrement, les tournures du diminutif. Elles sont utilisées dans le but de :

- valoriser ou dévaloriser soit une personne, soit une chose, soit une place ;
- marquer la diminution d'une quantité ou d'un espace ;
- désigner la courte durée ;
- désigner le rapprochement.

Parmi les schèmes selon lesquels le diminutif est construit : « فُعَيل », « فُعَيْعِل » et « فُعَيعيل ». Examinons les exemples suivants (11) :

- سويسرا دويلة بارعة في صناعة الساعات.
- هذا قليم سيئ الخط.
- يجري في الوادي نهير يسقي المزارع.
- نذهب إلى المسجد قبيل الصلاة.

Dans la première phrase, le diminutif connote la valorisation. « دويلة » est dérivé de « دولة / pays ». Pourtant, dans la deuxième phrase, le lexème « قليم » qui est dérivé de « قلم / crayon » exprime la dévalorisation. Quant à la troisième phrase, le lexème souligné est le diminutif du lexème « نهر / fleuve ». Il reflète le sens « petit fleuve ». Enfin, le diminutif dans la quatrième phase exprime la petite durée existant entre l'acte d'aller à la mosquée et la prière. « قبيل » est dérivé de « قبل / avant », il est paraphrasé par « juste avant ».

De ce qui précède, il paraît clair que le diminutif représente un phénomène linguistique particulier vu les valeurs sémantiques contradictoires qu'il exprime. Ce qui par conséquent justifie l'importance de l'analyse linguistique qui doit précéder tout processus de formalisation. Car, dans ce cas, ce n'est pas le phénomène qui sera mathématisé – comme l'hyperbole – mais la valeur sémantique exprimée par la forme lexicale. Ce qui veut dire que chaque valeur sémantique sera formalisée par une FL reflétant son sens. La valorisation sera formalisée par la FL **(Pos)** qui exprime l'évaluation positive alors que la dévalorisation sera représentée formellement par la FL composée **(AntiPos)**.

Pos (دولة) = دويلة

AntiPos (قلم) = قليم

En ce qui concerne la valeur diminutive, Anne-Laure Jousse y a accordé, dans sa thèse, la FL **(Dimin)**. [JOUSSE, 2010]

Dimin (نهر) = نهير

Reste la valeur exprimée dans la quatrième phrase, Jousse a parlé de la FL composée **(AntiMagn$_{\text{temps}}$)** qui connote la courte durée par opposition à **(Magn$_{\text{temps}}$)** qui désigne la longue durée.

6 CONCLUSION

Pour conclure, le présent article a comme objectif principal de jeter la lumière sur le formalisme des fonctions lexicales en tant qu'un système de mathématisation rigoureux qui est capable d'assimiler les différents phénomènes linguistiques d'une langue naturelle. Les principales catégories du formalisme des FL ont été explicitées en premier lieu. Puis, des exemples qui cernent les liens linguistiques fondamentaux modélisés par les FL ont été énumérés en deuxième lieu. De même, il était nécessaire de mettre en évidence les atouts du formalisme de la TST : l'analyse sémantique représentant le socle de ce système permet de comprendre les nuances de sens ainsi que le sens composé des unités lexicales.

A travers l'étude de trois phénomènes arabes (le cliché, l'hyperbole et le diminutif), nous avons démontré que les fonctions lexicales sont applicables sur la langue arabe. Mais les études sont encore peu nombreuses, beaucoup de champs ne sont pas encore découverts. Alors, nous espérons que cette brève étude encouragera les chercheurs à se lancer dans ce domaine riche et d'étudier avec profondeur le formalisme des FL.

NOTES

([1]) « Gérer une langue naturelle par la machine » ne signifie pas seulement le fait que la machine comprenne la langue humaine mais aussi qu'elle puisse réagir à l'égard des ordres donnés par l'utilisateur.

([2]) http://www.atilf.fr/

([3]) Le laboratoire ATILF a élaboré le projet RELIEF (REssource Lexicale Informatisée d'Envergure sur le Français). Pour des extraits clarifiants, veuillez visiter le lien suivant : https://spiderlex.atilf.fr/fr/id/35415

([4]) Le terme « syntactique » a été utilisé pour la première fois par Mel'čuk afin de désigner la combinatoire lexico-syntaxique d'un signe linguistique.

([5]) Nous visons par « le sens contextuel » le sens de l'unité lexicale au sein du contexte. Autrement dit, le sens qu'elle acquiert de sa combinatoire avec d'autres unités lexicales.

([6]) Les liens paradigmatiques représentent les relations sémantico-syntaxiques de base d'une langue naturelle.

([7]) Les liens syntagmatiques concernent la constitution des collocations ou des syntagmes semi-figés qu'une unité lexicale peut former avec les autres unités lexicales d'une langue naturelle.

([8]) Le chiffre employé dans la FL **(IncepReal)** dénote le premier actant du verbe. Dans ce cas, c'est le sujet ou la personne qui fond en larmes.

([9]) Nous rappelons que l'entrée dans un dictionnaire arabe est une racine et non un lexème. Ce qui fait que sous la même entrée nous trouvons tous les lexèmes qui sont dérivés de la même racine.

([10]) Selon le dictionnaire arabe *المكنز الكبير*, version numérisée, 2000, p. 459.

([11]) Les exemples sont extraits du site https://www.tunisia-sat.com/forums/threads/1712077

RÉFÉRENCES

[1] A. Clas, I. Mel'čuk et A. Polguère, *Introduction à la lexicologie explicative et combinatoire*. Editions Duculot, 1995.

[2] A. Polguère, *Liste de fonctions lexicales fréquemment utilisées dans le Réseau Lexical du Français (RL-fr)*, 2014.

[3] A.-L. Jousse, *Modèle de structuration des relations lexicales fondé sur le formalisme des fonctions lexicales*, Thèse de doctorat, Université Paris 7, 2010.

[4] I. Mel'čuk et J. Milicévic, *Introduction à la linguistique*, volume 1. Hermann, 2014.

[5] J.-B. Grize, *Langues logico-mathématiques et langues naturelles*, in Revue française de pédagogie, vol.23, 1973. Version électronique in https://www.persee.fr/doc/rfp_0556-7807_1973_num_23_1_1831 , consulté le 12/12/2019.

[6] R. Salem, *Modélisation de la polysémie : approche contrastive arabe – français basée sur la Théorie Sens – Texte*. Thèse de doctorat. Université d'Alexandrie, 2017.

[7] S. Auroux, *Mathématisation de la linguistique et nature du langage*, in Histoire Epistémologie Langage, tome 31, 2009. Version électronique in https://www.persee.fr/doc/hel_0750-8069_2009_num_31_1_3105 , consulté le 12/12/2019.

[8] *أنواع الترادف في اللغة العربية* , منتدى مجمع اللغة العربية على الشبكة العالمية http://www.m-a-arabia.com/vb/showthread.php?t=26627 , consulté le 17/1/2020

[9] L. Moussaoui, *دراسة دلالية تقابلية للتعابير المسكوكة بين العربية و الفرنسية* , https://www.academia.edu/13850490/%D8%AF%D8%B1%D8%A7%D8%B3%D9%80%D9%80%D9%80%D8%A9_%D8%AF%D9%84%D8%A7%D9%84%D9%8A%D8%A9_%D8%AA%D9%82%D8%A7%D8%A8%D9%84%D9%8A%D8%A9_%D9%84%D9%84%D8%AA%D8%B9%D8%A7%D8%A8%D9%8A%D9%80%D9%80%D8%B1_%D8%A7%D9%84%D9%85%D8%B3%D9%83%D9%88%D9%83%D8%A9_%D8%A8%D9%8A%D9%86_%D8%A7%D9%84%D8%B9%D8%B1%D8%A8%D9%8A%D8%A9_%D9%88%D8%A7%D9%84%D9%81%D8%B1%D9%86%D8%B3%D9%8A%D8%A9_A_Semantic_and_comparative_study_of_fixed_expressions_between_Arabic_and_French , consulté le 5/1/2020.

السيرة الذاتية

رشا محمد سالم

مدرس بقسم اللغة الفرنسية، كلية الآداب، جامعة الإسكندرية. تخصص لغويات حاسوبية. حاصلة على درجة الدكتوراه عام ٢٠١٧ و موضوعها "المعالجة الآلية للألفاظ متعددة المعاني : دراسة مقارنة بين اللغتين العربية و الفرنسية". أثناء تواجدها في فرنسا في مهمة علمية عام ٢٠١٦ ، شاركت في العديد من ورش العمل و الدورات التدريبية الخاصة بعلم المعاني الحاسوبية و التي نظمها معمل ATILF التابع لجامعة لورين. تم نشر بحثين لها في مجال الترجمة الآلية.

BIOGRAPHY

Lecturer, Department of French Language, Faculty of Arts, Alexandria University. Computer Linguistics. She holds a PhD in 2017 with the topic "Automated word processing with multiple meanings: a comparative study between Arabic and French". While in France on a scientific mission in 2016, she participated in many workshops and training courses on computer meanings organized by the ATILF laboratory of the University of Lorraine. Two papers have been published in the field of machine translation.

Application of Lexical Functions Formalism on the Arabic Language

Racha MOHAMMAD SALEM
Department of French Language and Literature, Faculty of Letters, University of Alexandria
salem.racha@live.com

Abstract: Research on the formalization of the Arabic language began in the nineties of the previous century. Modern Arab linguists seek to create a rigorous formal system which is inspired by the formalization systems of European languages and which is capable of managing the Arabic language with all its riches. On the following pages, we will propose a method of formalizing Arabic according to the lexical function system (encoding system of the Meaning-Text theory). The latter will be applied to three linguistic phenomena: cliche, hyperbole and diminutive.

Keywords*:* **Lexical functions, Arabic linguistics, language mathematization, Meaning-Text theory, automatic processing natural language.**

تطبيق اللغة الحاسوبية القائمة على الوظائف المعجمية على اللغة العربية

رشا محمد سالم
مدرس بقسم اللغة الفرنسية و آدابها، كلية الآداب، جامعة الإسكندرية
salem.racha@live.com

الملخص:
بدأ البحث في حوسبة اللغة العربية في تسعينيات القرن الماضي. يسعى اللغويون العرب المحدثون إلى إنشاء نظام حاسوبي قوي مستوحى من أنظمة حوسبة اللغات الأوروبية ، يكون قادار على إستيعاب اللغة العربية بكل ثرواتها. في الصفحات التالية ، سنقترح حوسبة اللغة العربية وفقًا لنظام الوظائف المعجمية (نظام الحوسبة الخاص بنظرية المعنى- النص). سيتم تطبيق هذا الأخير على ثلاث ظواهر لغوية عربية : التعبيرات المسكوكة، صيغ المبالغة، صيغ التصغير.

الكلمات المفتاحية : الوظائف المعجمية ، اللغويات العربية ، لغة حوسبة ، نظرية المعنى- النص ، المعالجة الآلية للغة الطبيعية.

Quelles Contraintes pour Traduire la Morphologie et la Syntaxe ?

Asmaa Gaafar Abdel-Rassoul
Faculte de Lettres, Universite de Menoufia

gmasmaa@yahoo.com

***Résumé*—– Cette recherche vise à étudier l'importance de la morphologie et la syntaxe sur La traduction. Nous allons distinguer la morphologie lexicale de la morphologie dérivationnelle ainsi que le micro-syntaxe de la macro-syntaxe. Par conséquent, notre problématique consiste à répondre aux questions suivantes en appliquant la théorie interprétative dépendant de l' ISIT à la Sorbonne : les deux sont-elles nécessaires pour le texte ? Devons-nous les traduire littéralement d'une langue à l'autre ? Alors, quelles sont les répercussions de les traduire littéralement sur le texte ?**

***Keywords :* syntaxe, morphologie, traduire littéralement.**

1 INTRODUCTION

Bien entendu, la tâche de **la morphologie** est basée sur le morphème qui se réalise au niveau de deux échelles : en premier lieu, **le morphème zéro ou la morphologie lexicale** qui s'intéresse à l'étude de la racine et de l'étymologie de l'unité lexicale. En second lieu, **les morphème (s) ou la morphologie dérivationnelle** qui consiste à étudier les changements intervenus aux unités minimales à travers quelques procédés comme la flexion, la dérivation et la composition. Par ailleurs, la syntaxe est caractérisée par deux notions fondamentales : premièrement, **le micro-syntaxe** qui porte sur l'étude des clauses soit séparément, soit dans une phrase simple. Deuxièmement, **le macro-syntaxe** qui a pour objectif d'analyser la période à travers l'enchaînement des clauses dans une phrase complexe en formant implicitement un sens logique, ce qui nous incite à dire qu'il y a une étroite relation entre le macro-syntaxe et la sémantique.

2 LA RATIONALISATION

La rationalisation que nous visons dans notre étude, c'est toute modification au sein du texte sur le plan structural pour conserver un équilibre communicationnel entre le traducteur et le destinataire comme dans le cas de l'orthographe qui est un outil pour une bonne organisation du texte et par conséquent une bonne compréhension.

Selon Podeur :

> **«La rationalisation correspond à la simplification de l'organisation morpho-syntaxique d'un énoncé et à la modification de la ponctuation» [1].**

À cette période-là où l'œuvre a été traduite, l'utilisation de la ponctuation n'a pas été connue. Le décalage temporel joue un rôle très important à ce sujet. Nous ne pouvons pas dire que c'est une lacune de la part des traducteurs. Cela est probablement dû à la nature de l'imprimerie qui n'était pas développée. **Zaki Pacha** [2] est le premier qui a écrit un livre parlant de la ponctuation qui est très proche de celle qui a été utilisée dans les journaux étrangers et il a souligné qu'il était utile de les imiter. **Zaki Pacha** a aussi indiqué, à ce sujet, que les Turcs ont également utilisé la ponctuation dans leurs journaux mobiles. En revanche, nous trouvons que **El-Hamouz** [3] a dit en bref que les anciens grammairiens arabes ont connu cette ponctuation mais par des signes qui diffèrent de ceux que nous utilisons aujourd'hui.

Examinons quelques exemples où le traducteur n'a pas utilisé la ponctuation que dans les plus rares cas :

Ex1 : «Mon pauvre enfant… nous n'avons qu'un parti à prendre. **Tu es robuste, brave, intelligent** ; […]. Le recrutement t'atteindra l'an prochain ; devance le moment, fais-toi soldat, tu pourras du moins choisir ton arme…**»**. SUE (Eugène), *L'Orgueil*, p. 3.

«"قال لهُ" وقد بحثت ونقبت فما وجدت لك سوى سبيلين إما انك تلج أبواب الجندية وتتهياء لمراقيها قبل ان تدهمك بالقرعة وهى تليق بك **لأنك قوى باسل وذكى»**. حضرة ديمترى أفندي خلاط، *عزة النفس*، الأهرام، عدد١١٦١، ١٨٨١.

Évidemment, le traducteur, dans cet exemple, n'a pas pu retransmettre littéralement la ponctuation de la phrase source car la ponctuation arabe suit autrement le découpage de la phrase. Par la suite, le traducteur a pris en conscience la spécificité de la ponctuation de la langue arabe en suivant les normes réceptrices afin d'unifier, dans sa version, entre l'ordre discursif et le rythme thématique exprimés dans le texte de départ.

Ex2 : **«**Ne m'avouez-vous pas vos peines ? Aimez-vous quelqu'un **?** **»**. DUMAS, *Vicomte de Bragelonne II*, p. 107.

«قال لماذا تكتمين عنى أسرارك اتحبين**.**». نجيب الحداد، عود *على بدء*، الأهرام ٤٤٧٥، ١٨٩٢.

Quant à cet exemple, le traducteur a amalgamé les deux phrases françaises en une seule arabe avec un point final. Nous pouvons aussi ajouter, à ce sujet, que la version du traducteur n'a aucun effet négatif sur le sens de la phrase cible vu que cette forme arabe «**اتحبين**» porte en soi la forme d'interrogation selon la chaîne syntaxique arabe. Cette forme «اتحبين» suit la flexion des formes dérivées de l'arabe de [4] «أتفعلين».Enfin de compte, le sens est retransmis clairement.

Ex3 : «Vos revenus se montent à la somme de *trois millions cent vingt mille francs* environ, ce qui vous fait à peu près *huit mille francs* par jour. Rien que cela, **—**, a ajouté le notaire en riant, **—** aussi êtes-vous ***LA PLUS RICHE HÉRITIÈRE DE FRANCE*»**. SUE (Eugène), *L'Orgueil*, p. 52.

«وقال لى ان دخلك يا حضـــرة الفتاة بالغ زهاء ثلاثة ملايين ومائة وعشـــرين الف فرنك بموجب الحســـاب المدقق فيكون دخلك اليومى نحو ثمانية الاف فرنك وبناءً عليهِ انت **اثرى وريثة فى فرنســـا**». حضـــرة ديمترى أفندى خلاط، *عزة النفس*، الاهرام، عدد١١٩٩، ١٢٠١، ١٨٨١ .

Dans l'exemple ci-dessus, l'auteur a voulu focaliser l'attention du lecteur sur le fond de la phrase en ayant recours au phénomène du majuscule tout en amplifiant les lettres. En contrepartie, l'arabe ne possède pas le phénomène du majuscule, mais quand même le sens est retransmis vu que le traducteur, avec une intelligence professionnelle, a opté pour la forme connue par «l'élatif» [5], pour remplir la même fonction de la phrase source en insistant sur la tonalité et la version rythmique afin de piquer la curiosité du lecteur.

Ex4 : «————Oh ! ma mère… ma mère ! ——— murmura Gerald avec un accent d'ineffable reconnaissance, en tombant aux genoux d'Herminie et couvrant ses mains de larmes et de baisers». SUE (Eugène), *L'Orgueil*, p. 136.

«فانكب جرال على اقدام ارمنا وطفق يبوس يديها ويغســـلهما بدموعه ويقول لامه بلســـان الامتنان **-لاعدمتك يا امى»**. حضـــرة ديمترى أفندى خلاط، *عزة النفس*، الاهرام، عدد ١٣٣٦، ١٨٨٢ .

Dans cet exemple, nous pouvons constater qu'il y a une coordination sémantique entre la répétition et l'orthographe. En plus, nous trouvons que la répétition, représentée dans le syntagme nominal «ma mère» et le point d'exclamation (!), reflète un effet très fort. Mais, quand il s'agit de la ponctuation arabe et la ponctuation française, il n'y a aucune similitude, ce qui constitue une pierre d'achoppement d'une part, et ce qui explique le choix de cette locution «**لاعدمتك يا أمى**» de la part du traducteur d'autre part, vu que cette locution bien connue, embrasse le sens et l'effet du texte source car l'énoncé français montre que Gerald a entendu une nouvelle presque impossible.

Avant d'aborder le problème de la **hamza** dans l'exemple ci-dessous, nous pouvons donner quelques observations : en feuilletant ***Al-Ahram*** à cette époque-là, nous avons remarqué que la hamza n'a pas été écrite que dans les plus rares cas. C'était un trait caractéristique de l'époque du roman.

Ex5 : «— Mademoiselle ! Oh ! ne m'appelez pas ainsi, — s'écria mademoiselle de Beaumesnil, — **ne suis-je donc plus** votre Ernestine, l'orpheline à qui vous avez promis votre amitié… parce que vous la croyiez malheureuse ?... ». SUE (Eugène), *L'Orgueil*, p. 121.

«**الست** صديقتكِ ارنسة الست تلك الفتاة اليتيمة التى حنَّ قلبكِ عليها ووعدتِها بمودتك». حضرة ديمترى أفندى خلاط، *عزة النفس*، الاهرام، عدد١٣١١، ١٨٨٢.

Il est à signaler que «**la hamza**» est un alphabet indispensable dans l'écriture de la langue arabe car à travers laquelle, nous pouvons mettre la main sur le sens exact. Quant à l'exemple ci-dessus, le traducteur n'a pas écrit le syntagme arabe «الست» avec la hamza comme c'était la coutume de cette période-là parce qu'il comptait sur le contexte qui déchiffrait, à son tour, le sens voulu.

3 LES TEMPS VERBAUX

De prime abord, nous travaillons sur deux systèmes temporels différents. L'arabe qui appartient aux langues sémitiques et le français qui appartient aux langues indo-européennes. En outre, l'arabe connaît deux formes accomplie et inaccomplie, tandis que le français connaît aussi l'accompli et l'inaccompli mais en ajoutant que ce dernier comporte plusieurs formes contrairement à l'arabe.

Chaïret a indiqué que :

> ***«Le système verbal de l'arabe, comme celui des langues sémitiques d'une manière générale, a la réputation de reposer sur l'opposition aspectuelle accompli-inaccompli. Les distinctions temporelles n'y seraient assurées que par l'environnement contextuel et éventuellement par des éléments auxiliaires»*** **[6].**

Examinons quelques exemples en la matière :

Ex1 : «Que l'on juge de l'étonnement du marquis et de Gerald. Tous deux **arrivaient** pâles… effarés… comme des gens qui accouraient sauver quelqu'un d'un grand danger…**»**. SUE (Eugène), *L'Orgueil*, p. 145.

«فليتصـور القاري مقدار انذهال دى ملفور وجرال اللذين **اقبلا** مصـفرين منزعجين خوفاً على الفتاتين من خطر جسيم». ديمترى أفندى خلاط، *عزة النفس*، الأهرام، عدد ١٣٥٣، ١٨٨٢.

Selon la règle arabe, «l'imparfait» en français se traduit par «**كان+المضارع** ». Par conséquent, la traduction devrait être «**كانوا يقبلون**». Le traducteur a opté pour le substantif «**اقبلا**» sous cette forme pour avoir un niveau de langage soutenu du point de vue poéticité. Par la suite, le traducteur a pu garder la même valeur morpho-syntaxique évoquée par le texte source.

Ex2 : «Je **resterai** toute la journée dans ma chambre**»**. SUE (Eugène), *L'Orgueil*, p. 71.

«ولا اخرج من غرفتي». ديمترى أفندي خلاط، *عزة النفس*، الأهرام، عدد١٢١٧، ١٨٨١.

Quand le traducteur aborde les temps verbaux, il prend en considération leurs contextes qui précisent la traduction pertinente des temps verbaux. Il est à noter que chaque temps fait une partie intégrante de sa langue.

Dans l'exemple ci-dessus, le traducteur a restitué le futur simple par le présent. Il est à signaler que dans ce contexte le présent donne le sens du futur à travers la formule de négation «لا». Le traducteur a choisi un autre temps verbal qui justifie sa compétence langagière et qui véhicule l'effet de sens particulier évoqué dans le texte source.

> **«*La traduction des temps verbaux s'appuie forcément sur des équivalences susceptibles de rendre un effet de sens particulier révélé dans le texte source*»[7].**

Ex3 : «Je **me moquerais** fort de la noblesse en général**»**. SUE (Eugène), *L'Orgueil*, p. 129.

«فانى **اضحك** على الحسب وأهله». ديمترى أفندي خلاط، *عزة النفس*، **الأهرام**، عدد١٣٢٣، ١٨٨٢.

La correspondance morpho-syntaxique totale est presque inaccessible de deux cultures assez distanciées l'une de l'autre comme notre cas. Chaque temps a ses particularités et sa fonction qui se diffèrent d'une langue à l'autre.

L'existence de l'adjectif «fort» a neutralisé la formule du conditionnel et son effet vu que la notion de doute - l'essence du conditionnel - se contredit avec le sens voulu de l'adjectif «**fort**». Il est à signaler que le conditionnel neutre équivaut au présent. Donc, le traducteur a capté cette modulation.

> ***«Dans toutes les langues, le système verbal diffère passablement, et dans un contexte bilingue qui est celui de la traduction, de nombreux problèmes d'équivalence se posent»[8].***

4 LA TRANSPOSITION

La transposition est une sorte de la traduction libre travaillant sur le signifiant. En d'autres termes, c'est une traduction linguistique à travers laquelle le traducteur adopte des changements aux catégories grammaticales. Par conséquent, **Oustinoff** [9] l'appelle «une récatégorisation».

> **Selon Forges et Braun**, la transposition consiste à **«*remplacer une partie de discours par une autre, sans changer le sens du message*»[10].**

Ex1 : «Ernestine **sourit tristement»**. SUE (Eugène), *L'Orgueil*, p.121.

«فابتسمت ارنسة **ابتسامة الحزن»**. ديمترى أفندي خلاط، *عزة النفس*، الأهرام، عدد ١٣١١، ١٨٨٢.

Le traducteur a eu recours à la transposition pour éviter toute littéralité qui pourrait porter atteinte au sens. Il a opté, avec un savoir-faire, pour le complément absolu outre le nom, ce qui nous a donnés un sens qui s'accorde avec le génie de la langue arabe.

Oustinoff a mentionné à ce propos :

> ***«Chaque fois que la traduction "directe" ou "littérale" aboutit à un énoncé équivalent sur le plan linguistique et stylistique, on le maintiendra ; dans le cas inverse, il faudra recourir à la traduction oblique»[11].***

Ex2 : «Désespérés ! ... mais pourquoi cela ? ...[…] ——— s'écria inconsidérément Ernestine […]». SUE (Eugène), *L'Orgueil*, p. 119.

«قالت ارنسة لمَ انت مستمرة على **القنوط** يا ارمنا». ديمترى أفندى خلاط، *عزة النفس*، الأهرام، عدد ١٣٠٩، ١٨٨٢.

Quant à cet exemple, le traducteur a évité le calque syntaxique de la phrase source, qui serait, à titre d'exemple : **«يائسة ! لمَ هذا اليأس».** Le traducteur a sollicité un changement grammatical - **«désespérés»**, un adjectif remplacé en arabe par le syntagme arabe **«القنوط»** - qui convient avec le tissu morpho-syntaxique arabe en reproduisant cette phrase.

D'ailleurs, l'ajout de l'adjectif **«مستمرة»** donne un excès d'éclaircissements au sens voulu afin de souligner la frustration durable d'Herminie.

Oustinoff a aussi ajouté à ce sujet :

> **«La traduction doit donner l'impression que l'original a été écrit directement en français : la visée est "cibliste"»[12].**

5 LE LANGAGE ET LA PENSÉE

En traitant ce point très délicat, nous pouvons mettre l'accent sur le rapport indissociable entre le signifiant et la pensée. À travers leur enchaînement découle deux notions fondamentales : **l'intertextuelle et l'interdiscursive** qui montrent que le traducteur ne traite pas le texte en s'appuyant sur les signes linguistiques seulement, mais à travers une sorte de corrélation entre ces signes et la culture. En d'autres termes, le traducteur essaye dans ce cas, de restituer le texte conformément au contexte social voulu de la part de l'auteur. En somme, la notion du génie de la langue en général fait allusion à ne pas traiter les langues comme des stéréotypes, mais plutôt le traducteur doit bien manipuler l'ensemble **référentiel/inférentiel**.

Voici quelques exemples à cet égard :

Ex1 : «Le roi, […], descendant de cheval au moment où l'on ouvrait la portière du carrosse, il lui avait offert la main. […] En voyant le roi entrer bravement dans **le bois** avec La Vallière, […]**»**. DUMAS, *Vicomte de Bragelonne II***,** p. 18.

«وترجل الملك عن جواده واقبل حتى اخذ بيد لويزا فانزلها من المركبة وســـار بها إلى **احد جوانب الغابة**». نجيب الحداد، *عود على بدء*، الأهرام، ٤٤٦٧، ١٨٩٢.

Le fait de retransmettre l'article défini «**le bois**» en un article indéfini **«احد جوانب الغابة»**, dans **le premier exemple**, montre que le traducteur a sollicité sa pleine visualisation de la situation pour une bonne restitution de la situation. Il est évident que le roi en entrant dans le bois avec La Vallière, ils se sont assis dans une place précise. Dans cette perspective, nous pouvons souligner que cette version a jeté la lumière sur la fonctionnalité suprême de la traduction qui donne la priorité aux éléments suprasegmentaux.

Ex2 : «Herminie, jusqu'alors craintive, accablée releva orgueilleusement **la tête»**. SUE (Eugène), *L'Orgueil*, p. 135.

«فلما سمعت ارمنا عبارة الأميرة المهينة رفعت **راسها** بعزة وتوردت وجناتها». ديمترى أفندي خلاط، *عزة النفس*، الأهرام، عدد ١٣٣٠، ١٨٨٢.

Il en est de même pour **le deuxième exemple**, la traduction du syntagme français **«la tête»** par «**رأسها**» est plus précise car il parle d'une chose particularisante : de sa tête.

« L'arabe préfère les formules concrètes et personnalisées»[13].

Ex3: « Votre médecin ne vous a-t-il pas déclaré devant moi que, sans les moyens héroïques auxquels il venait de recourir, **vous risquiez de perdre votre fils d'une fièvre cérébrale ?** ». SUE (Eugène), *L'Orgueil*, p. 127.

« قال لكِ انهُ مصاب بحمى فى الدماغ كادت ان . . . **(لا سمح الله)** لولا اجراء الوسائط الفعالة لازالتها ». حضرة ديمترى أفندى خلاط، عزة النفس، الاهرام، عدد ١٣٢٠، ١٨٨٢.

Concernant **le troisième exemple** : le traducteur a omis volontairement cet énoncé «**vous risquiez de perdre votre fils d'une fièvre cérébrale**» car il fait appel à l'esprit une mauvaise augure. Le traducteur l'a remplacé par les points de suspension **(…)** et il a prédit le bon augure en disant «**لا سمح الله**». Ce changement textuel est dû à la dissemblance culturelle. Donc, le traducteur, en exploitant les aspects cognitifs du texte de départ, a prêté une attention particulière à l'unité-texte qui donne la priorité à la cohérence des idées d'une part, et qui vise à garder la spécificité de la langue arabe d'autre part.

Ex4 : « - Elle veut cela ?... **Oh** ! la vaillante et noble fille ! - s'écria le marquis, après un moment de surprise ». SUE (Eugène), *L'Orgueil*, p. 123.

ـــ « فصاح الامير بتعجب ـ **عافاك الله** ايتها الفتاة الباسلة يا ارمنا ». حضرة ديمترى أفندى خلاط، عزة النفس، الأهرام، عدد ١٣١٣، ١٨٨٢.

Quant à cet exemple, le traducteur, en restituant «**Oh !**» par «**عافاك الله**», il a pu atteindre la dimension esthétique de la langue arabe. Par la suite, il a voulu garder la force émotive de l'énoncé source.

Jean et Brisset ont mentionné à ce sujet :

> ***«Le traducteur, jusqu'alors invisible, acquiert le statut explicite d'un spécialiste de la communication interculturelle. On exige qu'il soit capable de déterminer les moyens de médiation les plus fonctionnels, c'est-à-dire les mieux adaptés aux objectifs de la communication dans un contexte socioculturel donné»***[14].

6 LE DISCOURS RAPPORTÉ

Nombreuses sont les formes du **discours rapporté** : **le discours direct, le discours indirect, le discours indirect libre et le discours narrativisé**. Il est à signaler que le sens rapporté est leur production commune. **Le discours direct** est le plus fidèle car le discours cité a la primauté sur le discours citant. Par contre, **le discours indirect** est le contraire de la première forme car il est connu sous la forme d'un récit et la voix du narrateur s'impose tout au long du roman au détriment des personnages. Autrement dit, le disours citant prend le pas sur le discours cité. En plus, **le discours indirect libre** fait un amalgame de deux premiers discours. La polyphonie est l'une des caractéristiques de ce discours car le narrateur participe au dialogue ou plus particulièrement au récit avec les personnages. Dernièrement, **le discours narrativisé** dont la tâche principale est la concision.

Examinons quelques exemples en la matière :

Ex1 : «Celui-ci, à peine entré, dit à la portière :

- Dans quelques instants une dame viendra demander mademoiselle Herminie… vous l'introduirez.

- Oui, monsieur, - répondit madame Moufflon en se retirant». SUE (Eugène), *L'Orgueil*, p. 134.

«ولما صــار في وسـط المحل قال للحاجبة سـتقدم بعد سـاعة سـيدة مصـونة وتلتمس مقابلة السـيدة ارمنا فادخليها فأجابت الحاجبة بالإيجاب وانصرفت». ديمترى أفندي خلاط، *عزة النفس*، الأهرام، عدد١٣٢٩، ١٨٨٢.

Ex2 : «À ce moment, le valet de pied, qui avait disparu avec la voiture, revint sur ses pas, avisa à travers la grille les personnages rassemblés sous la tonnelle, s'approcha, et mettant la main à son chapeau :

—— Messieurs, pourriez-vous, s'il vous plaît, me dire si ce jardin dépend de la maison numéro 7 ? ». SUE (Eugène), *L'Orgueil*, p. 8.

«فنزل المجرى بعد ما نقب في عنوان رقعةٍ بيده كان يبحث عن نمرة المحل المقصود ثم دنا من مقام الجلساء ورفع قبعتهُ قائلاً هل هذه الحديقة تابعة لبيت نمرة ٧». ديمترى أفندى خلاط، *عزة النفس*، الأهرام، عدد ١١٦٤، ١٨٨١.

Ex3 : «À la voix du commandant Bernard, madame Barbançon arriva en hâte, s'excusa auprès de son maître, et dit au domestique qui attendait :

– Vous avez une lettre pour moi… mon garçon ? et de quelle part ? ». SUE (Eugène), *L'Orgueil*, p. 8.

«فنادي القبطان المدبرة بالحضــور فاتت مســرعة وخاطبت المجرى». ديمترى أفندي خلاط، *عزة النفس*، الأهرام، عدد ١١٦٤، ١٨٨١.

Ex4 : «Puis, s'interrompant, l'ancienne sage-femme poussa une exclamation, comme si une idée subite lui eût traversé l'esprit, et elle dit à son maître :

- Monsieur…

- Eh bien ! ...

- Voulez-vous venir un instant avec moi dans le jardin ? J'ai à vous parler en secret, dans le plus profond secret». SUE (Eugène), *L'Orgueil*, p. 9.

«(...) ثم التمست منهُ ان يتنحى لتستشيرهُ بامر سرى جال في خاطرها». ديمترى أفندي خلاط، *عزة النفس*، الأهرام، عدد١١٦٤، ١٨٨١.

Bien entendu, chaque traducteur traduit selon sa formation méthodologique et son bagage cognitif qui doit être assez fort et varié. Le traducteur peut certainement emprunter quelques structures semblables mais en conservant le sens logico-sémantique du texte original.

D'ailleurs, le traducteur a eu recours au changement en traduisant le style direct en style indirect. Ce qui nous intéresse dans ce contexte est le sens logique des phrases et la retransmission du vouloir-dire de l'auteur.

Selon Rudy : ***«Traduire un texte, c'est transposer "transférer" un contenu informationnel d'un système linguistique A à un système linguistique B, chacun possédant ses propres règles. Le contenu informationnel, en d'autres termes le sens logico-sémantique, doit rester stable au cours de l'opération pour que la traduction soit la plus fidèle possible»*[15].**

Le traducteur n'était pas dans l'obligation de traduire le style direct par un autre indirect car le style direct est le style le plus préférable au sein de la chaîne syntaxique arabe, mais par rapport à la période pendant laquelle le roman a été publié, le style indirect était le style préférable. En même temps, le traducteur n'a pas illustré l'esprit du texte source en omettant la scène théâtrale existante dans les quatre exemples en question. Donc, il y a un acte de violence concernant la forme car le traducteur a effacé toute trace de pleine saturation avec le texte original.

7 LE CHANGEMENT D'UN ACTE DE PAROLE

L'acte de parole fait une partie intégrante de la pragmatique car le langage avec la pragmatique se transforme en action qui nécessite par suite une réaction de la part d'autrui. D'ailleurs, il y a trois types de l'acte de parole:

les actes locutoire, illocutoire et perlocutoire. Premièrement, **l'acte locutoire** consiste au sens même de l'énoncé. Deuxièmement, **l'acte illocutoire** est le but de la personne qui parle. Dernièrement, **l'acte perlocutoire** est l'influence de la personne sur l'autrui.

Voici quelques exemples à ce propos :

Ex1 : «-Vous m'avez dit tout à l'heure, madame, et très sagement, qu'il ne fallait plaisanter, ni avec la noblesse, ni avec la religion, **n'est-ce pas ?**

- **Oui, monsieur le marquis.**

- Vous avouerez qu'il ne faut pas non plus plaisanter avec le mariage ? **».** SUE (Eugène), *L'Orgueil*, p. 129.

«قد قلتِ لي الآن يا سـيدتي عدم جواز الهزء بالحسـب والدين **وانا أقول لكِ** انهُ لا يجوز الهزء بالزواج**»**. ديمترى أفندي خلاط، *عزة النفس*، الأهرام عدد ١٣٢٣، ١٨٨٢.

Il est à noter que «**n'est-ce pas ? Oui, monsieur le marquis**» est remplacé par «**وانا اقول لكِ**» car à travers cet énoncé, l'auteur n'a pas voulu désigner le doute mais plutôt la confirmation vu que la négation de la négation exprime la confirmation. En plus, le traducteur a eu recours à cette modulation pour mettre l'accent sur l'unité de pensée révélée dans le texte source.

Ex2 : «- Comte, dit-elle, ménagez-moi. Vous voyez que je souffre, **et vous n'avez aucune pitié**». DUMAS, *Vicomte de Bragelonne II*, p. 107.

«وقالت انك تقتلني يا كونت **أفلا تشفق عليَّ**». نجيب الحداد، عود *على بدء*، الأهرام، ٤٤٧٥، ١٨٩٢.

Quant au **deuxième exemple**, le traducteur a modulé la phrase source qui est en négation «**et vous n'avez aucune pitié**» en une phrase interrogative «**أفلا تشفق عليَّ**» car Madame «Beaumesnil» a voulu s'interroger sur la pitié du comte. Le traducteur a mis l'accent, dans ce contexte, sur l'unité de traduction en faisant abstraction aux unités simples. C'est la raison pour laquelle, le traducteur a donné la priorité à la traduction oblique en restituant cet énoncé. Par la suite, le traducteur, par son intelligence professionnelle, a pu prendre en considération l'acte de langage voulu et l'a traduit selon l'intention de l'auteur.

Autrement dit, le traducteur n'a totalement pas changé le texte, mais il s'est contenté de faire un changement sur le plan structural seulement, ce qui lui a aidé à garder la même allure du texte source.

Keightley a montré à ce propos :

> ***«Une traduction doit impérativement transmettre les mêmes informations que l'original mais il faut en même temps veiller au bon fonctionnement pragmatique dans la situation destinée»[16].***

8 LES NOMS DE NOMBRE

Dans la langue arabe, **les noms de nombre** sont différents des noms qui les suivent. Plus précisément, les noms de nombre de (1 jusqu'à 10) ne suivent pas les genres du nom qui les suivent.

> ***«Le nom de nombre ne devrait pas être influencé par le genre du nom compté, […]. Mais suivant la doctrine des grammairiens arabes, il a été nécessaire, pour confirmer la qualité de substantif à ces noms de nombre, de leur donner le genre inverse de celui qu'a le nom compté au sing. ; c'est-à-dire que les noms de nombre de 3 à 10, […], accompagnent les noms comptés féminins et que les noms de nombre prennent la désinence : quand ils sont construits avec des noms masculins»[17].***

Ex1 : «Une vieille ménagerie, […], était depuis **dix ans**, au service du commandant Bernard». SUE (Eugène), *L'Orgueil*, p. 1.

«(...) مع خادمة امينة على امتعتهِ صـادقة الود لهُ كرَّت عليها بخدمتِه **عشرة سنين**». ديمترى أفندى خلاط، عزة *النفس*، الاهرام عدد١٨٨١،١١٦١.

Dans **l'exemple N.1**, le traducteur a littéralement retransmis le nom de nombre «dix» en suivant le genre féminin du nom compté «années». Ce qui contredit les règles des grammairiens arabes. Le traducteur aurait dû le traduire par : **«عشر سنين».**

Ex2 : «Moi, qui ai **sept maisons** sur le pavé de Paris, je n'ai pas seulement de tapis dans mon salon». SUE (Eugène), *L'Orgueil*, p. 61.

«حال كونى املك **سبع منازل** في شوارع باريز وما عندي بساط مفروش في مسكنى». ديمترى أفندى خلاط، عزة *النفس*، الأهرام، عدد ١٢٠٧، ١٨٨١.

Dans cet exemple, le traducteur a suivi la même méthode de l'exemple précédent en traduisant le nom de nombre «sept» en suivant le genre masculin du nom compté «**la maison**» qui est un nom masculin en arabe contrairement au français. Par conséquent, Le traducteur aurait dû le traduire par : «**سبعة منازل**».

Ex3 : «Puis elle compta dix pas de la fenêtre à son lit, et écrivit encore : "**Dix pas**"». DUMAS, *Vicomte de Bragelonne II*, p. 236.

«ثم قاست من النافذة الى سريرها وكتبت "**عشر اقدام**"». نجيب الحداد، عود *على بدء*، الأهرام،٤٤٩٨، ١٨٩٢.

Dans cet exemple, le traducteur n'a pas réussi à retransporter le nom de nombre : le nom de nombre «**dix**» ne suit pas le genre masculin du nom compté «**pas**». Il est à noter que le substantif **«القدم»** - dans la langue arabe - comme une unité de mesure est considérée comme un nom masculin. À l'inverse, lorsque le substantif «**القدم**» signifie - dans la langue arabe - un membre du corps humain, «**القدم**» constitue un nom féminin [18]. Par la suite, Le traducteur aurait dû le traduire par : «**وكتبت عشرة أقدام**».

9 CONCLUSIONS

En guise de conclusion, nous pouvons conclure que toute traduction valable devrait dépasser inévitablement toutes barrières de nature morpho-syntaxique pour mieux manipuler le rapport explicite/implicite pour être à l'abri de toute opacité culturelle.

Références

[1] J. PODEUR, «La Traductologie entre description et évaluation», in *Repères Dorif Traduction*, médiation, interprétation, volet N.1, June 2013, in www.dorif.it/ezine/ezine_printarticle.php?dorif..

[2]زكى باشا (أحمد)، *الترقيم وعلاماته فى اللغة العربية*، قدم له واعتنى بنشره عبد الفتاح أبو غدة، الطبعة الأولى، المطبعة الأميرية، مصر، ١٩١٢، ص ١٦. (نسخة مصورة).

«ورأيت من المفيد استعمال العلامات الإفرنجية، (...) وفوق ذلك قد استخدمها الأتراك فى (...) جرائدهم السيارة (...)». (Traduit par nos soins)

[3]الحموز (عبد الفتاح أحمد)، *فن الترقيم فى العربية أصوله وعلاماته*، دار عمار، عمان، الأردن، ١٩٩١، ص ٢٠.

«وبعدّ فلعلّك تتّفق معى فى أنَّ المصنِّفين والناسخين القدامى لم يتناسوا تلك العلامات والرموز التوضيحيّة التى لابدُ منها فى الكتابات المختلفة، (...)، وليس بخافٍ أنَّ ما مرَّ يُعد دليلاً بيناً وإشارةً ساطعةً إلى أن للعرب علاماتٍ وأماراتٍ كتلك التى تطالعنا فى الكتابات فى عصرنا، على الرغم من أنَّ هناك اختلافاً فى الأشكال والرموز (...)». (Traduit par nos soins).

[4] R. BLACHÈRE, M. GAUDEFROY-DEMOMBYNES, *Grammaire de l'arabe classique*, Maisonneuve et Larose, France, 1975, pp. 49-73.

[5] *Ibid*., p. 97.

[6] M. CHAÏRET, *Linguistique contrastive et traduction, Fonctionnement du système verbal en arabe et en français*, N. spécial, Ophrys, Paris, 1996, p. 7.

[7] L. AVENDAÑO ANGUITA, «Perspective et temps verbaux : problèmes de traduction», in *CLE des langues*, Avril 2010. **In cle**.ens-lyon.fr/.../com.univ.collaboratif.utils.LectureFichiergw?ID... (La dernière consultation était le 15-3-2015).

[8] J. MOESCHLER, C. GRISOT, B. CARTONI, «Jusqu'où les temps verbaux sont-ils procéduraux ?», in *Nouveaux cahiers de linguistique française*, N. 30, 2012, pp. 119-139.

[9] M. OUSTINOFF, *La Traduction*, *Que sais-je ?*, Presses Universitaires de France, 2003, p. 75.

[10] G. FORGES, A. BRAUN, *Didactique des langues, traductologie et communication*, De Bock Université, Paris, 1998, p. 15.

[11] M. OUSTINOFF, *Op.cit.*,, p. 72.

[12] *Loc.cit.*

[13] C. HECHAÏMÉ, *La Traduction par les textes*, 2e édition revue, Dar el-Machreq éditeurs, Beyrouth, p. 69.

[14] M. Y. JEAN, A. BRISSET, «La Notion de culture dans les manuels de traduction : domaines allemand, anglais, coréen et français », in *META*, V. 51, N. 2, Juin 2006, pp. 389-409.

[15] R.LOOCK, «"Parce qu'en plus il faut traduire la syntaxe ?!" : Contraintes et stratégies dans la traduction de la structuration d'un texte», éd. D'Amélio, *Actes du colloque international "la forme comme paradigme du traduire", Mons, CIPS*, 2009, pp. 173-190.

[16] S. KEIGHTLEY, *Changements syntaxiques, modulations et adaptations dans un texte médical*, un mémoire, Linnæus University, 2013, pp. 17-18.

[17] R. BLACHÈRE, M.GAUDEFROY-DEMOMBYNES, *Op.cit.*, p. 368.

[18] Thaqafat.com › إضاءات (La dernière consultation était le 27-3-2015).

«القَدَمُ: ما يَطَأُ بِهِ الإِنْسانُ الأَرْضَ، وَجَمْعُها "أَقْدامٌ" (إِذا قُصِدَ بِها وَحْدةُ القِياسِ المَعْرُوفةُ فَإِنَّها تُذَكَّرُ)».

What Constraints to Translate Morphology and Syntax ?

Asmaa Gafar Abdel-Rasoul

Faculty of Literatures

gmasmaa@yahoo.com

***Abstract*—The aim of this research is to study the importance of morphology and syntax in translation. We will distinguish the lexical morphology from the derivative morphology as well as the micro-syntax of the macro-syntax. Therefore, our problem is to answer the following questions by applying the interpretative theory dependent on ISIT in the Sorbonne: are both necessary for the text ? Are both necessary for the text ? Do we have to translate them literally from one language to another ? So what are the implications of literally translating them into the text ?**

Keywords: syntax, morphology, literally translation.

ما المعايير المُتبعة لترجمة القواعد والتراكيب النحوية ؟

أسماء جعفر عبد الرسول

كلية الآداب- جامعة المنوفية

gmasmaa@yahoo.com

***الملخص* — الهدف من هذا البحث هو دراسة أهمية القواعد والتراكيب النحوية وبناء الجملة في الترجمة. وسنميز المورفولوجيا المعجمية التى تهتم بدراسة أصل الكلمات عن المورفولوجيا الاشتقاقية التى تؤثر على بناء الجملة وكذلك بناء الجملة البسيطة من تلك المركبة. ولذلك، فإن إشكالية هذه الدراسة تكمن فى الإجابة على الأسئلة التالية من خلال النظرية التفسيرية التابعة لمدرسة المترجمين العليا بالسوربون : هل كلاهما ضروري للنص ؟ هل يجب علينا ترجمتهما حرفيا من لغة إلى أخرى ؟ ما هي إذن الآثار المترتبة على ترجمتهما حرفيا على النص ؟**

الكلمات المفتاحية : القواعد، التراكيب النحوية، الترجمة الحرفية.

السيرة الذاتية :

أسماء جعفر عبد الرسول

حاصلة على ليسانس آداب وتربية ثم على ليسانس آداب. حصلت على الماجستير فى الترجمة وتعمل حالياً مدرس مساعد بكلية الآداب، جامعة المنوفية. لقد شاركت بالبحث المعنون «الاختلاف الثقافى» فى المؤتمر الدولى للترجمة الذى أقيم فى المجلس الأعلى للثقافة والمجلس القومى للترجمة فى نوفمبر ٢٠١٦. ومن جانب آخر، شاركت بالبحث المعنون «حركة الترجمة وتأثيرها على الأدب العربى» فى ملتقى العلاقات الثقافية الفرنسية-المصرية والذى أقيم فى المجلس الأعلى للثقافة يومى ٢١ و٢٢ مايو ٢٠١٧. ولها بحث منشور فى مجلة كلية الآداب، جامعة المنوفية، تحت عنوان «إعادة قراءة الروايات الفرنسية المترجمة إلى العربية فى الصحافة المصرية فى الفترة من ١٨٨١ حتى ١٨٩٣. دراسة فى ترجمة الثقافة». وشاركت فى المؤتمر الدولى الثانى للغات الأوروبية بكلية الآداب، جامعة المنوفية المنعقد فى الفترة من ٣ إلى ٥ ديسمبر ٢٠١٧، بالبحث المعنون «المترجم والوساطة الثقافية». وشاركت أيضاً فى المؤتمر السابع عشر لهندسة اللغة والذى أقيم فى جامعة مصر الدولية يومى السادس والسابع من ديسمبر ٢٠١٧ بالبحث المعنون «دور السياق فى صياغة المعنى فى الترجمة». وشاركت فى المؤتمر الدولى الثانى عن التراث العربى والإسلامى، الذى أقيم بمعهد المخطوطات العربية يومى ٢١ و٢٢ فبراير ٢٠١٨، بالبحث الموسوم «حركة الترجمة فى جريدة الأهرام فى الفترة من ١٨٨١ حتى ١٨٩٣». وشاركت فى مؤتمر «اتجاهات معاصرة فى دراسات المستعربين» والذى أقيم بكلية الآداب، جامعة القاهرة فى الفترة من ٣ إلى ٥ أبريل ٢٠١٨ بالبحث المعنون «التباعد الزمنى والترجمة». وشاركت فى المؤتمر العلمى السابع لكلية الآداب بقنا جامعة جنوب الوادي العربية والدراسات الإنسانية والاجتماعية المنعقد فى الفترة من ١١ إلى ١٣ نوفمبر بالبحث الموسوم«العربية واللغات الأخرى : دراسة فى التعريب والتأصيل والاشتقاق». شاركت أيضا فى ٢٠١٩ فى عدة مؤتمرات تابعة لجامعة قناة السويس. وشاركت أيضاً فى المنتدى السادس لشباب الباحثين المنعقد فى مارس ٢٠١٩. ورشحت لدورة تدريبية لغوية وتربوية من الحكومة الفرنسية ومقرها مدينة فيشى بفرنسا فى أغسطس ٢٠١٩. عضوة فى عدة نشاطات تابعة للجمعية المصرية لأساتذة اللغة الفرنسية، وقد حضرت الكثير من المؤتمرات والندوات واللقاءات على هامش هذه الجمعية. وقد حصلت على شهادة تفيد بإجادتها للمستوى اللغوى *Delf B2* من وزارة التربية والتعليم الفرنسية. وسجلت مؤخراً حلقتين على النيل تى فى *NILE TV* فى البرنامج المعنون «***Bibliotheca***».

Asmaa Gaafar Abdel-Rassoul

She holds a Bachelor of Arts and Education and then a Bachelor of Arts.She holds a master’s degree in translation and is currently an assistant lecturer at the Faculty of Arts, Menoufia University. She participated by the research entitled "Cultural Difference" in the International Conference on Translation held in the Supreme Council of Culture and the National Council for Translation in November 2016.On the other hand, she participated by the research entitled "The movement of translation and its impact on Arabic literature" at the French-Egyptian Cultural Relations Forum, which was held at the Supreme Council of Culture on May 21 and 22, 2017.She has a research published in the Journal of the Faculty of Arts, Menoufia University, entitled "Re-reading French novels translated into Arabic in the Egyptian press in the period from 1881 to 1893. A study in translating culture". She participated in the Second International Conference of European Languages, Faculty of Arts, Menoufia University, held from 3rd to 5th December 2017, with research entitled "Translator and cultural mediation". She also participated in the 17th conference for language engineering, which was held at Misr International University (MIU) on the 6th and7th of December 2017, with research entitled "The role of context in formulating meaning in translation".

She participated in the second international conference on the Arab and Islamic heritage, which was held at the Institute of Arabic Manuscripts on February 21 and 22, 2018, with the research titled “The translation movement in Al-Ahram newspaper from 1881 to 1893. She participated in the conference "Contemporary Trends in Arabist Studies", which was held at the Faculty of Arts, Cairo University from 3 to 5 April 2018 with research entitled "Time Spacing and Translation". And participated in the 7th scientific conference of the Faculty of Arts in Qena, South Valley University, Arab and humanities and social studies held from 11 to 13 November with research titled «Arabic and other languages: a study in Arabization and derivation». In 2019 she also participated in several conferences affiliated with the University of the Suez Canal. She also participated in the sixth forum for young researchers held in March 2019.She was nominated for a linguistic and educational training course from the French government based in Vichy, France in August 2019. She is a member of several activities of the Egyptian Association of French Language professors and has attended many conferences, seminars and meetings on the sidelines of this association. She obtained a certificate of proficiency in the language level Delf B2 from the French Ministry of Education. And recently recorded two episodes on the NILE TV in the program entitled «**Bibliotheca**».

Artificial Intelligence and NLP

Chatbot System Architecture

Moataz Mohammed[*1] and Mostafa M. Aref[**2]

[*]*Computer Science Department, Faculty of computer and Information Sciences, Ain-shams University*

[**]*Computer Science Department, Faculty of Computer and Information Sciences, Ain-shams University*

[1]moataz.mohammed@cis.asu.edu.eg

[2]aref_99@yahoo.com

***Abstract*— The conversational agents is one of the most interesting topics in computer science field in the recent decade. Which can be composite from more than one subject in this field, which you need to apply Natural Language Processing Concepts and some Artificial Intelligence Techniques such as Deep Learning methods to make decision about how should be the response. This paper is dedicated to discuss the system architecture for the conversational agent and explain each component in details.**

***Keywords:* Conversational Agents, Chatbots, System agent, Dialog System, Natural Language Understanding, Natural Language Processing, Deep Learning, System Architecture**.

2 INTRODUCTION

A chatbot (conversational agent (CA), dialogue system) is a computer software that acts as an interface between human users and a software application, using spoken or written natural language as the primary means of communication. Examples for that Apple Siri, Google Assistant and Amazon Alexa. In the recent decade most of companies throughout the world use it as a service to improve the way of communicating with clients in any time and the response is instantly which there is no delay. That provide the client with the comfortableness. So this essential technology service should be improved continuously. In this paper we will discuss the CA system architecture. There are many variant system architectures which the developers can follow them to develop the chatbot(conversational agent) but any one of them must have three significant core components: Natural Language Understanding(NLU) is responsible for understanding the user's utterances meaning and put them in representational format, then The Dialog Manager(DM) which is the most important component about acting as a mediator which receive the representational format from the NLU and process them .After that, send the responses for the Natural Language Generator(NLG) which is the last core component in any Conversational Agent(CA) architecture. It takes the responses from the Dialogue Manager (DM) and check if there are more than one valid response taking the one with highest priority. Finally, produce the response in the final format which may be text or Speech.

Social chatbots' appeal lies not only in their ability to respond to users' diverse requests, but also in being able to establish an emotional connection with users. The latter is done by satisfying users' need for communication. How to make the social chatbot more Human Like about Emotional Quotient (EQ) **[1]**. Also it can make an examination on the influence of its responsiveness and embodiment on the answers. people give in response to sensitive and non-sensitive questions **[2]**. AlI and deep learning can help us build such chatbots that improve the lives of people who have busy schedule to easily keep a check on their health **[3]**. The access to large-scale data and real-world feedback can drive faster progress in research **[4]**.

3 LITERATURE REVIEW

When we made a deep reading to a survey papers, an articles and journals. We found many conversational agents each one with its own architecture. Like Amazon Alexa bot which take the data of the user that may be voice and make an Automatic Speech Recognition ASR about the Amazon ASR service. Then make a processing to the received data about some Amazon Web Services AWS and using Amazon DynamoDB to store the conversations and its state**[4]**. Google Assistant the most successful bot at all.Which receive the recordings from users, then sending these recordings to google's servers which works on making processing. It makes break down the voices into individual sounds then try to match every single sound with the most similar word's pronunciation one which is stored on google database. Then it identifies the required task through some matched key words. And other more agents so we mentioned just two examples. From all these different architectures we will talk in this paper about the most commonly generic one which include three main components Natural Language Understanding, Dialog Manager, and Natural Language Generator.

4 SYSTEM ARCHITECTURE

In this section we will discuss each component in the architecture in details .We will talk about the three core components of any chatbot (Conversational Agent) and its subs, First Component is the Natural Language Understanding(NLU) and its role in the system. After that, we will talk about the subcomponents of the NLU such as Topic Detection, Intent Analysis, and Entity Linking. Second Component is the Dialog Manager (DM) and its subs such as Rule-base, Knowledge-based, Neural Network Reply Generation and the Online Information Retrieval. Finally the Natural Language Generation (NLG) or Reply Generator and it subs like Content Filter and Engagement Ranking. the following figure1 **[5]** can be considered as a simulation for these components.

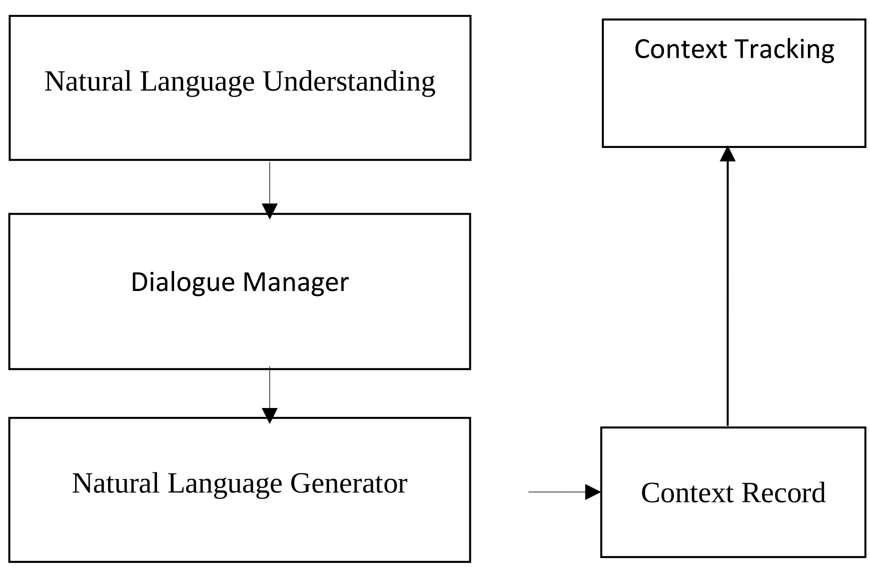


Figure 17: System Architecture of the Conversational Agent [e.g. chatbot]

A. Natural Language Understanding

Natural language understanding (NLU) is the first core component of the conversational agents which is responding about providing a semantic representation for user utterance **[6]** such as an in form of logic or class' intent, extracting the "meaning" of an utterance **[7]**. Parsing is the main task of - an NLU, which takes the string of words and provides a linguistic structure for the utterance. Implementation-dependent is the method which an NLU uses it to parse the input and can utilize context free grammars, pattern matching, or data-driven approaches as we can see in figure2. NLU outputs have to able to be tackled by a dialogue manager **[8]**.

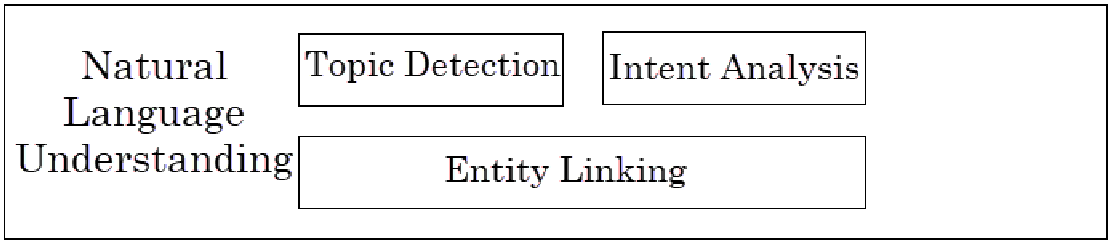


Figure 18: Natural Language Understanding [First Core Component]

1) *Topic Detection: from figure2. Topic Detection is One of the most important steps in developing the conversational agents NLU stage is to identify the Topic of the text (Conversation) for natural free-form interactions. Accurate tracking of the conversation topic can be a valuable signal for a system for dialog generation* **[9]**. *Examples on techniques can be used for topic classification are DANs and ADANs.*
2) *Intent Analysis: An intent is a group of utterances with similar meaning. Notice the following example of these two sentences: "I want to make a reservation in an Italian restaurant" and "I need a table in a pizzeria". They both have the same meaning. That means the intent analysis goal is to identify the similarities between words in meaning. If you can't understand a user, your bot will be useless whatever the effort you consume to develop the other components such as dialogue management. How we can Teach semantics to a machine. We can do this with word vectors algorithms such as Word2Vec or Glove.* [10]

3) *Entity Linking: the bottom component in figure2 consists of a Disambiguation model and Named Entity Recognition (NER) and a template selection model. NER [11] links entity mentioned in a text to a dictionary or knowledge base, local or remote such as the whole web, to make sense of an entity and know what it is. This is a significant step for allowing the chatbot to understand conversation topics and generate appropriate responses. As it links the words of the user's utterances with concepts and subtexts in the real word. Options are to use StanfordNLP, TAGME [12] or web mining via a search engine API. TAGME can take the input text and returns a list of entities with their concept titles from wikipedia, which in turn can be converted to nodes in the Wikidata knowledge graph.*

B. Context Tracking

When a sentence from a user appears, the chatbot obtains the most recent utterances of that user from the chat history database. The Stanford CoreNLP toolkit [13] can be used to resolve coreference. If a coreferent was identified, the pronouns and the mentions of entities in the new sentence will be replaced.

C. Dialogue Manager

Dialogue Manager is the second core component in any chatbot and we can differentiate the chatbots through this component which have many parts can be improved or adding some parts in the future if we discover that, it will serve the DM. DM receives a user input from the NLU and produce the system responses at a concept level to the natural language generator (NLG). the response which the DM will choose, depends on the strategy that has been chosen. Strategies are related to maintaining conversational state and the ability to model the dialogue structure beyond that of a single utterance **[6]**. the strategies are rule-based, knowledge-based, retrieval based, and generative. The rule-based strategies are backstory, intent templates, and entity-based templates ordered by their priorities. Because rule-based strategies encode human knowledge into the form of templates, they provide the most accurate responses. The system will adopt a template reply if input is recognized by one of these strategies. If there is no matching template for the input, the system can try to get an answer from a knowledge-based question answering (Q/A) engine **[5]**. Failing that, the input is handled by an ensemble of neural network models and information retrieval modules to create a general conversation output

.

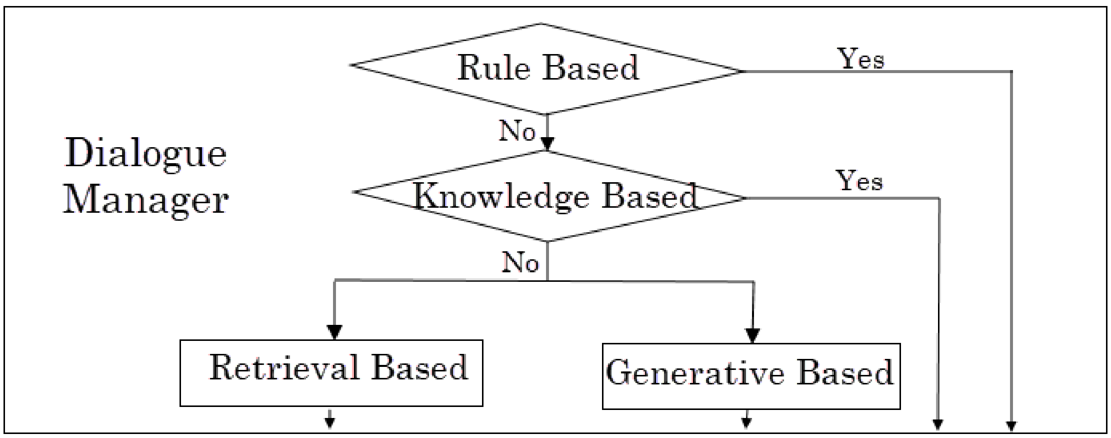


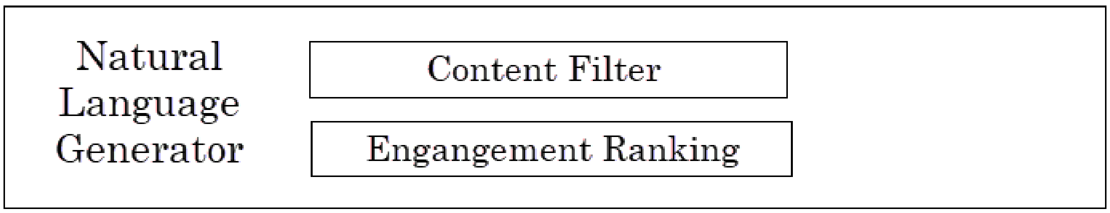
Figure 19: Dialogue Manager [Second Core Component]

1) *Online Information Retrieval: This module can be used when the rule-based and the knowledge-based is fail to provide responses ,it tries to provide more human-like, more concrete, and fresher responses compared to the rule-based templates and neural dialogue generation modules. The source of information for this module can be the most recent tweets provided by Twitter search API.**[14]** employed tweets as the source because they are usually short sentences closer to verbal language of most users compared to long written posts.*

2) *Neural Network Reply Generation: Generating sensitive context responses related to the conversation which helps make the user feel more engaging. This should be used for taking into account the recent utterances. Deep Learning approach and other Machine Learning techniques can be applied to develop.**[15]***

D. Natural Language Generator (Generator Reply)

The last main core component of any chatbot. It receives a communicative act from the Dialogue Manager (DM) and generates a matching textual representation. There are two functions that the NLG must perform: content planning (Content Filter | Engagement Ranking) and language Generation using just Text| Speech using Text to Speech). After going through one or more of strategies of the DM, the pipeline proceeds to the reply generator. This generator firstly will apply a content filter out incoherent or questionable candidates. If there are more than one valid response, a ranking process is used to reorder candidate utterances firstly according to the priority and then according to the engagement ranking. Finally, the chatbot will send the selected utterance as a text in the final output format or sending it to the Text to Speech to generate the final



output in Speech form. Simultaneously, all conversations are tracked in a history.

Figure 20: Natural Language Generator [Last Core Component]

5 CASE STUDY

KLM is a Dutch airline has an enormous number of flights around the world which was founded since 1919 so it was considered the world's oldest airline. KLM's owners thought about how they could improve their services for their clients. They found out that using the technology will be the key of success for this improvement. They are knowledge experts about having the information related to the area of study. They discovered that the social media is the most engagement way to the people. So they decided to ask the help from an AI experts about how they could benefit from the social media in the impact on the KLM airline services improvement.

A. Discussion

The main goal was to refining communication with their clients, making its client experience better by making it easier for people to talk to its agents via social media. KLM's owners studied their customers and knew that them spend a lot of time on social media like Facebook and Messenger. So, it can be the entry point and this about enabling the clients from sending any inquiry and may be reservation message privately to KLM's social media page. Because of the huge number of messages, the AI experts come with some bots such as KLM bot which provide as many as services like immediate reply on messages that can reach to 60000 message, Try before you fly with augmented reality, Tip advisor, and booking your favorite seat with more comfortability, privacy and more. You can use this bot through Facebook Messenger, Telegram, Twitter, and more other Social Media.

B. Evaluation

They made a survey at the end of each Messenger conversation, about asking the client some questions like "How you can feel comfortability with our Messenger bot?", "Are you like to recommend this bot for a friend?", and so on. And the answers were options to choose from them scaled from [1-10].

This Messenger bot achieved a breakthrough through both meeting and exceeding the expectations. Since January 2017, the airline has achieved:

1) Increasing in the interaction with client through the Messenger reached to 40%.
2) The success of online boarding booking was 15% about this Messenger bot.

6 CONCLUSION

The discussion was around the chatbot or CA. Firstly we talked briefly about what is the CA and then moved on the main purpose of the discussion, which is about what is the chatbot system architecture and its main components. After that, we discussed in details NLU and it subs, DM and it subs, then finally the NLG or reply generator. Notice that the DM is the most important part of any chatbot system because its components can be improved continuously or we can develop some alternative subs. If you are interested in chatbots, you can use this paper as a reference either in developing chatbot system or working on improving the Dialog Manager components.

REFERENCES

[1] Shum, H., He, X. & Li, D. From Eliza to XiaoIce: challenges and opportunities with social chatbots. Frontiers Inf Technol Electronic Eng. 19, 10–26 (2018).

[2] Schuetzler, Ryan M.; Grimes, G. Mark; Giboney , Justin Scott; and Nunamaker, J ay F . Jr., " The Influence of Conversational Agents on Socially Desirable Responding", "in *proc. of the 51St Hawaii International Conference on System Sciences", (HICSS),* University of Hawaii at Manoa, 2018

[3] S. Rai, A. Raut, A. Savaliya and R. Shankarmani, "Darwin: Convolutional Neural Network based Intelligent Health Assistant," Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2018, pp. 1367-1371.

[4] Ashwram, Roprasad, Ckhatri, Anuvenk, Raeferg, Qqliu, Jeffnunn, Behnam, Chengmc, Nashish, Kinr, Kateblan, Warticka, Yipan, Hasong, Skj, Ehwang, Pettigru "*Conversational AI: The Science Behind the Alexa Prize*" 1st Proceedings of Alexa Prize (Alexa Prize 2017), 2018

[5] Boris Galitsky, *Developing Enterprise Chatbots, Learning Linguistic Structures, Springer Nature Switzerland AG 2019*

[6] Jurafsky D., Martin J. H., Speech and language processing (Pearson International), 2nd edn. Pearson/Prentice Hall, Upper Saddle River. ISBN 978-0-13-504196-3, 2009

[7] Skantze G., "Error handling in spoken dialogue systems-managing uncertainty, grounding and miscommunication". Doctoral thesis in Speech Communication. KTH Royal Institute of Technology. Stockholm, Sweden,2007

[8] Lee C, Jung S, Kim K, Lee D, Lee G. G., "Recent approaches to dialog management for spoken dialog systems". *Journal of Computing Science and Engineering* (JCSE), 4(1):1–22,2010

[9] C. Khatri *et al*., "Contextual Topic Modeling For Dialog Systems, "*IEEE Spoken Language Technology Workshop (SLT)*, Athens, Greece, 2018, pp. 892-899. doi: 10.1109/SLT.2018.8639552, 2018

[10] Botfront.(2016), How intent classification works in NLU, https://botfront.io/blog/how-intent-classification-works-in-nlu#teaching-semantics-to-a-machine, (accessed 9 Jan 2020).

[11] Haptik Open source chatbot NER(2013) : https://haptik.ai/tech/open-sourcing-chatbot-ner/(accessed 12 January 2020)

[12] Ferragina P., Scaiella U. Tagme: "on-the-fly annotation of short text fragments (by Wikipedia entities)". In: Proceedings of the 19th ACM *international conference on information and knowledge management (CIKM). ACM*, New York, pp 1625–1628, 2010

[13] Manning CD., Surdeanu M., Bauer J, Finkel J, Bethard SJ, McClosky, "in proc. 52nd The Stanford CoreNLP natural language processing toolkit". *Annual Meeting of the Association for Computational Linguistics: System Demonstrations (ACLSD)*, pp 55–60, Baltimore, Maryland USA, June 23–24,2014

[14] Liu H, Lin T, Sun H, Lin W, Chang C-W, Zhong T, Rudnicky A. RubyStar: “A non-task oriented mixture model dialog system”. *First Alexa Prise comptions proceedings*, 2017

[15] *A. Sordoni, M. Galley, M. Auli, C. Brockett, Y. Ji, M. Mitchell, J.-Y. Nie, J. Gao, B. Dolan. A Neural Network Approach to Context-Sensitive Generation of Conversational Responses. In Proc. of NAACL-HLT. Pages 196-205,2015*

BIOGRAPHY

Moataz Mohammed is a teaching assistant of Computer Science department, Ain Shams University, Cairo, Egypt. Master candidate student in NLP and DL fields, B.Sc. of Computer Science Dept., Faculty of Computer and Information Sciences, in 2017, Ain Shams University, Cairo, Egypt.

Prof. Mostafa M. Aref is a professor of Computer Science and Chairman of Computer Science department, Ain Shams University, Cairo, Egypt. Ph.D. of Engineering Science in System Theory and Engineering, June 1988, University of Toledo, Toledo, Ohio. M.Sc. of Computer Science, October 1983, University of Saskatchewan, Saskatoon, Sask. Canada. B.Sc. of Electrical Engineering - Computer and Automatic Control section, in June 1979, Electrical Engineering Dept., Ain Shams University, Cairo, EGYPT.

بنية نظام المحادثة الآلية

[2]مصطفى محمود عارف ,[1]معتز محمد

*جامعة عين شمس, كلية الحاسبات والمعلومات ,*قسم علوم الحاسب*

*جامعة عين شمس, كلية الحاسبات والمعلومات ,**قسم علوم الحاسب*

[1]moataz.mohammed@cis.asu.edu.eg

[2]aref_99@yahoo.com

نبذه مختصرة **ـ في واحدة من أكثر المجالات اهتماما في الفترة الأخيرة ألا وهي المحادثات الآلية حيث تستطيع ان تفتح البحث في أكثر من مجال في مجالات علوم الحاسب حيث انها تحتاج الى فهم كيف للحاسب ان يتعامل ويفهم لغة الإنسان وكيف يمكننا ان نجعل هذا الحاسب ان يطبق خصائص الذكاء الاصطناعي والتعلم العميق في إمكانية اخذ القرار والرد المناسب. في هذه الورقة البحثية سوف نناقش معا كيف يكون الهيكل البنائي لبناء أي محادثة آلية وسيتم هذا عن طريق مناقشة تفصيلية لكل مكون فيه.**

الكلمات الرئيسة:ـ

وكيل المحادثة الآلية, المحادثات الآلية, نظام وكيل المحادثة الآلية, نظام الحوار, فهم اللغات الطبيعية, معالجة اللغات الطبيعية, التعلم العميق, بنية النظام

Arabic Optical Character Recognition Using Sequence to Sequence Models

Mohamed Sobhi*[1], Yasser Hifny**[2], Saleh Mesbah***[1]

[1]*Arab Academy for Science, Technology and Maritime Transport, Alexandria, Egypt.*

[2]*University of Helwan, Cairo, Egypt.*

* mohamed.sobhi@bibalex.org
** yhifny@fci.helwan.edu.eg
*** saleh.mesbah@aast.edu

Abstract—Optical character recognition (OCR) software is used to convert scanned documents into text. Arabic OCR is an active area of research where high accuracy is demanding. This paper focuses on building a model for converting images that contain Arabic text into their corresponding text using a deep learning approach. This model does not require any knowledge of the underlying language and it is simply trained end-to-end on different datasets. It combines several standard neural components from vision and natural language processing. Features are extracted from images using Convolutional Neural Networks (CNNs) where the features are arranged in a grid. Each row is then encoded using a Recurrent Neural Networks (RNNs). An RNN decoder with a visual attention mechanism is used to generate the output text. Our preliminary experiments show that the presented approach is effective. The obtained accuracy is in the range of 97.5% - 99.1%.

Keywords: **Sequence to Sequence Model, Arabic OCR, Convolutional Neural Networks (CNNs), Recurrent Neural Network (RNN), Attention Mechanism, Long Short Term Memory (LSTM) .**

1 INTRODUCTION

Print disability represents an obstacle against some people concerning gaining information from printed material in an identical method that makes them rely upon alternative ways to access these printed data using special format (i.e. braille, large print, audio, digital text). Print disabilities involve Visual impairments, learning disabilities, and physical disabilities that prevent the ability to read a book in some way. Print disabilities guidelines suggest presenting information in various modalities, e.g. through vision, hearing or touch. These guidelines present an accessible format for users. For example, visual texts enable users to enlarge texts, amplify sounds, and click for supporting information like definitions and images.
To satisfy these requirements, there must be a high-quality application of optical character recognition OCR in order to analyze the picture of the page, find the letters and words and generate text file paragraphs and pages because the image cannot transfer into Braille or synthetic voice output directly [1].
Arabic OCR has many challenges such as the variations of Arabic characters according to their position in the word where every character may have two or four different forms. Thus, the number of classes will increase to be recognized from 28 to 100. Moreover the cursive nature of Arabic writing does not allow direct implementation of many algorithms formulated for other [2].
This paper presents a model that is built to transfer the Arabic text images to their corresponding texts depending on the recent deep-learning techniques. The model uses different Arabic datasets. Section 2 reviews the related work of OCR. Section 3 explains the used model for supporting the current proposed work. Section 4 describes the used datasets and how these data sets are preprocessed to be used in the current model. Section 5 details the experiments used for building the model. Section 6 reviews the results of the proposed work and finally. Finally, Section 7 concludes the paper.

2 RELATED WORK

This section provides the appropriate background on previous work on image to text generation. Recently, numerous methods have been proposed for generating the text from image. Many of these methods are based on recurrent neural networks (RNN) and inspired by the successful use of sequence to sequence Models with neural networks.

A synthetic text generation engine to train three different models for scene text recognition is proposed: 90K-way dictionary encoding, character sequence encoding and bag-of-N-grams encoding. The experiments showed that the synthetic dataset is realistic and sufficient that it can efficiently replace real data [3].

A method to eliminate the need of pre-processing stages and the use of multiple neural networks in OCR by training a single neural network is proposed [4]. The network is used to detect and recognize text relying on semi-supervised learning. The proposed network was firstly evaluated on SVHN dataset, then it was evaluated on ICDAR 2013, SVT and IIIT5K datasets with accuracies 90.3%, 79.8% and 86% respectively. Finally, it was evaluated on FSNS dataset with 97% accuracy for character recognition and 71.8% accuracy for words.

A framework for lexicon-free OCR is proposed [5]. The proposed method mainly uses: recursive CNNs for image feature extraction, RNNs for character-level language modeling with soft-attention mechanism. However, the proposed method did not use LSTM. The framework was evaluated on ICDAR 2003, ICDAR 2013, Street View Text and IIIT5K datasets with accuracies 88.7%, 90%, 80.7% and 78.4% respectively for unconstrained text recognition.

In [6], they demonstrated the convenience of sequence to sequence learning in OCR by proposing a recurrent encoder-decoder framework. Attention based model was not used, but the paper shows the effect of LTSM in the sequence to sequence learning. The proposed model outperforms LSTM with CTC output layer. However, it requires more memory space.

In [7], they proposed a method to identify multiple scripts at the same text-line level. The method is based on sequence learning model with LSTM capabilities, where a 1D-LSTM architecture is used. The model was evaluated on an English-Greek data and it gives 98.186% accuracy.

In [8], they presented WYGIYS (What You Get Is What You See) model for OCR of presentational markup. The model is based on convolutional networks with visual attention. They introduced a new dataset IM2LATEX-100K, besides a synthetic dataset of webpages paired with HTML, for training and evaluating the model. The model gives 75% accuracy.

In [9], they proposed a model based on sequence-to-sequence architecture and a convolutional network to recognize handwritten text. The proposed model mainly consists of: A convolutional network for features extraction, an LSTM-based RNN for encoding and another LSTM-based RNN with attention for decoding. The model was evaluated on IAM and RIMES databases on which it gives 12.7% and 6.8% for word error rate, respectively.

In [10], they introduced ASTER (Attentional Scene Text Recognizer with Flexible Rectification) to recognize text with distortions and irregular layout. ASTER is an end-to-end model that includes both a rectification network and recognition network. The rectification network is based on Spatial Transformer Network, while the recognition network is based on attentional sequence-to-sequence model. The proposed model is trained on two synthetic datasets: Synth90K and Synth-Text, and is evaluated on: IIIT5K, SVT, ICDAR 2003, ICDAR 2013, SVTP and CUTE with accuracies 92.67%, 91.16%, 93.72%, 90.74%, 78.76% and 76.39%, respectively.

In [11], they proposed a system of three-neural network models for Arabic OCR. The first network is a three-layer neural network to recognize the text font size, then the text is normalized to 18pt font size to be fed to the second two networks. The second network model is a multi-channel neural network with three-window input. Finally, the third network is a two-layer convolutional network with two max pooling layers. The whole pipeline was evaluated on APTI dataset and gives an accuracy of 94.8%.

3 MODEL

This model is mainly based on combining a number of standard neural components for vision and natural language processing. Firstly, it uses a Convolutional Neural Network (CNN) for features extraction from the image, which are then arranged in a grid. Secondly, it uses two Recurrent Neural Networks (RNN) for as an encoder-decoder. The first RNN encodes each row in the feature grid, then the second RNN, which is enhanced with a visual attention mechanism, decodes the encoded features. The full structure is illustrated in Figure1.

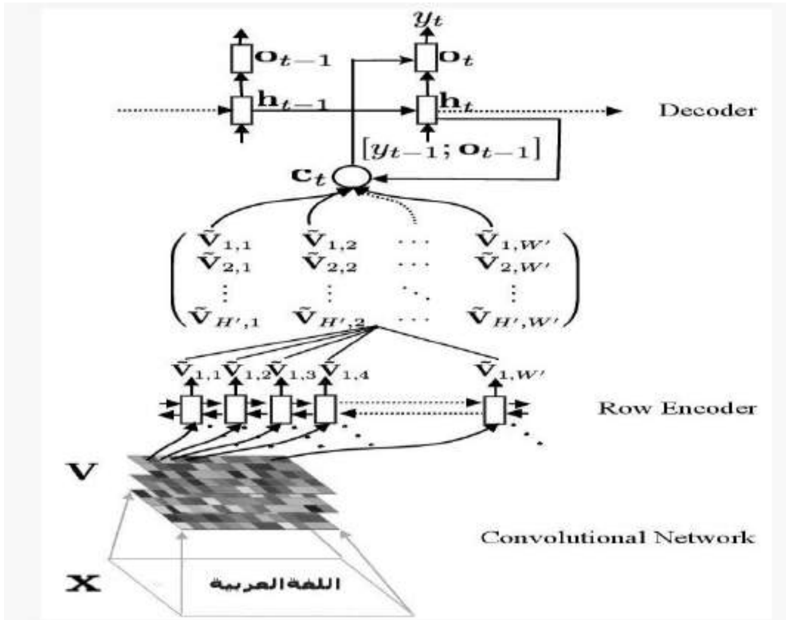


Figure 1. The full structure [12]

Model Structure:

- A CNN is used to extract a feature map V from the input image.
- Arrange the features in a grid.
- An RNN is used to encode spatial layout information for each row in the feature map.
- An RNN is used to decode with a visual attention mechanism to produce final outputs.

Convolutional Network: A Convolutional Neural Network (CNN, or ConvNet) is a sort of deep learning, feedforward Artificial Neural Network used extensively in analyzing visual image recognition tasks. It is mainly composed of an input layer, an output layer, and multiple hidden layers: convolutional layer, pooling layer, loss layer (drop out), fully connected layer (dense layer – Multi-layer Perceptron). [13] The visual feature extraction in this model is achieved through a multi-layer convolutional neural network interleaved with max-pooling layers. This network is considered a standard architecture nowadays. The convolution layer neurons execute a dot product between the image pixels and a filter. Then a Rectified Linear Unit (ReLu) and pooling layer follow each convolution layer. The ReLu is a non-linear activation function performed on each element to detect the negative values and threshold them to zero, such as max(0,x). For example: the output of ReLu (2, -3) will be (2,0). The pooling layer mainly reduces the dimensions of the image as the computation moves towards successive layers, and its output is then the input for down-sampling. The image-dimensions reduction results in reducing the number of parameters which helps in controlling the problem of overfitting. The CNN is designed such that pooling layers are interleaved in-between successive convolutional layers. The pooling layer acts on each features map independent of others and resizes it by performing a MAX operation. For example, in Figure 2, an image of dimensions 4*4 will be reduced by a pooling layer having a filter of dimensions 2*2 and stride 2. The output image from this layer will be 2*2, thus it has been reduced to 50% of its previous size. The ***max*** is performed to get the greatest number from the numbers that are in the filter's window [14].

$$\max(2, 1, 0, 3) = 3$$

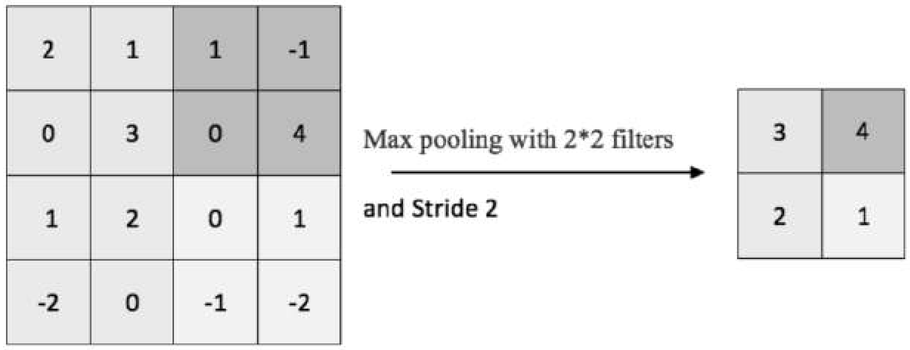


Figure 2: Pooling layer operation [14]

In this model, as we use visual attention model, we have to preserve the locality of CNN features, so final fully-connected layers are not used. The CNN accepts the raw input $R^{H \times W}$ and outputs a grid of features V of size $D \times H^1 \times W^1$, where D is the number of channels, H^1 is the reduced-sized grid height and W^1 is its width.

Row Encoder: Unlike image captioning, where the CNN features are used as they are, in OCR the encoder has to localize the relative positions within the image. This is achieved through running RNNs over each row of the CNN features.

A recurrent neural network (RNN) is a neural network characterized by the concept of 'internal memory', which differentiate it from other feedforward networks. The RNN internal memory is created by cyclic connections through its units. RNN is composed of three input layers (X_{t-1}, X_t, X_{t+1}), three hidden layers (h_{t-1}, h_t, h_{t+1}), three output layers (y_{t-1}, y_t, y_{t+1}), and weight matrices (W, U, V). The RNN accepts one input at time, for example, at time t the network is fed with input X_t, then it goes through the hidden layers for output prediction. The hidden layers represent the core of the network, the reason behind their importance is that they keep track of previous inputs. Each hidden layer accepts input from its previous hidden layer, in addition to its input unit, to predict the output. The weigh matrix of the hidden layer must be squared to maintain the same number of inputs as there are outputs.

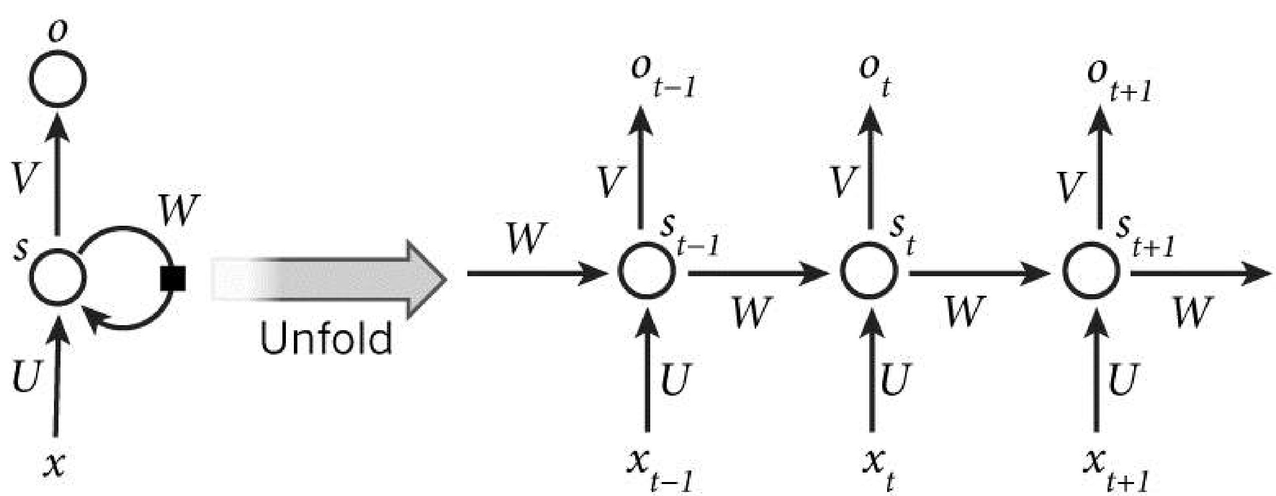


Figure 3: RNN Architecture [15]

The weight matrices (W, U, V) are initially set randomly. As for the first hidden layer (h_{t-1}), it is initialized by the dot product performed between its current input X_{t-1} at time t-1 and its weight matrix U, since it has no prior hidden layer and thus no contributions for the output prediction. The activation function such as sigmoid function then accepts the resulted dot product to generate values for the first hidden layer. Generally, data is processed in RNN by accepting an input X_t at time t, multiplying it with the weight matrix U and then passing it through hidden layer h_t. The hidden layer h_t also accepts another input from its prior hidden layer h_{t-1}, which has been parameterized with the weight matrix W. These two inputs contribute in predicting the output y_t, where a dot product is performed on the hidden layer h_t and the weight matrix V. This flow is repeated until all layers are covered to predict the final output.

Among the different variants of RNN, long short-term memory networks (LSTMs) [16] have shown effectiveness in the majority of NLP tasks. Thus, in this model the experiments use RNN with LSTM networks.

The long-short-term memory (LSTM) neural network is an extended version of simple RNN, where the past and current memories are related through linear dependence, instead of the non-linear connection between the past and current layers' activity. Furthermore, and most importantly, an LSTM has an introduced forget gate to adjust each of the past memory elements to contribute to the current memory cell.

Generally, each layer of the LSTM network accepts three inputs: X_t, h_{t-1} and C_{t-1}, which denotes the input at time t, the output of the previous LSTM unit, and the memory of the previous unity, respectively. The memory unit represents the most important unit in LSTMs, as it differentiates them from RNNs. The current layer has output h_t and the current unit has memory C_t. LSTMs have shown their effectiveness in captioning long term temporal dependencies. This obviously helped in improving the state-of-the-art for many difficult problems including: handwriting recognition and generation, language modeling and translation, acoustic modeling of speech, speech synthesis, protein secondary structure prediction and analysis of audio and video data among others. [17]

In this model, the new feature grid $\mathbf{V}$ is created from $\tilde{\mathbf{V}}$ by running an RNN across each row of that input. Recursively for all rows $h \in \{1, \ldots, H\}$ and columns $w \in \{1, \ldots, W\}$, the new features are defined as

$$\mathbf{V}_{hw} = \mathbf{RNN}\,(\mathbf{V}_{h,w-1}, \mathbf{V}_{hw}) \quad (1)$$

In order to capture the sequential order information in vertical direction, i use a trainable initial hidden state $\mathbf{V}_{h,0}$ for each row, which we refer to as positional embeddings.

This model uses an attention-based encoder-decoder composed of a pair of recurrent neural networks (RNNs). As for the encoder, it maps the variable-length input sequence to a vector. The decoder then maps the vector back to a variable length output sequence. To achieve maximum conditional probability of an output sequence given an input sequence, the two networks are trained jointly. [18]

Decoder Only considering the grid $\mathbf{V}$, the decoder predicts the tokens of output text $\{\mathbf{y}_t\}$. The decoder is trained as a conditional language model in order to predict the probability of the following token taking in consideration the history and annotations. The mentioned language model is described on top of a decoder RNN,

$$p(y_{t+1}|y_1, \ldots, y_t, \mathbf{V}) = \mathrm{softmax}(\mathbf{W}^{out}\,\mathbf{o}_t) \quad (2)$$

Where $\mathbf{o}_t = \tanh(\mathbf{W}^c[\mathbf{h}_t; \mathbf{c}_t])$ and $\mathbf{W}^{out}, \mathbf{W}^c$ are learned linear transformations. The vector h_t is used to summarize the decoding history: $\mathbf{h}_t = \mathbf{RNN}(\mathbf{h}_{t-1}, [y_{t-1}; \mathbf{o}_{t-1}])$. The context vector $\mathbf{c}_t$ is used to capture the context information from the annotation grid.

Attention Model: This model follows the mechanism of highly resolving a certain image region and fading the surrounds to focus on the specified region. Using the attention method eases the long-term dependencies modeling by adjusting the focal point over the period. This mechanism arises direct dependence between the model states at different time, which consequently introduces a hidden state h_t at each time step. [19]

4 DATA DESCRIPTION AND PREPROCESSING

The used datasets in the current work was selected from two sources. The first one is the LDC's Arabic Treebank PART 3 (ATB3). It is about 280,000 words that is select from "An-Nahar" Lebanese News. The second one is the Holy Quran without the Verses Numbers. Two versions of each dataset are used; diacritized and undiacritized with different formats.

Before beginning the experiments, each dataset is preprocessed as following:

1. A python script is used to separate the texts into lines where the maximum number of characters per each line is 80 characters by spaces. Each line is generated in Traditional Arabic with different sizes.
2. Another python script is used to generate the corresponding image of each line with different Dots Per Inch (DPI) to have different image qualities. A random name is generated to each image file and it is linked to its correspondent line.
3. A vocab list of the set of the date (unique characters) is generated.
4. Spaces in each line is marked by "<SP>" and a space is added after each character to be used in the proposed model.

5 EXPERIMENTS

For training the system, 70-80% of the images are selected from the data sets randomly and 10-15% of the images are selected for testing and validating the system. The training data sets are used in four experiments using Torch [20] based on the Open NMT (Neural Machine Translation) system [21]. These experiments are run on a 24GB NVIDIA Tesla K80 GPU [22].

In the first experiment 5197 images of the holy Quran without diacritics are used; 4158 images for training, 520 images for validation and 519 images for testing. The second experiment uses 8940 images of the holy Quran without diacritics; 6258 images for training, 1341 images for validation and 1341 images for testing. The third experiment aims at training the model on different format of the diacritized holy Quran as table 1 shows. Equal amount of images number for training, validation and testing are used in each adopted format. In the run time process, both data sets are combined in training, validation and testing.

TABLE VIII

EXPERIMENT 3 ON THE HOLY QURAN WITH DIFFERENT FORMATS

Experiment	Description	Total	Train	Val	Test
3	Holy Quran with diacritics and different format	8940	7152	894	894
3.1	Bold Italic Underline	2980	2384	298	298
3.2	Bold Italic Underline Strikethrough	2980	2384	298	298
3.3	Bold Italic Underline Double Strikethrough	2980	2384	298	298

For conducting the fourth experiment the LDC data set is used and it is dealt with twice. At first, the undiacritized LDC dataset with different format is used. Then, diacritized LDC data set with different format is used. In the run time process, both data sets are combined in training, validation and testing. As the total of the images is 39454, and Table 2 shows the division of these images.

6 RESULTS

The performance measures used to evaluate the system is the Word Error Rate (WER); the percentage of words has at least one error. A python script is developed to calculate the WER of the system depending on the Levenshtein distance [23] to calculate the insertion, deletion and substitution percentage. In the first experiment the rate yielded 99.92% of accuracy. In the second experiment the rate yielded 99.02 of accuracy and the third one yielded 98.55% of the rate (part one 97.49% - part two 99.58% - part three 98.57). The main problem in the second and third experiments' results is the insertion of diacritics for undiacritized words (Surah's name) which arise the need to conduct the fourth experiment that combines between diacritized and undiacritized words. As a result, the fourth experiment yielded to an accuracy of 97.56%. Some sample test images and their prediction are shown in Figure 4

TABLE II
EXPERIMENT 4 ON LDC WITH DIFFERENT FORMATS

Experiment	Description	Total	Train	Val	Test
4	LDC	39454	31564	3946	3944
4.1	without diacritics Regular	1640	1312	164	164
4.2	without diacritics Bold	1640	1312	164	164
4.3	without diacritics Italic	1640	1312	164	164
4.4	without diacritics Underline	1640	1312	164	164
4.5	without diacritics Double Strikethrough	1640	1312	164	164
4.6	without diacritics Bold Italic	1640	1312	164	164
4.7	without diacritics Bold Underline	1640	1312	164	164
4.8	without diacritics Bold Double Strikethrough	1640	1312	164	164
4.9	without diacritics Italic Underline	1640	1312	164	164
4.10	without diacritics Italic Double Strikethrough	1640	1312	164	164
4.11	without diacritics Underline Double Strikethrough	1640	1312	164	164
4.12	without diacritics Full Format	1687	1350	169	168
4.13	with diacritics Regular	1640	1312	164	164
4.14	with diacritics Bold	1640	1312	164	164
4.15	with diacritics Italic	1640	1312	164	164
4.16	with diacritics Underline	1640	1312	164	164
4.17	with diacritics Double Strikethrough	1640	1312	164	164
4.18	with diacritics Bold Italic	1640	1312	164	164
4.19	with diacritics Bold Underline	1640	1312	164	164
4.20	with diacritics Bold Double Strikethrough	1640	1312	164	164
4.21	with diacritics Italic Underline	1640	1312	164	164
4.22	with diacritics Italic Double Strikethrough	1640	1312	164	164
4.23	with diacritics Underline Double Strikethrough	1640	1312	164	164
4.24	with diacritics Full Format	1687	1350	169	168

Figure 4. Samples of some scanned images that contain Arabic text into their corresponding text.

7 CONCLUSION

This paper introduces a deep learning framework for converting the scanned images that contain Arabic text into their corresponding text. This model does not require any knowledge of the underlying language. It is simply trained end-to-end on different datasets. Convolutional Neural Networks (CNNs) are used to extract salient features from images and an RNN decoder with a visual attention mechanism is used to generate the output text. The preliminary experiments show that the presented approach is effective. The obtained accuracy is from 97.5% to 99.1%. This work can be extended further in numerous ways, such as training this model to identify other fonts.

ACKNOWLEDGEMENT

The authors would like to thank Bibliotheca Alexandrina for granting the permission to use its computational resources provided by the High-Performance Computing (HPC). (hpc.bibalex.org)

REFERENCES

[1] M. Hersh and M. A. Johnson, Assistive Technology for Visually Impaired and Blind People, (2008).

[2] A. M. Zeki and M. S. Zakaria, "Challenges in Recognizing Arabic Characters," *, (*2004).

[3] M. Jaderberg, K. Simonyan, A. Vedaldi and A. Zisserman, "Synthetic Data and Artificial Neural Networks for Natural Scene Text Recognition," *arXiv preprint arXiv:1406.2227, (*2014).

[4] C. Bartz, H. Yang and C. Meinel, "STN-OCR: A single Neural Network for Text Detection and Text Recognition," *arXiv preprint arXiv:1707.08831, (*2017).

[5] C.-Y. Lee and S. Osindero, "Recursive Recurrent Nets with Attention Modeling for OCR in the Wild," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* (2016).

[6] D. K. Sahu and M. Sukhwani, "Sequence to Sequence Learning for Optical Character Recognition," *arXiv preprint arXiv:1511.04176, (*2015).

[7] A. Ul-Hasan, M. Z. Afzal, F. Shafait, M. Liwicki and T. M. Breuel, "A sequence learning approach for multiple script identification," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR),* (2015).

[8] Y. Deng, A. Kanervisto and A. M. Rush, "What You Get Is What You See: A Visual Markup Decompiler.," *arXiv: Computer Vision and Pattern Recognition, (*2016).

[9] J. Sueiras, V. Ruíz, Á. Sánchez and J. F. Vélez, "Offline continuous handwriting recognition using sequence to sequence neural networks," *Neurocomputing,* vol. 289, pp. 119-128, (2018).

[10] B. Shi, M. Yang, X. Wang, P. Lyu, C. Yao and X. Bai, "ASTER: An Attentional Scene Text Recognizer with Flexible Rectification," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 41, no. 9, pp. 2035-2048, (2019).

[11] M. A. Radwan, M. I. Khalil and H. M. Abbas, "Neural Networks Pipeline for Offline Machine Printed Arabic OCR," *Neural Processing Letters,* vol. 48, no. 2, pp. 769-787, (2018).

[12] Y. Deng, A. Kanervisto, J. Ling and A. M. Rush, "Image-to-Markup Generation with Coarse-to-Fine Attention," in *International Conference on Machine Learning*, (2017).

[13] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," *Intelligent Signal Processing,* pp. 306-351, (2001).

[14] A. Karpathy, "Stanford university cs231n: Convolutional neural networks for visual recognition," 2018. [Online]. Available: http://cs231n. stanford. edu/syllabus. html/.

[15] D. Britz, "Recurrent neural networks tutorial, part 1–introduction to rnns.," (2015). [Online]. Available: http://www. wildml. com/2015/09/recurrent-neural-networkstutorial-part-1-introduction-to-rnns/. [Accessed 1 February 2019].

[16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation,* vol. 9, no. 8, pp. 1735-1780, (1997).

[17] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink and J. Schmidhuber, "LSTM: A Search Space Odyssey," *IEEE Transactions on Neural Networks,* vol. 28, no. 10, pp. 2222-2232, (2017).

[18] K. Cho, B. v. Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk and Y. Bengio, "Learning Phrase Representations using RNN Encoder--Decoder for Statistical Machine Translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (2014).

[19] C. Raffel and D. P. W. Ellis, "Feed-Forward Networks with Attention Can Solve Some Long-Term Memory Problems," *arXiv preprint arXiv:1512.08756, (*2015).

[20] R. Collobert, K. Kavukcuoglu and C. Farabet, "Torch7: A Matlab-like Environment for Machine Learning," in *BigLearn, NIPS Workshop*, (2011).

[21] G. Klein, Y. Kim, Y. Deng, J. Crego, J. Senellart and A. M. Rush, "OpenNMT: Open-source Toolkit for Neural Machine Translation," *arXiv preprint arXiv:1709.03815, (*2017).

[22] "nvidia Web Site," [Online]. Available: https://www.nvidia.com/en-gb/data-center/tesla-k80/. [Accessed 1 February 2019].

[23] K. U. Schulz and S. Mihov, "Fast string correction with Levenshtein automata," *International Journal on Document Analysis and Recognition,* vol. 5, no. 1, pp. 67-85, (2002).

Biography

Mohamed Sobhi

Sobhi is an Advanced Specialist in Assistive Technology Training at the Bibliotheca Alexandrina, he graduated from the faculty of Arts (Arabic Section) in 2004 and completed his premaster in linguistics in 2005. He also holds a master degree in Phonology with an overall grade: Excellent. In 2016 he completed a diploma in Information system and is currently a postgraduate researcher at the college of computing and Information Technology at the Arab Academy for Science, Technology and Maritime Transport.

Dr. Yasser Hifny

He is an associate professor at the University of Helwan, Egypt. Dr. Hifny received his M.Sc. and B.Sc. degrees in electronics and communication engineering in 1995 and 2000 respectively, from the University of Cairo, Egypt. In 2006, he received the PhD degree from the University of Sheffield, U.K. In 2000, he joined Research and Development International (RDI), where he worked on research in text-to-speech (ArabTalk), limited domain speech compression, and speech verification (HAFS) projects. In addition, he was a postdoctoral fellow in the Human Language Technologies Group at IBM T.J. Watson Research Center in 2007.

Dr. SALEH MESBAH ELKAFFAS

Dr. Saleh Mesbah Elkaffas is an Associate Professor at the Department of Information Systems, College of Computing and Information Technology, Arab Academy for Science, Technology and Maritime Transport (AAST). He is currently the director of the Remote Sensing and Spatial Studies Unit, College of Engineering and Technology, AAST. He received his B.Sc. in Electronics and Telecommunications Engineering and M.Sc. in information content in remote sensing. He received his Ph.D. degree under the channel system supervision from the Center for Remote Sensing, Imperial College, UK, and Alexandria University. He has over 35 years of experience in the field of Remote Sensing, GIS, Computing and Information Technology. Over the span of his career, Dr. Saleh Mesbah has published a number of papers and book chapters in the fields of Remote Sensing, GIS, Digital Image Processing, Advances in Intelligent Systems and Computing, Applied Geomatics, Sensing and Imaging and Environmental Management. His research interests include Satellite Remote Sensing, Digital Image Processing, Geographic Information Systems (GIS), Global Positioning Systems (GPS), Information Technology and Environmental Manageme

التعرف الضوئي على الحروف العربية باستخدام نماذج التسلسل إلى التسلسل

محمد صبحي *[١] ، ياسر حفني **[٢] ، صالح مصباح ***[١]

١ الأكاديمية العربية للعلوم والتكنولوجيا والنقل البحري ، الإسكندرية ، مصر.

٢ جامعة حلوان ، القاهرة ، مصر.

mohamed.sobhi@bibalex.org *

yhifny@fci.helwan.edu.eg **

saleh.mesbah@aast.edu ***

الخلاصة — يستخدم برنامج التعرف الضوئي على الأحرف (OCR) لتحويل المستندات الممسوحة ضوئيًا إلى نص. تعد التعرف الضوئي على الحروف العربية مجالًا نشطًا للبحث حيث تتطلب الدقة العالية. توقدم هذه الورقة بناء نموذج لتحويل الصور عربي الى النص المقابل لها باستخدام نهج التعلم العميق. لا يتطلب هذا النموذج أي معرفة باللغة الأساسية ويتم تدريبه ببساطة من طرف إلى طرف على مجموعات بيانات مختلفة. فهو يجمع بين العديد من المكونات العصبية القياسية من الرؤية ومعالجة اللغة الطبيعية. يتم استخراج الميزات من الصور باستخدام الشبكات العصبية التلافيفية (CNNs) حيث يتم ترتيب الميزات في شبكة. ثم يتم تشفير كل صف باستخدام الشبكات العصبية المتكررة (RNNs). يتم استخدام وحدة فك ترميز RNN مع آلية الاهتمام البصري لإنشاء نص الإخراج. تظهر تجاربنا الأولية أن النهج المقدم فعال. الدقة التي تم الحصول عليها هي في حدود ٩٧,٥ ٪ الى ٩٩,١ ٪.

الكلمات المفتاحية: **نموذج التسلسل إلى التسلسل ، التعرف الضوئي على الحروف العربية ، الشبكات العصبية التلافيفية (CNN) ، الشبكة العصبية المتكررة (RNN) ، الاهتمام البصري.**

Whale Swarm Algorithm Methodology for Text Mining

Amr M Sauber*1, Passent M. El-Kafrawy*,**2, Amr. F. Shawish*3

*Faculty of Science, Menoufia University, Egypt
1amr@science.menofia.edu
2basant.elkafrawi@science.menofia.edu.eg
3amr_shawish@yahoo.com

**School of Information Technology and Computer Science, Nile University, Egypt
2pelkafrawy@nu.edu.eg

Abstract: **growing nature-inspired meta-heuristic algorithms are used to resolve real-world optimization issues, as they have some benefits over the classical techniques of numerical optimization. This paper explains the methodology of nature-inspired meta-heuristic called Whale Swarm Algorithm (WSA) for feature optimization, which is inspired by the whales' behavior of communicating with every different other via ultrasound for hunting. Text mining is used in every field for business intelligence, social media analysis, sentiment analysis, biomedical analysis, software process analysis and even for security analysis. This paper discusses different application of WSA as an optimization algorithm specifically the automation of understanding of Arabic text into ontology construction in the state-of-the-art literature. The key issues which are involved in the WSA enhancement models are also discussed here. This paper presents an up-to-date review over the uses of WSA in different fields to improvement of swarm optimization applications with focus on ontology learning from Arabic text.**

Key words: **optimization, swarm intelligence, whale swarm algorithm, whale optimization algorithm, ontology, Arabic text mining, concepts mapping.**

1 INTRODUCTION

Optimization is utilized in different applications. In the manufacturing of a new device, in a new artificial intelligence method, in big data application or in deep learning network, optimization is the most vital phase of any application. In order to develop a device with optimum sizes utilizing minimal power, to train a network, in order to limit the desired between the desired output and actual output values, optimization is desired. Text Mining has become an important research area. Text Mining is the discovery by computer of new, previously unknown information, by automatically extracting information from different written resources. Text mining can work with unstructured or semi-structured data sets such as emails, full-text documents and HTML files etc. Text Mining is widely used in field of Natural Language Processing and Multilingual Aspects. The Data Mining Optimization Ontology (DMOP) has been developed to support informed decision-making at various choice points of the data mining process. An evolutionary approach that combines information extraction technology and genetic algorithms can produce a new, integrated model for text mining. Text mining discovers unseen patterns in textual databases. The major uses of a text mining tool are for: Text Analytics, Text Processing, Classification/Categorization, Sentiment Analysis and Knowledge Discovery. Genetic Algorithms are the algorithms used to solve optimization problems. These algorithms are search based algorithm used to generate useful solutions for search problems [1].

Nature-inspired algorithms are becoming vigorous in solving numerical optimization issues, specially the NP-hard issues for example, the travelling salesman issue [3], vehicle routing [4], classification issues [5], routing issue of wireless sensor networks (WSN) [6] and multiprocessor scheduling issue [7], etc. These real world optimization issues often likely come with more global or local optima of a given mathematical model. Whereas, a point-by-point classical technique of numerical optimization is utilized for this task, the classical technique has to try repeatedly for locating various optimal solutions in every iteration [8], which takes a lot of time and work. Thus, using nature-inspired meta-heuristic algorithms to solve these issues has become an important research topic, as they are simple to execute and can converge to the global optima with high probability. In this research, we shall discuss a new nature-inspired meta-heuristic called Whale Swarm Algorithm (WSA) or Whale Optimization Algorithm (WOA) for function optimization; rely on the whales' behavior of communicating with every other through ultrasound for hunting. Hence, a brief overview of the nature-inspired meta-heuristic algorithms is explained. GA and WSA where merged in a new methodology, G-WSA, to extract concepts in text mining. The approach where utilized to construct Arabic ontologies from Arabic text [35]. The concepts where automatically extracted by optimizing the identification of related concepts and their relationships to parent concepts.

In this paper after the introduction in section 2 the natural phenomena of Whales Swarm will be retrieved, then detailed definition of the WSA is explained with the algorithm details and the mathematical background is defined in sections 3 and 4 consequently. In section 5 applications of WOA is listed with related state of the art publications. Finally, a discussion and conclusion in sections 6 and 7.

2 BACKGROUND

A. Whale Hunting Behavior

Social animals which live in groups inside the sea are called Whales. They make various sounds to demonstrate their movement, sustaining and mating designs. Whales decide nourishment azimuth and stay in contact with one another from enormous separations by ultrasound.

For example, pregnant females will assemble with other female whales and calves in order to improve protection abilities. Also, sperm whales are frequently seen in gatherings of somewhere in the range of 15 to 20 population, as shown in figure. 1. The whale sounds are delightful tunes in the sea and their sound range is exceptionally wide. As of not long ago, researchers have found 34 types of whale sounds, for example, whistling, squeaking, moaning, yearning, thundering, chattering, clicking, humming, churring, talking, trumpeting, clopping, etc. These sounds made by whales can frequently be connected to significant capacities, for example, their relocation, encouraging and mating designs. What's more, whales decide food azimuth and keep in contact with one another from an extraordinary separation by the ultrasound which are past the extent of human hearing [2].

Figure 1: The swarm of sperm whales.

B. Swarm Behavior

Swarming, or swarm behavior is an aggregate swarm behavior shown by creatures of comparative size which total together, maybe processing about a similar spot or maybe moving as a group or relocating toward some path. As a term, swarming is connected especially to mealy bug, likewise can be connected to some other creature that shows swarm conduct. The term rushing is typically used to allude explicitly to swarm conduct in fowls, grouping to allude to swarm conduct in quadrupeds, shoaling or tutoring to allude to swarm conduct in fish. Phytoplankton likewise assembles in tremendous swarms called blossoms, despite the fact that these living beings are green growth and are not self-impelled the manner. The term swarm is connected likewise to lifeless substances which display parallel practices, as in a robot swarm, a seismic tremor swarm, or a swarm of stars. From an increasingly theoretical perspective, swarm conduct is the aggregate movement of

an enormous number of self-pushed elements. From the viewpoint of the numerical modeler, it is a developing conduct emerging from straightforward principles that are traced by people and does not include any focal coordination. Swarm conduct was first reproduced on a PC in 1986 with the reproduction program boids. This program recreates basic operators that are permitted to move as per a lot of fundamental standards. The model was initially intended to impersonate the rushing conduct of winged animals; however it tends to be connected additionally to trained fish and other swarming elements [2].
The boids PC program, made by Craig Reynolds in 1986. Many consequent and current models use minor departure from these guidelines, frequently executing them by methods for concentric "zones" around every animal. In the "zone of aversion", near the creature, the central creature will look to remove itself from its neighbors to maintain a strategic distance from impact. Second, in the "zone of alignment", the central creature will try to adjust its heading of movement to its neighbors. Third, "zone of fascination", which reaches out as far away from the central creature as it can detect, and the central creature will try to move towards a neighbor.

C. Swarm Intelligence

Swarm Intelligence (SI) is a sort of man-made consciousness that intends in order to mimic the conduct of swarms or social mealy bug. Swarm alludes to any inexactly organized accumulation of cooperating specialists. Actually swarms are viewed as decentralized self-sorted out frameworks. Swarm knowledge has a multidisciplinary character its investigation gives bits of knowledge that can enable people to oversee complex frameworks. There is no reasonable definition for swarm insight. Developing conduct, self-sorted out conduct and aggregate knowledge are the related terms. Shockingly swarm insight framework can act in a planned manner with no facilitator or outside controller.

3 WHALE SWARM ALGORITHM

A. Overview of Whale Swarm Algorithm

Whale swarm algorithm is developed for comprehending work enhancement issue, have romanticized a few chasing principles of whale. At the point when a whale has discovered food source, it will make sounds to tell different whale's close-by of the quality and amount of food. So every whale will get loads of warnings from the neighbors and after that transition to the legitimate spot to discover food dependent on these warnings. The behavior of whales connecting with one another through sound for hunting fill with us to build up another meta-heuristic algorithm for function optimization issues as shown figure 2.

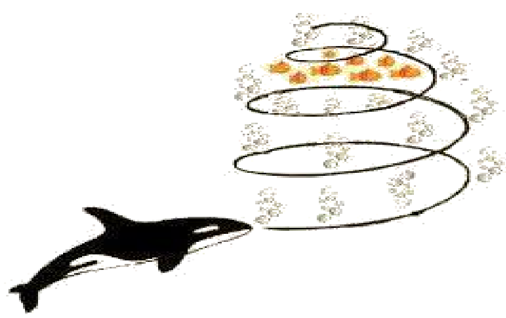


Figure 2: Method of WOA for accessing best solution

B. Whale Swarm Algorithm Rules

There are four rules are very important in building successful whale optimization algorithm as follows:

1. All the whales should connect with every other through ultrasound in the search region;
2. Every whale has a certain degree of computing capability in order to calculate the distance to other whales
3. The quality and quantity of food found through every whale is assigned to its fitness.
4. The motion of a whale is guided through the nearest one amongst the whales that are better (judged by fitness) than itself.

The flow chart of whale Swarm Algorithm is explained in figure 3 as follows:

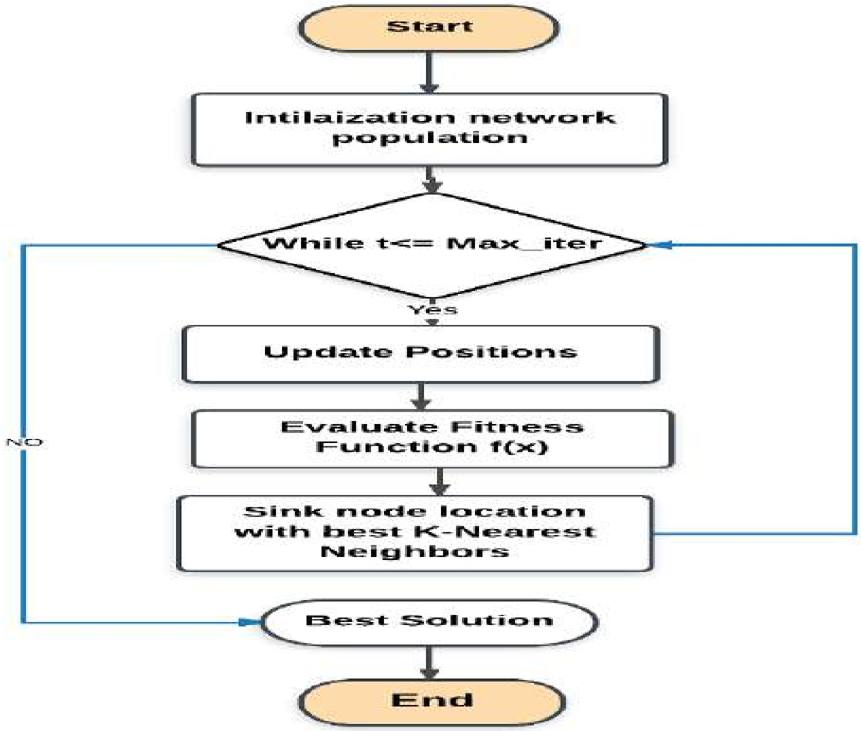


Figure 3. The flow chart of WSA technique

4 THE MATHEMATICAL MODEL OF WOA

The mathematical model for whale optimization algorithm is presented in details. The following functions describe the behavior of WOA to achieve encircling prey, feeding of spiral bubble-net maneuver, and search for prey.

The WOA algorithm supposes that the current best candidate solution is the target prey or is close to the optimum. After the best search agent is defined, the different search agents will hence attempt to update their positions in the direction of the best search agent. This conduct is represented with the aid of the following equations [9]

$$\vec{D} = |\vec{C} . \vec{X}^*(t) - \vec{X}(t)| \quad (1)$$

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A} . \vec{D} \quad (2)$$

Where t suggests the current iteration, A □ and (C) □are coefficient vectors, X* is the position vector of the best solution acquired so far, X □ is the position vector, | | is the absolute value, and • is an element-by-element multiplication. It is well worth citing here that X∗ should be updated in every new release if there is a better solution.

The vectors $\vec{A}$ and $\vec{C}$ are calculated as follows:

$$\vec{A} = 2\vec{a} . \vec{r} - \vec{a} \quad (3)$$

$$\vec{C} = 2.\vec{r} \quad (4)$$

Where (a) □ is linearly reduced from 2 to 0 over the direction of iterations (in each exploration and exploitation stages) and r □ is a random vector in [0, 1].

For a 2D issue, the position (X, Y) of a search agent can be up to date according to the position of the current best record (X *, Y*). Various places round the best agent can be done with respect to the current position through using adjusting the value of A □ and (C) □ vector. It should be mentioned that with the aid of defining the random vector (r □) it is viable to reach any function in the search space defined among the key-

points. Therefore, Eq. (2) permits any search agent to update its position in the neighborhood of the current best solution and simulates encircling the prey.
The same concept can be prolonged in order to a search space with n dimensions, and the search agents will go in hyper-cubes round the best solution got yet. As above-mentioned in the previous part, the humpback whales as well attack the prey with the bubble-net technique.
Then, Calculates the distance between the whale placed at (X, Y) and prey placed at (X*, Y*). A spiral equation is then created between the position of whale and prey to mimic the helix-shaped motion of humpback whales as follows:

$$\vec{X}(t+1) = \vec{D'} \cdot ebl. \cos(2\pi l) + \vec{X}^*(t) \qquad (5)$$

Where (D') □ = |X □^* (t) - (X() □t)| shows the distance of the i-th whale to the prey (best solution acquired yet), b is a regular for defining the shape of the logarithmic spiral, l is a random quantity in [−1, 1], and * is an element-by-element multiplication.
Humpback whales swim round the prey inside a shrinking circle and among a spiral-shaped path with each other. To mannequin this simultaneous behavior, we suggest that there is a probability of 50% to select between either the shrinking encircling mechanism or the spiral model to update the position of whales throughout optimization. The mathematical model is as follows:

$$\vec{X}(t+1) = \begin{cases} \vec{X}^*(t) - \vec{A} \cdot \vec{D} & \text{if } P < 0.5 \\ \vec{D'} \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) & \text{if } P \geq 0.5 \end{cases} \qquad (6)$$

Where p is a random number in [0, 1].
Humpback whales search randomly in accordance to the position of every other. Thus, we utilize A □ with the random values greater than 1 or less than −1 to force search agent to get about far away from a reference whale. In contrast to the exploitation stage, we update the position of a search agent in the exploration phase in accordance to a randomly chosen search agent instead of the best search agent found yet. This mechanism and | A □ | > 1 confirm exploration and permit the WOA algorithm to execute a global search. The mathematical model is as follows:

$$\vec{D} = |\vec{C} \cdot \overrightarrow{X_{rand}} - \vec{X}(t)| \qquad (7)$$

$$\vec{X}(t+1) = \overrightarrow{X_{rand}} - \vec{A} \cdot \vec{D} \qquad (8)$$

where $\overrightarrow{X_{rand}}$ is a random position vector (a random whale) chosen from the current population.

5 USES OF WOA

There are many uses for whale swarm optimization algorithm in various areas of life. Stochastic nature-inspired meta-heuristic algorithms have validated their power on the closing two decades in dealing with global optimization issues bobbing up often in engineering. Whale optimization algorithm is a recent swarm based meta-heuristic. It copies the pattern of spiral bubble net hunting pattern of humpback. Mohit Verma and Amit Kumar (2018) [10] offered a brief survey on the whale optimization algorithm with focus on its several applications over single objective and multi-objective optimization issues. First of all, the goal of single objective optimization is to attempt to uncover the exceptional solution that corresponds to the minimal or the most well worth of one objective that lumps all definitely various objectives into one. WOA has been utilized for the various single objective optimization issue. Table I lists some of the application of WOA in solving single-objective optimization issues.

C. Engineering Applications

TABLE I: APPLICATION OF WOA IN SINGLE-OBJECTIVE OPTIMIZATION

S. No.	Application name	Reference
1.	Bearing fault diagnosis using a WOA	[11]
2.	Welded beam design	[12]
3.	The Workflow Planning of Construction Sites using WOA	[13]

4.	Optimal siting of capacitors in radial distribution network using WOA	[14]
5.	A hybrid WOA and pattern search technique for optimal power flow problem	[15]
6.	Combined Emission Constrained Economic Dispatch with Valve Point Effect Loading Problem Solution using WOA	[16]
7.	An emission constraint environment dispatch problem solution with microgrid using WOA	[17]

On the other hand, there exists no most suitable solution in case of multi-objective optimization with contradictory objectives [18]. The co-occurrence of definitely distinctive targets leads to a collection of trade-off solutions popularly known as non-dominated or Pareto-optimal solutions. Multi-Objective Whale Optimization Algorithms (MOWOA) has been utilized for the different issues with multi objective optimization. Table II lists some of the applications of WOA in solving multi objective optimization issues.

TABLE II: APPLICATION OF WOA IN MULTI-OBJECTIVE OPTIMIZATION

S No	Application	Reference
1.	Economic and Emission Dispatch using WOA	[19]
2.	Multi-Objective Optimal Vehicle Fuel Consumption based on WOA	[20]
3.	Multi-objective optimal mobile robot path planning base on WOA	[21]
4.	An Ameliorative WOA for Multi-Objective Optimal Allocation of Water Resources	[22]
5.	A MOWOA for Solving Engineering Design Problems	[23]
6.	WOA for combined heat and power economic dispatch	[24]

D. Feature Selection Applications

Choosing of relevant benefits of a dataset is vital in high dimensional datasets to keep away from the curse of dimensionality. Feature Selection is carried out to decrease overfitting, to enhance accuracy and to decrease the training time of the algorithms. Bing Zeng, Liang Gao and Xinyu Li used WSA for feature subset selection [2]. P. Anuradha and Dr. Vasantha Kalyani David (2019) [25] focused on choosing a features subset utilizing Whale Swarm Algorithm (WSA) where Logistic Regression (LR), Random Forest (RF) and k-Nearest Neighbor (KNN) are utilized as fitness functions. These WSA-LR, WSA-RF and WSA-KNN combinations generate different feature subsets for different number of iterations. Then training and trying out is accomplished on the dataset with the subset of selected features using LR, RF, Support Vector Classifier (SVC) and Gaussian Naive Bayes (GNB) and prediction accuracies generated are analyzed.
In High dimensional datasets, Feature Selection pursuits at decreasing the redundant and irrelevant features. The refined dataset with only the relevant features would enhance the learning accuracy and decrease the learning time [25]. The features which are utilized to train the machine learning model highly influence the efficiency of the model. Irrelevant features can bring down the efficiency of the model. Feature Selection can be greatly labeled into filter technique, wrapper method and Embedded method. In Filter technique, different statistical tests are utilized to choose the features that rely on their correlation with the outcome or dependent variable. In wrapper technique, a subset of features is utilized to train a model. Based on the efficiency of the model, features will be added or deleted to/from the subset. In embedded technique, both the benefits of filter and wrapper techniques are combined. The embedded technique algorithm operates using subset selection, train a model and additionally execute a penalization function to limit overfitting.

The pseudo code of finding a whale's better and nearest whale [2]:

Input: The whale swarm n, a whale u.

Output: The better and nearest whale u.

1: begin

2: Define an integer variable v initialized with 0;

3: Define a float variable temp initialized with infinity;

```
for i=1 to n do
if i≠u then
    if f(whale i )
        if dist(whale i , whale u )
            v=i;
            temp=dist(whale i , whale u );
        end if
    end if
  end if
end for
return whale v;
end
```

The authors experimented three various classifiers for the fitness function namely, Logistic Regression, Random Forest and K-Nearest Neighbors and the subsets are obtained from various numbers of iterations. The dataset rely on the chosen subsets are then utilized for classification. The classification accuracy of four various classifiers namely, Random forest, Logistic Regression, Support Vector Classification and Gaussian Naive Bayes are compared. Among these WSA with Logistic Regression as the fitness function (WSA-LR) gives a subset of eight features on an average and the accuracy of Random Forest Classifier is found to be 85.7% which is better than the other classifiers. In future, other classifiers can be tried as fitness function in the WSA, and the prediction accuracy can be compared among other classifiers [26].

E. Clustering Applications

Clustering is a powerful method in data-mining, which entails identifying homogeneous corporations of objects based totally on the values of attributes. Meta-heuristic algorithms for example, particle swarm optimization, artificial bee colony, genetic algorithm and differential evolution are now becoming powerful techniques for clustering.

Clustering is aggregating unlabeled objects into corporations with similarities between these objects. Such that the objects in the identical clusters are extra similar to every different object in distinct clusters in accordance to some predefined criteria [27] and [28]. A variety of algorithms have been proposed that take into account the nature of the data, the volume of the information and different enter parameters in order to cluster the data. The similarity standards in clustering are a range of in different researches. Most of the clustering troubles have exponential complexity in terms of the quantity of clusters.

Lately, Mirjalili and Lewis [29] described a new swarm based meta-heuristic optimization algorithm that mimicks the social behavior of humpback whales in searching. The algorithm is inspired through the bubble net hunting delineation. They have tested the WOA algorithm with 29 mathematical benchmark optimization issues and compared the efficiency of WOA algorithm with other traditional current heuristic algorithms for example, Particle Swarm Algorithm (PSO) [29], Differential Evolutional(DE) [31], Gravitational Search Algorithm (GSA) [32] and Fast Evolutionary Programming [33]. WOA was identified to be ample aggressive with different general and popular meta-heuristic techniques.

Jhila Nasiri1 and Farzin Modarres Khiyabani (2018) [34] proposed a new meta-heuristic clustering technique, the Whale Clustering Optimization Algorithm, primarily rely on the swarm foraging behavior of humpback whales. After a detailed formula and explanation of its implementation, they compared the proposed algorithm with different existing well-known algorithms in clustering, including PSO, Artificial Bee Colony (ABC), GA, DE and k-Means. Proposed algorithm was once examined with the usage of one synthetic and seven real benchmark data units from the UCI computer mastering repository. Simulations exhibit that the proposed algorithm can efficiently be utilized for data clustering.

The consequences of their algorithm were contrasted with customary k-means clustering strategy and other popular stochastic algorithms such as PSO, ABC, DE, and genetic algorithm (GA) clustering. The Preliminary computational experience in terms of the intra-cluster distance function and standard deviation revealed that the whale optimization algorithm can successfully be applied to solve clustering issues. Furthermore, the results from the proposed algorithm was once effective, simple to execute and robust as compared with different strategies. There are some directions that can enhance the overall performance of the suggested algorithm in the future. The aggregate of WOA clustering algorithm with different clustering strategies and the usage of different fitness functions in clustering strategy should be considered in future researches.

F. Text Analysis

Rania M. Ghoniem et al. (2019) [35] proposed an optimized ontology learning from Arabic text. Ontology is a technique for extending web syntactic interoperability to semantic interoperability. Ontologies are exploited to signify massive information in such a way that permits machines to interpret its meaning, allowing it to be reused and shared. Their work was done in two phases. First, a text mining algorithm is proposed for extracting concepts and their semantic relations from text documents. The proposed algorithm calculated the *concept frequency weights* using the term frequency weights. Afterwards, they calculated the weights of thought similarity utilizing the facts of the ontology structure, involving (i) the concept's route distance, (ii) the concept's distribution layer, and (iii) the mutual mother or father concept's distribution layer. Then, feature mapping is carried out via assigning the concepts' similarities to the concept features. The second phase, a hybrid genetic-whale optimization algorithm was once proposed to optimize ontology learning from Arabic text. The operator of the G-WOA is a hybrid operator integrating GA's mutation, crossover, and selection processes with the WOA's procedures to achieve the stability between both exploitation and exploration, and to locate the solutions that showcase the best possible fitness. For estimating the overall performance of the ontology learning method, widespread comparisons are carried out utilizing extraordinary Arabic corpora and bio-inspired optimization algorithms. Moreover, two publicly accessible non-Arabic corpora are utilized to compare the performance of the proposed method with those of different languages. To validate the performance of the proposed G-WOA algorithm in mastering ontology from Arabic text, they compared the solution returned to those returned through the normal GA and WOA. The G-WOA starts off evolved to search for the fine answer via a set of iterations, which include embedding the genetic operators into the WOA architecture. Eventually, the algorithm returns the answer which recommends the great set of concepts/relations that can contribute to the ontology. The outputs revealed that the proposed genetic-whale optimization algorithm outperforms contrasting algorithms throughout all the Arabic corpora in precision, recall, and F-score measures. Moreover, the proposed approach outperforms the latest strategies of ontology learning from Arabic and non-Arabic texts in terms of these three measures.

This research contributes to todays Arabic ontology learning by the following: text mining algorithm is proposed specifically for extracting the ideas and their semantic relations from the Arabic documents. The extracted set of principles with the semantic relations constitutes the shape of the ontology. In this regard, the algorithm operated on the Arabic documents with the aid of calculating the concept frequency weights depending on the term frequency weights. Thereafter, they calculated the weights of concept similarity, using the information-driven from the ontology structure involving the concept's path distance, the concept's distribution layer, and the mutual parent concept's distribution layer. Eventually, it performs the mapping of aspects by means of assigning the notion similarity to the thinking features.

This study benefits from a prior knowledge (initial concept set obtained from the text mining algorithm) to create progressive solutions for the fine concept/relation set that can constitute the ontology. Proposed ontology learning strategy is applicable on different languages; it can be utilized to extract the most appropriate ontology structure from the non-Arabic texts. The proposed algorithm extracts standards and their semantic members of the family that constitute the ontology from every record of Arabic text, in three steps: term weighting, notion similarity weights, and feature mapping. Genetic algorithms (GA) were embedded into the WOA algorithm in order to improve a wide variety of whales (search agents) in the form of chromosomes.

The evaluation to the model was composed of three experiments: (i) comparisons with different bio-inspired optimization algorithms existing in the literature involving Arabic ontology learning, (ii) comparisons with previous published approaches on Arabic ontology gaining knowledge from text, and (ii) comparisons with modern day on gaining knowledge of ontology from non-Arabic settings. Eventually, the proposed ontology getting to know strategy is relevant to the non-Arabic texts too. It achieved higher performance that outperformed the contemporary processes on gaining knowledge of ontology from Arabic and non-Arabic text [36].

6 DISCUSSION

Whale Optimization Algorithm (WOA) is a recent swarm intelligence based meta - heuristic optimization algorithm, which simulates the natural behavior of bubble - net hunting technique of humpback whales and has been correctly applied to clear up complicated optimization troubles in a huge variety of disciplines. Therefore, when applied to large size issues, its efficiency and performance degrades due to the huge computational work load required. Distributed computing is one of the effective ways to improve the scalability of WOA for resolving large - scale issues. Whale swarm algorithm can be applied to solve nearly any optimization issues. Whale Optimization Algorithm (WOA) is one of the newly proposed algorithms belonging to the class of swarm intelligence. The humpback whale is simulated so as to find optimum solutions to various optimization issues. There are nonetheless some shortcomings that are handy to fall into nearby top of the line or slow convergence, which want to be always improved and innovated. There are many complex optimization problems such as clustering, Hadoop MapReduce, Dynamic software rejuvenation in web…etc. So, in the future, we will improve the whale swarm algorithm to solve the previous issues.

7 CONCLUSIONS

New swarm intelligence based meta-heuristic called Whale Swarm Algorithm, inspired using the whales' behavior of communicating with every other through ultrasound for hunting, is explained in this paper. We showed the methodology and strategy of whale swarm algorithm in detail. Finally, we explained several uses of WSA in many problems such as Clustering, Prediction Diseases, Mechanical, Text Mining and Production Engineering. Semantic understanding of textual knowledge to find concepts of thought could be detected by the use of genetics and WOA. This will provide a new layer of knowledge understanding through an optimization mechanism for big data analytics. It is still required to find means of advanced processing of such optimization technique to enhance the performance of WOA and text mining of huge amount of documents.

REFERENCES

[1] S.M. Khalessizadeh, R.Zaefarian, World Academy of Science, Engineering and Technology, 2006, "Genetic Mining: Using Genetic Algorithm for Topic based on Concept Distribution"

[2] Bing Zeng. Liang Gao. Xinyu Li. 2017. Whale swarm algorithm for function optimization. LNCS, 10361:624-639

[3] M. Mahi, Ö.K. Baykan, H. Kodaz, A new hybrid method based on Particle Swarm Optimization, Ant Colony Optimization and 3-Opt algorithms for Traveling Salesman Problem, Applied Soft Computing, 30 (2015) 484-490.

[4] K.C. Tan, Y. Chew, L.H. Lee, A hybrid multi-objective evolutionary algorithm for solving truck and trailer vehicle routing problems, European Journal of Operational Research, 172 (2006) 855-885.

[5] S.N. Qasem, S.M. Shamsuddin, S.Z.M. Hashim, M. Darus, E. AlShammari, Memetic multi objective particle swarm optimization-based radial basis function network for classification problems, Information Sciences, 239 (2013) 165-190.

[6] B. Zeng, Y. Dong, An improved harmony search based energy efficient routing algorithm for wireless sensor networks, Applied Soft Computing, 41 (2016) 135-147.

[7] E.S. Hou, N. Ansari, H. Ren, A genetic algorithm for multiprocessor scheduling, Parallel and Distributed Systems, IEEE Transactions on, 5 (1994) 113-120.

[8] B.-Y. Qu, P. Suganthan, S. Das, A distance-based locally informed particle swarm model for multimodal optimization, Evolutionary Computation, IEEE Transactions on, 17 (2013) 387-402.

[9] Nasiri, J., & Khiyabani, F. M. (2018). A whale optimization algorithm (WOA) approach for clustering. Cogent Mathematics & Statistics, 5(1), 1483565.

[10] Mohit Verma, Amit Kumar," A Brief Review of Applications of Whale Optimization Algorithm to Mechanical and Production Engineering", in: International Journal of Pure and Applied Mathematics, 119(18), pp. 1953-1960, 2018.

[11] Zhang, X., Liu, Z., Miao, Q. and Wang, L., Bearing fault diagnosis using a whale optimization algorithm-optimized orthogonal matching pursuit with a combined time– frequency atom dictionary, Mech. Syst. Signal Process., vol. 107, pp. 29–42.
[12] Mirjalili, S. and Lewis, A., 2016. The Whale Optimization Algorithm, Adv. Eng. Softw., vol. 95, pp. 51–67.
[13] Rohani and Mohammad. 2016. The Workflow Planning of Construction Sites using whale optimization algorithm (WOA), Turkish Online Journal of Design Art and Communication 6: 2938-2950.
[14] Prakash, D.B. and Lakshminarayana, C., 2016. Optimal siting of capacitors in radial distribution network using Whale Optimization Algorithm. Alexandria Engineering Journal.
[15] Bentouati, B., Chaib, L. and Chettih, L., 2016. A hybrid whale algorithm and pattern search technique for optimal power flow problem, Modelling, Identification and Control ICMIC 8th International Conference on. IEEE 2016.
[16] Buch, and Hitartch., 2016. Combined Emission Constrained Economic Dispatch with Valve Point Effect Loading Problem Solution using Whale Optimization Algorithm.
[17] Trivedi, Indrajit N., 2016. An emission constraint environment dispatch problem solution with microgrid using Whale Optimization Algorithm, Power Systems Conference (NPSC), 2016 National IEEE, 2016.
[18] Jangir, P. and Jangir, N., 2017. Non-Dominated Sorting Whale Optimization Algorithm (NSWOA): A Multi-Objective Optimization Algorithm for Solving Engineering Design Problems, Global Journal of Researches in Engineering: F Electrical and Electronics Engineering Volume 17 Issue 4 Version 1.0.
[19] Faseela, C.K. and Vennila, H., 2018. Economic and Emission Dispatch using Whale Optimization Algorithm (WOA), 2018 in IJECE (Vol 8, No 3). 2018.
[20] Horng. andMong, Fong., 2016. A Multi-Objective Optimal Vehicle Fuel Consumption Based on Whale Optimization Algorithm, Advances in Intelligent Information Hiding and Multimedia Signal Processing: Proceeding of the Twelfth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2016, Kaohsiung, Taiwan, (Volume 2. Springer International Publishing). 2017.
[21] Dao, Thi-Kien, Tien-Szu Pan, and Jeng-Shyang Pan., 2016 A multi-objective optimal mobile robot path planning based on whale optimization algorithm. Signal Processing (ICSP), 2016 IEEE 13th International Conference on. IEEE, 2016.
[22] Yan, Z., Sha, J., Liu, B., Tian, W. and Lu, J., 2018. An Ameliorative Whale Optimization Algorithm for Multi-Objective Optimal Allocation of Water Resources in Handan, China, water, 10,87.
[23] Marler, R. T. and Arora, J.S., 2004. Survey of multi-objective optimization methods for engineering, Structural and Multidisciplinary Optimization, vol. 26, no. 6. pp. 369–395.
[24] Kalaipriyan, T., Amudhavel, J. and Sujatha, P., 2017. Whale Optimization Algorithm for combined heat and power economic dispatch, Advances and Applications in Mathematical Sciences Volume 17, Issue.
[25] Jie Cai. Jiawei Luo. Shulin Wang. Sheng Yang. 2018. Feature selection in machine learning: A new perspective. Neurocomputing, 300:70–79.
[26] Anuradha, P. and Dr. Vasantha Kalyani David," International Journal of Research and Analytical Reviews (IJRAR), 6(2), 2018.
[27] Elhag, A., & Ozcan, E. (2018). Data clustering using grouping hyper-heuristics. Evolutionary Computation in Combinatorial Optimization, LNCS, 10782, 101–115.
[28] Zhang, C., Ouyang, D., & Ning, J. (2010). An artificial bee colony approach for clustering. Experiments System Applications, 37, 4761–4767. doi: 10.1016/j. eswa.2009.11.003
[29] Mirjalili, S., & Lewi, A. (2016). The whale optimization algorithm. Advancement Engineering Softwares, 95, 51–67. doi:10.1016/j.advengsoft.2016.01.008
[30] Kenedy, J., & Eberhart, R. (1995). Particle swarm optimization. Proceeding of the 1995 IEEE international conference on neural network, 194–208.
[31] Storn, R., & Price, K. (1997). Differential evolution- a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization,11, 341–359. doi:10.1023/A:1008202821328
[32] Rashedi, E., Nezamabadi- Pour, H., & Saryazdi, S. (2009). GSA: A gravitational search algoritm. Information Sciences, 179, 2232–2248. doi:10.1016/j.ins.2009.03.004
[33] Yao, X., & Liu, Y. (1999). Evolutionary programing made faster. IEEE Transactions Evolution Computer, 3, 82– 102. doi:10.1109/4235.771163
[34] Jhila Nasiri and Farzin Modarres Khiyabani," A whale optimization algorithm (WOA) approach for clustering", in: APPLIED & INTERDISCIPLINARY MATHEMATICS, 2018.

[35] Mezghanni, I.B.; Gargouri, F. CrimAr: A Criminal Arabic Ontology for a Benchmark Based Evaluation. Procedia Comput. Sci. 2017, 112, 653–662. [CrossRef]
[36] Rania M. Ghoniem , Nawal Alhelwa and Khaled Shaalan," A Novel Hybrid Genetic-Whale Optimization Model for Ontology Learning from Arabic Text", In: the journal of algorithms,12(9),2019.

Creating and Implementing ArSL Corpus for Deaf Drivers

Samah A. Abbas[*1], Hassanin M. Al-Barhamtoshy [**2], Fahad M. Al-Otaibi [***3]

[*]*Management Information Systems Department, Faculty of Economics and Administration, King Abdulaziz University*
Jeddah, Saudi Arabia
[1]sabbas@kau.edu.sa

[**]*Information Technology Department, Faculty of Computing & Information Technology, King Abdulaziz University*
Jeddah, Saudi Arabia
[2]hassanin@kau.edu.sa

[***]*Information System Department, Faculty of Computing & Information Technology, King Abdulaziz University*
Jeddah, Saudi Arabia
[3]fmmalotaibi@kau.edu.sa

Abstract— **Sign languages are a general language that deaf people around the world use it for communication with others. However, normal people usually do not know the sign language and they do not have to learn their language for communicating with them in daily life. For supporting deaf people and facilitating their work, many technologies open up possibilities for overcoming such barriers, particularly through natural language processing (NLP), text understanding, machine translation, and sign language simulation. In this paper, we focus on the problem that faced the deaf community in Saudi Arabia as an important member of society who needs support in communicating with others, particularly in the field of work as a driver. Where they need a system that facilitates the process of communicating with passengers using NLP that helps translate Arabic sign language (ArSL) into voice and vice versa. In this paper, we discuss the linguistic background and automated ArSL. In addition, we provide a brief look at the studies that have used various techniques to identify the sign language in order to translate it into sound and vice versa. Moreover, we illustrate our corpus, data determination (deaf driver terminologies) and dataset creation and processing in order to implement our future proposed system. Therefore, the dataset evaluation will be presented and simulated.**

Keywords: **Arabic Sign Language, Speech Recognition, Sign Language Recognition, Natural Language Processing, Deaf Driver in Saudi Arabia.**

1 INTRODUCTION

Deaf people use sign language to communicate with their peers and normal people who know their sign language. Thus, sign language is the only way for communication with deaf although it is still getting less attention from normal people. Moreover, sign language is not a unified language between deaf around the world, where each country has its own sign language. For instance, American Sign Language (ASL), Indian Sign Language (ISL), Australian Sign language (ASL), British Sign Language (BSL) and Arabic Sign Language (ArSL). In particular, Arab countries who use ArSL like Gulf, Al-Sham, and some Arab countries in North Africa have some similarities and differences in sign language, the reason behind that is the differences in dialects. Therefore, deaf people face some communication problems with the community in many aspects while they are working or practicing their daily lives, such as health, education, and transportation. One of the solutions for these problems is using a sign language interpreter, which is a person that knows the sign language and can interpret it to normal people. However, this way is not an optimal way because of loss of privacy and working independence [1], [2]. So, researchers developed some technologies for providing a computer interpreter instead of using a human. For example, sign language recognition technique, machine translation (MT). In addition, there are some researchers and Arabic organizations like The Arab League Educational, Cultural and Scientific Organization (ALECSO) provided their effort for unifying the ArSL by introducing the first dictionary in 1999 [3].

In Saudi Arabia, deaf people have some difficulties in communicating with others while they are driving a vehicle like a car and the deaf or normal person sitting as a passenger. They cannot use Arabic Sign Language (ArSL) at that time, and even if they can use ArSL, the normal passenger may not understand it. Also, a normal passenger cannot describe the needed location for the deaf driver. However, there are many mobile applications that can be used to facilitate communication such as, Tawasol, and Turjuman but still not translating in real-time [4], [5]. Moreover, normal passengers or drivers do not necessarily download the application of sign language translation just for using one time. In the deaf driver domain, Saudi Arabia faced a lack of technologies that can improve the communication between deaf drivers and passengers. Based on our knowledge, there is no research done yet that can introduce a solution in the deaf drivers' domain in Saudi Arabia.

This paper aims to overview the ArSL recognition and MT. In addition, we present our corpus (texts and videos) in the deaf driver domain. This paper is organized as follows: The second section illustrates the work related to ArSL and Speech Recognition Systems and the Machine Translation (MT) systems. The third section illustrates the ArSL linguistic background where the fourth is about ArSL design architecture. In the fifth section, we discuss our data collection processing and modules along videos of deaf driver corpus evaluation that required implementing the ArSL system in the context of deaf driving. Finally, we illustrate the future research directions.

2 LITERATURE REVIEW OF ARSL RECOGNITION AND MT SYSTEM

This literature review investigates the different techniques for ArSL recognition and speech recognition. Also, it focuses on the researches that illustrate some different systems by using a different method for translation from text or speech to ArSL and from ArSL to text or voice in Saudi Arabia.

In terms of ArSL recognition researches, many researchers conducted several experiments using different methods for translating ArSL to text or voice. Some of them focused on implementing systems for real-time communication [6]–[18]. In contrast, others dedicated their efforts to develop the best recognition system but not in real-time. They used several techniques of the ArSL recognition system to achieve a high percentage of accuracy [19]–[26]. For instance, Support Vector Machine (SVM) with a camera that gained 98.8 % [19] Principal Component Analysis (PCA), and Hidden Markov Models that achieved 99.9% accuracy [24]. The techniques used for ArSL recognition by some researchers divided into two methods. The first method is a wearable-based, which builds based on special equipment, such as smart gloves or glasses. This method supports real-time communication [7], [10], [17], [18], [27]. The second method is the vision-based (sometimes called a camera, image or video-based). This method is implemented based on recognition of hand gestures by a camera that captures the images and then segments to analyze (preprocessing). After that, some feature vectors are extracted for classifying in order to build the final model and measure the accuracy. In the classification stage, different classifications are used by researches for building their experiments. Some researchers focused their experiments on one classifier, for instance, Hidden Markov Models (HMM), Microsoft Kinect Device (sensor), a different method of the neural network, or a colored image [6], [12]–[14], [22], [24], [25]. While the researchers of [8], [9], [19]-[21], [23], [26] used two or more classifiers for ArSL recognition. Some of these researchers, which are [11], [15] implemented their system on mobile devices. All of these studies recorded different degrees of accuracy, where the goal of each study was to use the most method that achieves higher accuracy as a key performance indicator (KPI).

In the translation system from text or speech to ArSL, researchers conducted their experiments in building the translation system based on MT. Where the output of this system can be either video or virtual animation called the avatar [28]. One of the researchers used both video and avatar as ArSL output in order to compare them based on WER [29].

During translation, the translation system from the Arabic text is very complex due to the writing rules. The natural Arabic text should consider the grammatical rule, syntax, and semantic. Also, these considerations can support converting the text into the right meaning. Researches of [29]–[33] used various methods for analyzing the text. For example, they used semantic, pragmatic, syntactic, and morphological analysis.

To sum up the previous related work, we can say that the Arab deaf community, who use ArSL, has been the center of interest for researchers. The researchers dedicated their effort in improving the performance of recognition and building a correct system. The reason behind that is to facilitate communication with the deaf in the working environment and also in their daily life.

As we saw previously, there is a lack of research that supports and enhances the method of communication between the deaf driver or passenger. Whereas, a lot of research done focused on proposing the best techniques or methods in many sectors, such as education and health. In our proposed system, we intend to choose the best method or technique for recognizing ArSL to speech by using a video camera. Also, we intend to choose the best method or technique for recognition of speech and translation of the text into ArSL.

3 ArSL and Linguistic Background

ArSL has massive complexities in phonology, morphology, and structure, which is not like other sign languages. These complicities are explained below.

G. Phonology of ArSL

The "‘phonemes" are mental representations, which is just a way to empty what's inside the brain. The phonemes consist of four elements: 1) shape of hand. 2) Orientation of hand. 3) Position of hand from the body. 4) Direction of hand while moving [34]. In phonology, these four elements known as Manual Features (MFs). Particularly in Sign Language there are MFs and also Non-Manual Features (NMFs) are involved. The (NMFs) refers to the emotional parts of the body, for instance, lip motion, facial expression, shoulder, head, eyelids and eyebrows movement. Usually in ArSL , we use both MFs and NMFs for giving the correct meaning, called essential NMF. Where if the signer uses just MFs, the meaning will change to another meaning that the signer did not mean [35].

H. Morphology and Structure of ArSL

The grammar rules of ArSL are not the same as grammar rules of the Arabic language. The differences are in the following points: verb tenses, singular and plural differences, prepositional and adverb rules and gender signs. In terms of the sentence structure, ArSL just uses Subject-Verb-Object (SVO) instead of structure SVO, OVS and VOS [30].

4 ArSL and Design Architecture

New technologies that support communication have a significant impact on human life. For deaf people, the developers and researchers tried to use some new technologies to facilitate deaf life by developing some automated systems that can support them in communication in different aspects of their life with others. In this section, we will illustrate the brief explanation of some techniques used in order to implement the automated system for better communication between the community and amongst themselves.

A. Machine Translation

Machine Translation (MT) is a standardized name for the system that builds based on the computer analysis in order to translate between two natural languages. It is used for both text and speech using Natural Language Processing (NLP) and Artificial Intelligence (AI). For example, translating from the source "English" into the target "Arabic" [36], [37]. Also, MT is used for translating text or speech to avatar or video sign language. There are several approaches in MT that can be used based on what we need to translate, and what is a better translation. One of the approaches is Direct-Based translation without considering any grammar rules. For improving the quality of MT, the Rule-Based approach introduced, which built to analyze the syntactic parsing for both source and target language. Another approach is Corpus-Based, which deals with massive data that contain sentences. Also, there exists Knowledge-Based approaches, which consider understanding both target and source text within linguistic and semantic knowledge. Lastly is the google translation which is developed by Google [37].

B. Arabic Speech Recognition

Speech Recognition (SR) is a computerized system that converts the speech into text or sign. This system is used to communicate between humans and machines. It is also known as automatic speech recognition (ASR). SR by machine is a complicated task because of the differences in dialects, contexts and speech styles. For reducing this complexity in SR, the system can exploit the repetition and speech signal structure of the token as multiple sources of knowledge. The sources of SR are built based on knowledge of phonology, syntax, phonemes, grammar and phrases [38], [39]. In addition, the SR system has multiple classifiers. 1) The utterance of speech, which is composed of separate, connected, continued and spontaneous speech. 2) The speaker model, which contains one of both dependently that designed for a specific speaker or independently for different speakers. 3) The size of vocabulary [40].

In terms of the SR process, there are four stages that SR can be implemented in. Analyze the signal speech and then extract the feature by using different techniques for identifying the vector, like MFCC (Mel-frequency cepstral coefficient). After that, we build a model using different techniques like HMMs with the training dataset. In the last stage, we test the model within matching, taking the dissection and measuring the performance based on the error rate [41], [42] as shown in Fig. 1.

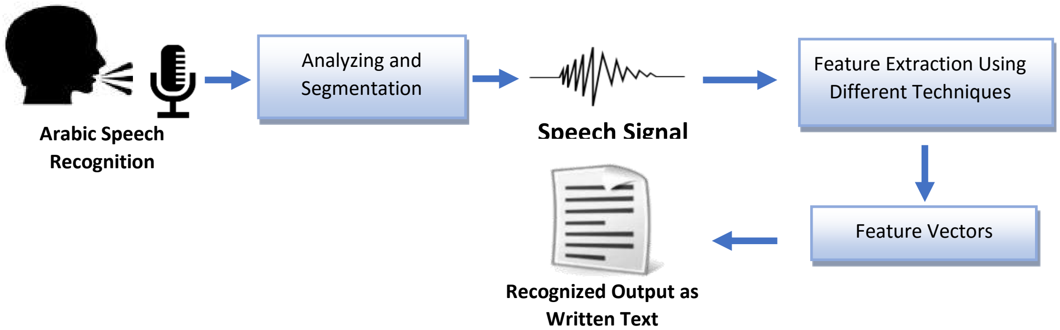


Figure 1: Block diagram to recognize spoken to Arabic written text

C. ArSL Gestures Recognitions

Gesture recognition is defined as the ability of the computer for understanding the gestures and executing the commands based on the performed gestures. The first gesture recognition system was developed in 1993 as a kind of user interface for perceptual computing that helps for capturing human gestures and then transfer it into commands by using a computerized system. It is used with many technologies, particularly in the games field, such as X-box, PlayStation, and Wii Fit. These games use Just Dance and Kinect Sports, which recognizes the hand and some parts of the body [43], [44].

In the sign language field, the gesture recognition system uses the following processes: 1) Recognize the deaf signs. 2) Analyzes the sign. 3) Converts this sign into the meaningful text (word or sentences), voice or expressions that non-deaf can understand. Moreover, there are two main methods for ArSL gesture recognition, which are wearable-based devices and vision-based devices (video or image based). Each of them have their advantages and disadvantages. One of the advantages of wearable-based does not need to look for changes in the background and lighting. The disadvantage is impeding movement. In contrast, the vision-based advantage is easy to move whereas the disadvantage is the effect of changing the background and lighting [45] [46].

Also, each of them has different processes and techniques. In the Vision-based process, one or more than one camera is the main tool that should be available for using this method. On the other hand, wearable-based method is dependent on some types of equipment and computers mainly. In terms of their processes, wearable-based method spells the alphabet by reading the particular information in each finger joint sensor or glove sensor. However, the vision-based has some stages, which are the following:

- The image capturing, using a camera for collecting data (building the corpus) and analyzing the collected images.
- The preprocessing, which starts to prepare the images and identify the information based on the color (segmentation).
- The feature extraction using some techniques like Root Mean Square (RMS) in order to identify the feature vector.
- The classification which classifies based on the feature vector in order to build the model [47] [48]. The process of vision-based gesture recognition is shown in the Fig.2.

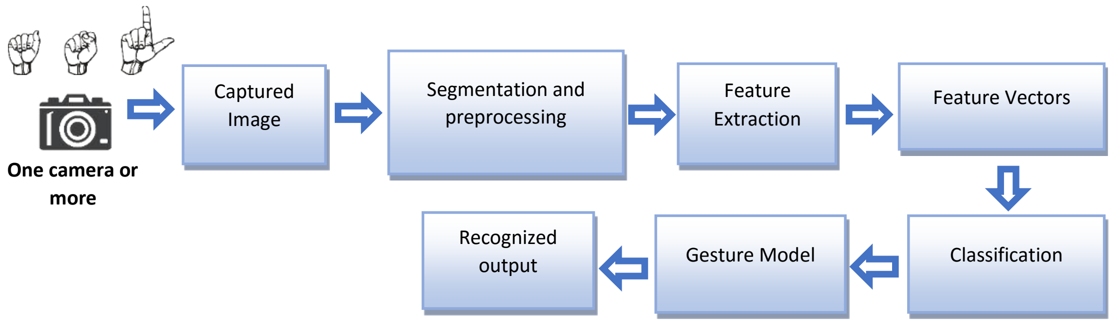


Figure2: Block diagram of the ArSL gesture recognition processing Adapted from [48]

5 DEAF DRIVER CORPUS

In order to create deaf driver corpus, we divided the processes that we will use for this creation into 4 modules. Which are preprocessing, recording, assessment and validation module. These high-level and low-level approaches are represented in Fig.3 and Fig.4.

Figure3: High-level approach of our dataset (corpus) collection

A. Data Collection and Creation

In this section, we explain our data collection, data determination (deaf driver terminologies) and dataset creation and processing in order to create our video corpus through the preprocessing and recording module.

In the preprocessing module, first, we gathered the data based on two level: a word or phrase-level dataset and a sentence-level dataset in order to create a small Arabic dictionary. This dictionary is divided into eight sections (categories). 1) Welcoming "Salam Alaikum [السلام عليكم] and How are you? [كيف حالك؟]". 2) Directions "Left [يسار] and right [يمين]". 3) Place "School [مدرسة] and Deaf Association [جمعية الصم]". 4) Traffic and Transportation "Driving License [رخصة سير] and Traffic light [إشارة ضوئية]". 5) Sentences that are used by deaf drivers when they need to talk with their passengers, for example, "We have arrived [لقد وصلنا]" 6) Sentences that are used by passengers when they need to talk with their deaf drivers, for example, "I do not have cash to pay the amount [لا أملك كاش لدفع المبلغ]". 7) General Words " No and Yes [نعم و لا] and In and On [في و على]". 8) Amount "Dollar [دولار], Riyal [ريال] and 2 Riyals until 100 Riyals [٢ ريال الى ١٠٠ ريال]".

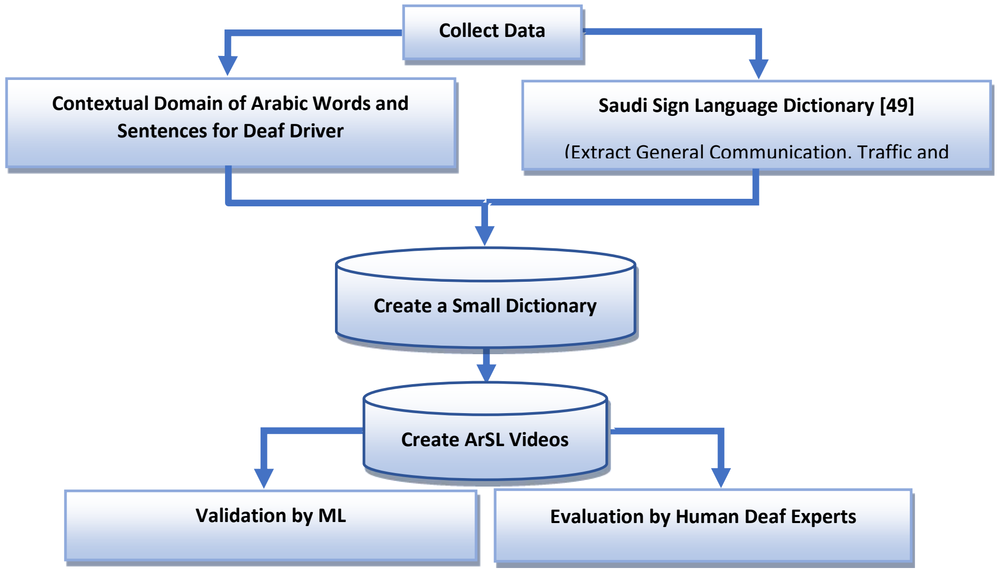


Figure 4: Low-level approach of our dataset (corpus) collection

The total words and sentences are 215. Some of these words and sentences are of the general communication, collected from 2018 edition of the Saudi Sign Language Dictionary [49] and some of them are collected from the contextual domain of the normal conversation between the taxi driver and their passengers. The flowing Table ١ shows part of our created Arabic dictionary.

TABLE IX
AN EXAMPLE OF OUR CREATED DICTIONARY

Section 1: Welcome	**القسم ١: الترحيب**
Word / Phrase / Sentences	**الجملة \ الكلمة \ المصطلح**
1. Salam Aleikum (peace be upon you)	١. السلام عليكم
2. How are you?	٢. كيف حالك؟
Section 2: Directions	**القسم ٢: الاتجاهات**
1. Left	١. يسار
2. Right	٢. يمين
Section 3: Place	**القسم ٣: الأماكن**
1. School	١. مدرسة
2. Deaf Association	٢. جمعية الصم
Section 4: Traffic and Transport	**القسم ٤: حركة المرور والنقل**
1. Driving License	١. رخصة سير
Section 5: Driver	**القسم ٥: السائق**
1. We have arrived	١. لقد وصلنا
Section 6: Passenger	**القسم ٦: الراكب**
1. I do not have cash to pay the amount	١. لا أملك كاش لدفع المبلغ

In the recording module, we did our videos at a rate of 29.97 FPS using one camera for our ArSL corpus (video capturing) with one of the expert's signer in ArSL. For recording this corpus, we took approximately 20 minutes continuously, where the total of our corpus is 215 words including sentences and signs. The expert signer tried to use one hand to be suitable with the deaf driver context unless if the sign required it to be necessary to use both hands. Then, we did video segmentation by video editing VEGAS (segment every single sign for word or sentences of our dictionary to one video where the total videos are 215). For supporting our future work, we added an Arabic audio and labeled each video with Arabic text that refers to the same ArSL that we recorded. Fig.5 represents one of our captured corpus.

Figure 5: One Captured Video from our ArSL Video Corpus

B. Dataset Evaluation

These collected and created data will be used for evaluation and validation of our corpus. In this section, we explained the evaluation module and the validation will be made in the future work.

In the assessment module, we evaluated each generated corpus' video based on our created dictionary using the human expert evaluation technique. The number of expert participants in ArSL was four and two of them were deaf. We used the quantitative approach (questionnaire) and we divided it into two sections. The first section was a demographic questionnaire (gender - age- education level - if he/she is deaf or not). The second section was a video evaluation based on the related word (phrase) or sentence. Each video attached to each related word or sentence. We asked the participants to evaluate the 215 sign videos if the video for each word or sentence is correct or not. If it is not correct, which means that the video is not related to that particular word or sentence. We asked the four participants to choose one of the correction types that they were supposed to do for each video (adding - replacing - deleting). The way of evaluation is explained in Table 2 using some videos evaluated by one of the experts.

TABLE X
AN EXAMPLE OF THE WAY OF EVALUATION WITH ONE EXPERT

Section One: Welcom					**القسم الاول: الترحيب**
Correction if the translation is wrong: The video of sign needs.			**Video evaluation based on the related word (phrase) or sentences.** √ أو ×	**Word \ phrase \ sentences**	**الكلمة \ المصطلح \ الجملة**
Deleting	**Replacing**	**Adding**			
			√	1- Salam aleikum	١-السلام عليكم
			√	2- Waleikum salam	٢-عليكم السلام
	√		×	3- Good morning	٣-صباح الخير

The deaf experts' evaluations results based on Word Error Rate (WER) for each category (section) like welcome, directions and place is shown in Fig.6. It means that for each category the result of the evaluation of these videos was wrong. Also, they must be replaced with the correct videos.

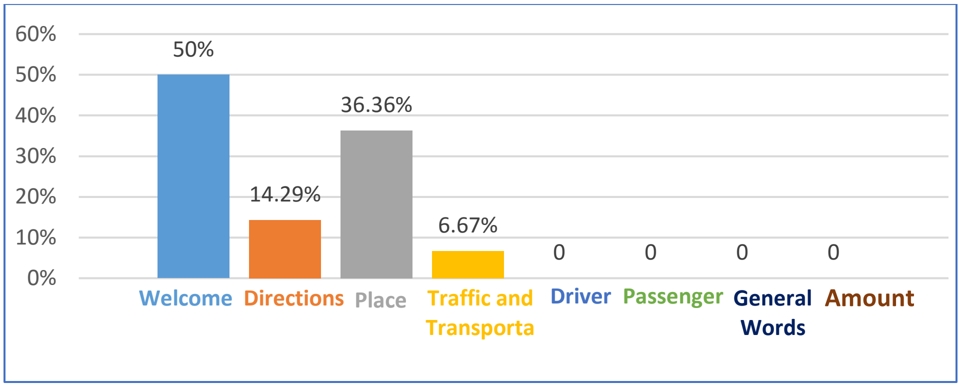


Figure 6: The Deaf Experts' Evaluations Results Based on (WER) for Each Category

As we can see in graph 6, first, the welcoming category has a high percentage of WER which is 50%. After that, the place category has approximately 36%. Thus, we need to reduce the WER in our video corpus that is related to these two categories (sections) for enhancing the communication between deaf drivers and passengers and also in order to describe the correct place for passenger's destination.

The total WER of our video corpus was 10.23% as shown in Table 3. For solving the wrong ArSL videos corpus, we re-captured these videos again based on the corrected signs that the evaluators explained for us.

TABLE III

TOTAL WORD CORRECT AND ERROR RATE FOR OUR CREATED VIDEO CORPUS

	Correct	**Wrong**
Signs' Videos Corpus	89.8%	10.2%

In the validation module, in the future work, we are going to implement ML technique using python as a programming language. For doing that, we recorded other videos with different signers to divide our data into training and testing datasets in order to measure the accuracy and error rate.

6 CONCLUSIONS AND FEATURE WORK

We have reviewed, through this research, previous studies that have been completed in the field of ArSL and speech recognition and also the translation system that converts ArSL into text and speech or vice versa. We also reviewed used methods that seek to achieve better performance while reducing the error rate in translation. We clarified the difficulties faced by translating ArSL from grammatical, semantic, and syntax. How they affect the accuracy of the translation and recognition. Finally, we described the ArSL dictionary for deaf drivers and we explained the data collection processes to construct our corpus videos. It was recorded by using one camera and then verified with 4 participants, experts in ArSL, of whom two were deaf.

The created ArSL corpus offers possibilities for testing various feature extraction methods and recognition techniques. The dataset extension and validation using Machine Learning (ML) will be implemented in future work. In addition to that, this corpus is going to be used to design our proposed system that facilitates communication with the deaf driver.

ACKNOWLEDGMENT

This work was supported by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah. The authors, therefore, gratefully acknowledge the DSR technical and financial support.

REFERENCES

[1] L. H. Forestal, "A study of deaf leaders' attitudes towards sign language interpreters and interpreting," New York University, 2001.

[2] B. Broecker, "Speaking the Language of Sign: The Art and Science of Signing," *American Annals of the Deaf*, vol. 131, no. 3, pp. 199-200, 1986.

[3] M. Al-Binali and S. Samareen, "Grammar of the Unified Qatari Arabic Sign Language," *Dar Al-Sharq, Doha, Qata*r, In Arabic, 2009.

[4] A. Al-Nafjan, B. Al-Arifi, and A. Al-Wabil, “Design and Development of an Educational Arabic Sign Language Mobile Application: Collective Impact with Tawasol,” in *Universal Access in Human-Computer Interaction. Access to Interaction*, 2015, pp. 319–326, doi: 10.1007/978-3-319-20681-3_30.

[5] Team Mind Rockets, "Mind Rockets Inc, Assistive Technologies for the deaf", Mindrocketsinc.com, 2017. [Online]. Available: http://mindrocketsinc.com. [Accessed: 08-Nov-2018].

[6] T. Aujeszky and M. Eid, "A gesture recogintion architecture for Arabic sign language communication system," *Multimedia Tools and Applications*, vol. 75, no. 14, pp. 8493-8511, 2016.

[7] M. A. Mohandes, "Recognition of two-handed Arabic signs using the CyberGlove," *Arabian Journal for Science and Engineering*, vol. 38, no. 3, pp. 669-677, 2013.

[8] M. F. Tolba, A. Samir, and M. Aboul-Ela, "Arabic sign language continuous sentences recognition using PCNN and graph matching," *Neural Computing and Applications*, vol. 23, no. 3-4, pp. 999-1010, 2013.

[9] N. B. Ibrahim, M. M. Selim, and H. H. Zayed, "An automatic arabic sign language recognition system (ArSLRS)," *Journal of King Saud University-Computer and Information Sciences*, vol. 30, no. 4, pp. 470-477, 2018.

[10] N. Tubaiz, T. Shanableh, and K. Assaleh, "Glove-based continuous Arabic sign language recognition in user-dependent mode," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 4, pp. 526-533, 2015.

[11] A. Eqab and T. Shanableh, "Android mobile app for real-time bilateral Arabic sign language translation using leap motion controller," in *2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, 2017, pp. 1-5: IEEE.

[12] E. E. Hemayed and A. S. Hassanien, "Edge-based recognizer for Arabic sign language alphabet (ArS2V-Arabic sign to voice)," in *2010 International Computer Engineering Conference (ICENCO)*, 2010, pp. 121-127: IEEE.

[13] M. Al-Rousan, K. Assaleh, and A. Tala’a, "Video-based signer-independent Arabic sign language recognition using hidden Markov models," *Applied Soft Computing*, vol. 9, no. 3, pp. 990-999, 2009.

[14] F. Guesmi, T. Bouchrika, O. Jemai, M. Zaied, and C. B. Amar, "Arabic sign language recognition system based on wavelet networks," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016, pp. 003561-003566: IEEE.

[15] F. Al Ameiri, M. J. Zemerly, and M. Al Marzouqi, "Mobile Arabic sign language," in *2011 International Conference for Internet Technology and Secured Transactions*, 2011, pp. 363-367: IEEE.

[16] D. Dahmani and S. Larabi, "User-independent system for sign language finger spelling recognition," *Journal of Visual Communication and Image Representation*, vol. 25, no. 5, pp. 1240-1250, 2014.

[17] M. Mohandes and M. Deriche, "Arabic sign language recognition by decisions fusion using Dempster-Shafer theory of evidence," in *2013 Computing, Communications and IT Applications Conference (ComComAp)*, 2013, pp. 90-94: IEEE

[18] F. N. H. Al-Nuaimy, "Design and implementation of deaf and mute people interaction system," in *2017 International Conference on Engineering and Technology (ICET)*, 2017, pp. 1-6: IEEE.

[19] H. Luqman and S. A. Mahmoud, "Transform-based Arabic sign language recognition," *Procedia Computer Science*, vol. 117, pp. 2-9, 2017.

[20] A. Tharwat, T. Gaber, A. E. Hassanien, M. K. Shahin, and B. Refaat, "Sift-based arabic sign language recognition system," in *Afro-european conference for industrial advancement*, 2015, pp. 359-370: Springer.

[21] N. A. Sarhan, Y. El-Sonbaty, and S. M. Youssef, "HMM-based arabic sign language recognition using kinect," in *2015 Tenth International Conference on Digital Information Management (ICDIM)*, 2015, pp. 169-174: IEEE.

[22] A. SamirElons, M. Abull-ela, and M. F. Tolba, "Pulse-coupled neural network feature generation model for Arabic sign language recognition," *IET Image Processing*, vol. 7, no. 9, pp. 829-836, 2013.

[23] M. Elpeltagy, M. Abdelwahab, M. E. Hussein, A. Shoukry, A. Shoala, and M. Galal, "Multi-modality-based Arabic sign language recognition," *IET Computer Vision*, vol. 12, no. 7, pp. 1031-1039, 2018.

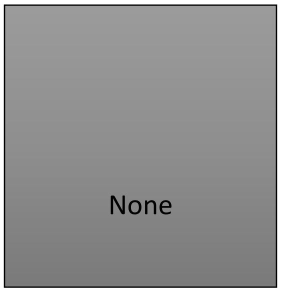
[24] A. A. Ahmed and S. Aly, "Appearance-based arabic sign language recognition using hidden markov models," in *2014 International Conference on Engineering and Technology (ICET)*, 2014, pp. 1-6: IEEE.

[25] M. ElBadawy, A. Elons, H. A. Shedeed, and M. Tolba, "Arabic sign language recognition with 3d convolutional neural networks," in *2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*, 2017, pp. 66-71: IEEE.

[26] M. Mohandes, S. Aliyu, and M. Deriche, "Arabic sign language recognition using the leap motion controller," in *2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE)*, 2014, pp. 960-965: IEEE.
[27] D. Dahmani and S. Larabi, "User-independent system for sign language finger spelling recognition," *Journal of Visual Communication and Image Representation,* vol. 25, no. 5, pp. 1240-1250, 2014.
[28] N. Aouiti, M. Jemni, and S. Semreen, "Arab gloss and implementation for Arabic Sign Language," in *2017 6th International Conference on Information and Communication Technology and Accessibility (ICTA)*, 2017, pp. 1-6: IEEE.
[29] H. M. Al-Barhamtoshy, N. E. Abuzinadah, A. 3, T. F. 4, A. A. 5 and, and A. A. Allinjawi, “Development Of An Intelligent Arabic Text Translation Model For Deaf Students Using State Of The Art Information Technology,” *Biosci. Biotechnol. Res. Commun.*, vol. 12, no. 2, pp. 338–345, Jun. 2019, doi: 10.21786/bbrc/12.2/17.
[30] H. Luqman and S. A. Mahmoud, "Automatic translation of Arabic text-to-Arabic sign language," *Universal Access in the Information Society,* vol. 18, no. 4, pp. 939-951, 2019.
[31] N. Aouiti, "Towards an automatic translation from Arabic text to sign language," in *Fourth International Conference on Information and Communication Technology and Accessibility (ICTA)*, 2013, pp. 1-4: IEEE.
[32] O. H. Al-Barahamtoshy and H. M. Al-Barhamtoshy, "Arabic Text-to-Sign (ArTTS) Model from Automatic SR System," *Procedia Computer Science,* vol. 117, pp. 304-311, 2017.
[33] A. M. Almasoud and H. S. Al-Khalifa, "Semsignwriting: A proposed semantic system for Arabic text-to-signwriting translation," *Journal of Software Engineering and Applications,* vol. 5, no. 08, p. 604, 2012.
[34] S. Ojala, "Towards an integrative information society: Studies on individuality in speech and sign," 2011.
[35] T. Johnston and A. Schembri, *Australian Sign Language (Auslan): An introduction to sign language linguistics*. Cambridge University Press, 2007.
[36] V. K. Verma, S. Srivastava, and N. Kumar, "A comprehensive review on automation of Indian sign language," in *2015 International Conference on Advances in Computer Engineering and Applications*, 2015, pp. 138-142: IEEE.
[37] I. Pradhan, S. P. Mishra, and A. K. Nayak, "A Collation of Machine Translation Approaches with Exemplified Comparison of Google and Bing Translators," Singapore, 2020, pp. 854-860: Springer Singapore.
[38] Y. Chow *et al.*, "BYBLOS: The BBN continuous speech recognition system," in *ICASSP'87. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1987, vol. 12, pp. 89-92: IEEE.
[39] A. Katyal, A. Kaur, and J. Gill, "Automatic Speech Recognition: A Review," *International Journal of Engineering and Advanced Technology (IJEAT) ISSN*, pp. 2249-8958, 2014.
[40] V. Radha and C. Vimala, "A review on speech recognition challenges and approaches," doaj. org, vol. 2, no. 1, pp. 1-7, 2012.
[41] S. K. Saksamudre and R. Deshmukh, "Isolated Word Recognition System for Hindi Language," *International Journal of Computer Science andEngineering,* vol. 3, no. 7, pp. 110-114, 2015.
[42] M. Yankayiş, T. Ensari, and N. Aydin, "Performance Evaluation of Feature Extraction and Modeling Methods for Speaker Recognition."
[43] S. Schechter, “What is gesture recognition? Gesture recognition defined,” Marxent, 24-Mar-2014. [Online]. Available: https://www.marxentlabs.com/what-is-gesture-recognition-defined/. [Accessed: 16-Jan-2020].
[44] T. Darrell and A. Pentland, "Space-time gestures," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1993, pp. 335-340: IEEE.
[45] P. Paudyal, J. Lee, A. Banerjee, and S. K. Gupta, "Dyfav: Dynamic feature selection and voting for real-time recognition of fingerspelled alphabet using wearables," in *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, 2017, pp. 457-467: ACM.
[46] R. Ambar, C. K. Fai, M. H. A. Wahab, M. M. A. Jamil, and A. A. Ma’radzi, "Development of a Wearable Device for Sign Language Recognition," in *Journal of Physics: Conference Series*, 2018, vol. 1019, no. 1, p. 012017: IOP Publishing.
[47] M. R. Mahmood and A. M. Abdulazeez, "Different Model for Hand Gesture Recognition with a Novel Line Feature Extraction," in *2019 International Conference on Advanced Science and Engineering (ICOASE)*, 2019, pp. 52-57: IEEE.

[48] K. G. Derpanis, "A review of vision-based hand gestures," *Department of Computer Science York University,* 2004.
[49] Saudi Association For Hearing Impairment, Saudi Sign Language Dectionary, Riyadh,2018.

BIOGRAPHY



Samah A. Abbas 1, received the B.A. degree in Economics and Administration, and M.Sc. degrees in Computer Science, Information System from King Abdulaziz University, Saudi Arabia. Currently, she is a Teacher Assistant of Management Information System with the Faculty of Economic and Administration, King Abdulaziz University, Jeddah, Saudi Arabia. She is an Associate member in the international conference USAR- IC 2020 as a leader of the session in ICSM 20 (Business field) in Athena. She is interested in big data, IoT, data science, machine learning and artificial intelligence.

Hassanin M. Al-Barhamtoshy , Hassanin M. Al-Barhamtoshy received the B.S. degree in electronic and communication engineering from Cairo University, and the M.S. degree in systems and computers engineering from the Azhar university, Cairo. He received the Ph.D. degree in systems and computers engineering from Azhar University, Cairo. He was a Professor in the Department of Systems and Computer Engineering at Azhar University. He is currently Professor in the Department of Computer Science and Information Technology at Faculty of Computing and Information Technology, KAU University (2003-now). His interests: language engineering, software engineering, special needs technologies, machine translation, dictionary and lexicon building, speech processing, image processing, image similarity, Sentiment Analysis, and Arabic OCR.

Fahad M. Al-Otaibi , received the B.A. and M.Sc. degrees in computer science from King Abdulaziz University, Saudi Arabia, and the Ph.D. degree in computer science and information systems from Montfort University, Leicester, U.K., in 2014. He is currently an Associate Professor with the Faculty of Computer and Information Technology, King Abdulaziz University. He is interested in smart learning, machine learning, and artificial intelligence.

إنشاء وتطبيق مجموعة لغة الإشارة العربية للسائقين الصم

[*]سماح أنور عباس، [**]الحسنين محمد البرهمتوشى، [***]فهد مزيد العتيبي

[١] قسم نظم المعلومات الإدارية، كلية الاقتصاد والإدارة، جامعة الملك عبد العزيز، جدة، المملكة العربية السعودية
[٢]قسم تقنية المعلومات، كلية الحاسبات وتقنية المعلومات، جامعة الملك عبد العزيز، جدة، المملكة العربية السعودية
[٣] قسم نظم المعلومات، كلية الحاسبات وتقنية المعلومات، جامعة الملك عبد العزيز، جدة، المملكة العربية السعودية

[1]sabbas@kau.edu.sa
[2]hassanin@kau.edu.sa
[3]fmmalotaibi@kau.edu.sa

الخلاصة

لغات الإشارة هي لغة عامة يستخدمها الأشخاص الصم حول العالم للتواصل مع الآخرين. ومع ذلك، لا يعرف الأشخاص العاديون عادة لغة الإشارة ولا يتعين عليهم تعلم لغتهم للتواصل معهم في الحياة اليومية. لدعم الصم وتسهيل عملهم، تفتح العديد من التقنيات إمكانياتها للتغلب على هذه العوائق، خاصة من خلال معالجة اللغة الطبيعية (NLP) وفهم النصوص والترجمة الآلية ومحاكاة لغة الإشارة. في هذه الورقة، نركز على المشكلة التي واجهت مجتمع الصم في المملكة العربية السعودية كعضو مهم في المجتمع يحتاج إلى دعم في التواصل مع الآخرين، لا سيما في مجال العمل كسائق. حيث يحتاجون إلى نظام يسهل عملية التواصل مع الركاب باستخدام البرمجة اللغوية العصبية التي تساعد على ترجمة لغة الإشارة العربية (ArSL) إلى صوت والعكس. في هذه الورقة، نناقش الخلفية اللغوية ولغة الإشارة الآلية. بالإضافة إلى ذلك، نقدم نظرة مختصرة على الدراسات التي استخدمت أساليب مختلفة لتحديد لغة الإشارة لترجمتها إلى صوت والعكس صحيح. علاوة على ذلك، فإننا نوضح مجموعتنا، وتحديد البيانات (مصطلحات السائق الأصم) وإنشاء مجموعة البيانات ومعالجتها من أجل تطبيق نظامنا المقترح في المستقبل. لذلك، سيتم تقديم تقييم مجموعة البيانات ومحاكاتها.

الكلمات الرئيسية

لغة الإشارة العربية، التعرف على الكلام، التعرف على لغة الإشارة، معالجة اللغة الطبيعية، سائقي الصم في المملكة العربية السعودية.

Corpus based NLP

A Critical Review of Language Resources and Tools for Arabic Sentiment Analysis

Miramar Etman[1], Sameh Alansary[2]

Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt

[1]miramar.etman@alexu.edu.eg

[2]sameh.alansary@bibalex.org

Abstract— **Sentiment Analysis field is growing expeditiously, nowadays researchers from a wide range of disciplines are devoting their time and efforts to extract and detect opinions, feelings, views and attitudes towards a specific topic, products, brands and services. With the booming number of people over the Internet where they spend much of their time expressing their opinions and reactions on social platforms, as a result researchers are heading for collecting data from social media platforms. Arabic sentiment analysis has been gaining much attention recently. However, there is insufficiently in the available language resources and tools when comparing Arabic language to English language. This paper will try to shed the light on the sentiment analysis resources and tools developed for Arabic language, so beginners to the Arabic sentiment analysis field could use this paper as a guide.**

Keywords: **Arabic Sentiment, Lexicons, Datasets, Corpora, Social Media, Dialectal Arabic, Modern Standard Arabic, Sentiment Tools.**

1 INTRODUCTION

Sentiment analysis recently has been the focus of many researchers from a wide range of backgrounds and it is now one of the most important tasks in natural language processing that everyone is seeking to work on. Sentiment analysis or opinion mining are two terms that researchers used interchangeably. However, some researchers declared the differences between them as [1], [2].

Opinion Mining extracts and analyzes people's opinion about an entity while Sentiment Analysis identifies the sentiment expressed in a text then analyzes it. Therefore, the target of SA is to find opinions, identify the sentiments they express, and then classify their polarity [1].

Unfortunately, Arabic language resources and tools are very few compared to English where there is a plenty of resources available free to use. This motivated English sentiment researches to work on sentiment analysis and the outcome results of their work were very satisfied and the progress in the field is growing rapidly and efficiently.

On the other hand, Arabic sentiment researchers devote their time to build resources and tools to help them to encounter the lack existed, but most of the resources made are not available for free and still need further improvements. This paper goal is critically reviewing the available lexicons, data sets, corpora and tools made for Arabic language.

The paper is organized as follows; Section 2 gives background information about Arabic language and challenges in sentiment analysis. Section 3 gives overview about Arabic sentiment analysis architecture, section 4 presents the available Arabic lexicons, Section 5 discusses the available Arabic datasets and corpora, section 6 gives an overview on the used tools, and finally section 7 will conclude the paper.

2 BACKGROUND ABOUT ARABIC LANGUAGE

Arabic language is one of the most used languages in the world; it is the official language of 27 countries and is spoken by more than 422 million people in the Arab world [3].

Arabic language has three different varieties; Classical Arabic (CA) which is the language of Quran (The Holy Book in Islam), Modern Standard Arabic (MSA) which is used in formal contexts (e.g. newspapers, education, television) and Dialectal Arabic (DA) which refers to the colloquial and informal Arabic used in daily communication and mostly seen on the social media platforms.

The number of Arabs on social media is growing rapidly and the amount of opinions, public views and behaviors Arabs express daily is very huge on social platforms and made many researchers head to social media data. The number of Arab Facebook users as of beginning of May 2014 is 81,302,064 up from 54,552,875 in May 2013 [4] and total number of Arab active Twitter users reached 5,797,500 users as of March 2014 [5].

Arabic language made sentiment analysis a very challenging task in terms of its morphological complexity which can be summed up in inflection, agglutination and derivation nature of Arabic, different Arabic dialects present a big challenge, even in MSA [6], Franco-Arabic or Arabic chat Alphabet which Arabs use to write both MSA and DA on social platforms have been one of the most challenging issues in Arabic sentiment Analysis, [7] worked on Franco-Arabic for Algerian Dialect and [8] worked on Franco-Arabic and DA.

3 ARABIC SENTIMENT ANALYSIS ARCHITECTURE

There are four main phases that constitute the cycle of sentiment analysis: preprocessing, feature extraction, feature selection and sentiment classification. Those four phases are similar across the languages; however the implementation of each phase will differ according to the language system. Figure 1 shows the Arabic sentiment architecture.

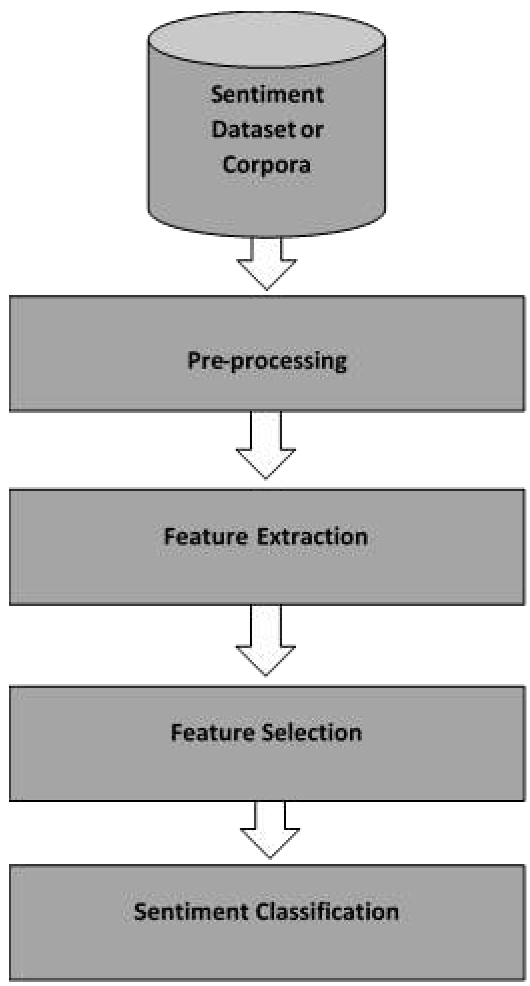


Figure 1: Arabic sentiment analysis general phases

A. Preprocessing

Data preprocessing is one of the most important steps in natural language processing tasks in general and sentiment analysis in particular. Arabic is a morphologically rich language that requires special care during the preprocessing stage [9].

Preprocessing phase includes de-noising (e.g., removing white spaces, HTML tags, special characters, emoticons and URLs), extending abbreviations, tokenization (splitting the text into words), stemming (reducing inflected or derivational words to their word stem), lemmatization (reducing the inflectional forms from each word to a common base or root), normalization (e.g. remove duplicate characters), stop-words removal and negation transformation.

B. Feature Extraction and Selection

The outputs of pre-processing are the extracted text features [10]. Feature extraction methods can be divided into two main approaches lexicon-based approaches and statistical-based approaches.

Lexicon-based approaches usually use a set of seed words or familiar hashtags when dealing with social media platforms and then expanding this seed set using synonyms, antonyms and other resources.

Statistical methods are fully automated methods, which work by extracting linguistic rules from domain-oriented corpus to detect candidate sentiment terms and structures [11].

Those selected text features mainly full under one of four aspects, the first aspect is extracting the terms presence and their frequency (e.g. words and n-grams), the second aspect is concerned with extracting the important parts of speech (e.g. adjectives), the third aspect deals with extracting lists of opinion words and phrases and the fourth aspect deals with negations.

Feature selection is used for filtering irrelevant or redundant features from all your extracted features. Generally, feature selection falls under two main approaches, filter approach and wrapper approach. Filter methods rank the features according to certain metric and select the top-ranked features. Wrapper methods, on the contrary, select the best subset of features by generation and evaluation of different subsets with a classifier [10].

A significance difference between feature extraction and selection is that feature extraction creates brand new features while feature selection keeps a subset of the original features.

C. Sentiment Classification

Sentiment classification deals with classifying a piece of text into positive, negative and neutral. Broadly sentiment classification techniques are divided into three main categories; lexicon-based approach, machine learning approach and hybrid approach.

Machine learning approach relies on machine algorithms to solve sentiment analysis task, machine learning algorithms fall under three main categories; supervised learning, semi-supervised learning and unsupervised learning. Supervised learning methods include four classifiers; decision trees classifiers, linear classifiers (e.g. Support Vector Machines (SVM) and Neural Network), rule-based classifiers, and probabilistic classifiers (e.g. Naïve Bayes, Bayesian and Maximum Entropy).

Lexicon-based Approaches are either Corpus-based approach or Dictionary-based approach. Corpus based approach fall under two methods; statistical and semantic.

Concerning the Arabic language sentiment classification, in [10] survey they declared that there is a dominance of supervised learning over other techniques (semi-supervised, unsupervised, and hybrid techniques) and widely used methods appear to be based on Support Vector Machines (SVM), Naive Bayes (NB), and K-Nearest Neighbors (KNN). Figure 2 shows sentiment classification techniques.

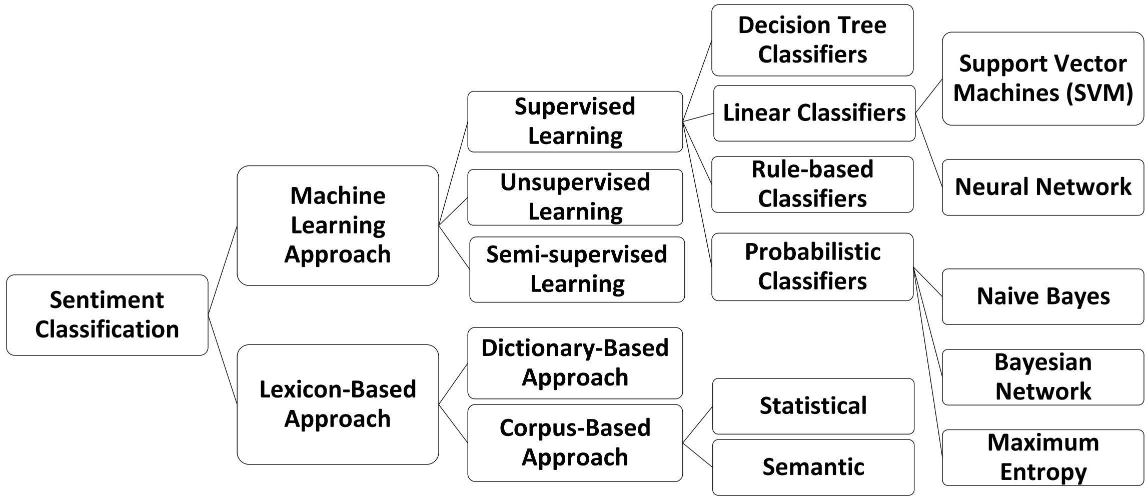


Figure 2: Sentiment classification approaches

4 ARABIC SENTIMENT LEXICONS

The Available Arabic Sentiment lexicons are either automatically or manually created, nowadays there is a trend towards the automatically created lexicons as they save both time and effort. Sentiment lexicons are lists of positive and negative words, optionally with a score indicating the degree of polarity [12].

Lexicons can be based on data collected from social media or reviews or other resources by using machine learning techniques or it can be a translated lexicon from another language (researchers create translated lexicons as a reflection of the limited Arabic resources).

Mohammed et al. [12] presented three sentiment lexicons for Arabic social media that were automatically created based on the idea that hashtag words and emoticons could act as sentiment labels for tweets; the Arabic Emoticon Lexicon was generated by collecting nearly one million tweets that had sad and happy emoticons, the Arabic Hashtag Lexicon was based on the NRC Canada System seed words which was translated into Arabic using Google Translate, the data was pulled from Twitter API and a positive seed hashtag was considered a positive label (pos) and a negative seed hashtag was considered a negative label (neg) [12], the Arabic hashtag lexicon (Dialectal) was man created using seed words taken from a previously created sentiment lexicon by [13]. And work is the best example for showing how researches could benefit from each other work when the resource is made for free use.

They also translated six English sentiment lexicons into Arabic (four of them were manually created and two were automatically created) and they used Google translation to translate the words in the English lexicons, however some words Google were unable to translate. One of the best finding of this work was the observation made about the accuracy of the automatically generated Arabic lexicons which provide accuracy similar to the manual ones around (63%) and the Arabic hashtag lexicon (Dialectal) obtained marked accuracy results (65%). All of the lexicons they presented are available for free use and their work is very valuable in Arabic sentiment analysis.

Badaro et al. [14] created the first publicity available large scale MSA lexicon and it was based on the idea of reusing existed resources to overcome the lack of Arabic resources. They used English WordNet (EWN), Arabic Word-Net (AWN), English SentiWordNet (ESWN) and Standard Arabic Morphological Analyze (SAMA). The creation was based on two approaches; Firstly, Arabic WordNet-based Approach in which a mapping between AWN to ESWN and then mapping between AWN and SAMA and named the resulting lexicon ArSenL-AWN, Secondly the English Gloss-based Approach in which they associated the SAMA lemma entries with English glosses and they named the resulting lexicon as ArSenL-Eng. The ArSenL lexicon was a result of combining the two approaches together.

Eskander & Rambow [15] Presented MSA large scale lexicon, the idea behind SLSA is linking the glosses of AraMorph to the synset terms in SentiWordNet. They also took in consideration while preparing the resources cleaning up the AraMorph especially in cases where part of speech tags was not optimal.

El-Beltagy [16] constructed a lexicon that was a turning point in the field of sentiment analysis; NileULex is an Arabic sentiment lexicon that composes both MSA and Egyptian Colloquial Arabic. Nearly 45% of the lexicon is colloquial and the rest 55% is MSA, NileULex is a phrase and a word level sentiment lexicon, Beside that the lexicon included the most common English transliteration terms and some of the most frequent misspelling words as both are heavily common on social media.

Youssef & El-Beltagy [17] Created MoArLex an Arabic lexicon for use in social media applications and they have taken a new approach in creating their lexicon, rather than creating the lexicon from scratch they extended another lexicon (NileULex [16]) as seed lexicon which showed improvement of the accuracy in sentiment analysis across various datasets. The idea of the expanded lexicon is based on finding the terms that are semantically similar to those in the original seed lexicon that can be found in some accurate manner, and then those terms will most likely be excellent candidates for addition to the lexicon. To the author's best knowledge this work is from the pioneers in expanding lexicon using automatically approaches.

From the previously mentioned work on lexicons it's apparently clear that there is a trend towards building lexicons from social media platforms based on dialectal Arabic data [12], [16], [17], [14], however some lexicons still created for MSA [14], [15]. Table I lists all of the lexicons reviewed with addition of another lexicons which were not reviewed here. Some of the lexicons in Table I are available for use (e.g. [12]).

TABLE XI
Arabic Sentiment Lexicons

Article	Lexicon	Size	Construction Approach	Source of Data	Year
[12]	Arabic Emoticon Lexicon	43,304 **Positive:** 22,962 **Negative:** 20,342	**Automatic** (Using distant supervision Techniques)	Twitter API	2016
	Arabic Hashtag Lexicon	21,964 **Positive:** 13,118 **Negative:** 8,846	**Automatic** (Using distant supervision Techniques)	Twitter API	
	Arabic Hashtag Lexicon (Dialectal)	20,128 **Positive:** 11,941 **Negative:** 8,179	**Automatic** (Using distant supervision Techniques)	Twitter API	
	English Lexicons Translated into Arabic • NRC Emoticon Lexicon • NRC Hashtag Lexicon		**Automatic**	English Lexicons	
	English Lexicons Translated into Arabic • AFINN • Bing Liu's Lexicon	2,476 **Positive:** 878 **Negative:** 1,598	**Manually**	English Lexicons	

	• NRC Emotion Lexicon • MPQA Subjectivity Lexicon	6,789 **Positive:** 2,006 **Negative:** 4,783 8,199 **Positive:** 2,718 **Negative:** 4,911 **Neutral:** 570 14,182 **Positive:** 2,317 **Negative:** 3,338 **Neutral:** 8,527			
[14]	Arabic sentiment lexicon (Ar-SenL)	**33,995**	**Automatic**	• English SentiWordnet (ESWN) • Arabic Word-Net (AWN) • English SentiWordNet (ESWN)	2014
[15]	SLSA (A Sentiment Lexicon for Standard Arabic)	**34,821**	**Automatic**	• English SentiWordnet (ESWN	2015
[16]	NileULex	**5953** **Compound Positive Phrases:** 416 **Compound Negative Phrases:** 563 **Single Negative Terms:** 369 **Single Positive Terms:** 1281	**Manually**	- Egyptian dialect dataset (NU_EG_Twitter_corpus) • Datasets is one that was collected at a research center in Saudi Arabia	2016
[17]	MoArLex	**36,775**	**Automatic** - Lexicon Expansion	• NileULex Lexicon	2017
[18]	ArSEL (Arabic Sentiment and Emotion Lexicon)	**32,196**	**Automatic**	- DepecheMood - English WordNet - ArSenL	2018
[19]	HILATSA		**semi- automatic** - Combines both lexicon based and machine learning approaches	- ASTD - Mini Arabic Sentiment Tweets Dataset (MASTD) - ArSAS - Arabic Gold Standard Twitter Data for Sentiment Analysis - Syrian Tweets Corpus	2019

				- Twitter dataset for Arabic Sentiment Analysis (ArTwitter)	
[20]	SentiRDI	18,164 **Positive:** 3,156 **Negative:** 4,169 **Neutral:** 10,839	**Automatic** - If word is not covered by semantic database, the polarity is set **Manually**	- Arabic Semantic Dtabase	2014

5 ARABIC SENTIMENT CORPORA AND DATASETS

There are a very limited available corpora and datasets that are suitable for sentiment analysis tasks. Most of the Arabic sentiment analysis researchers built their own resources as the amount and quality of the freely available are insufficient. Not all Arabic corpora available for text categorization could be used in sentiment analysis for example the Quranic corpus although it is morphologically and syntactically annotated it is not sentimentally tagged and does not provide opinions.

Nabil et al. [21] introduced an Arabic Sentiment Tweets Dataset (ASTD) that was obtained from twitter and its size about 10,000 tweets that were collected in two stages; the first stage by using SocialBakers to determine the most active Egyptian twitter account and in the second stage they used the top trending hashtags in Egypt. They annotated the data using Amazon Mechanical Turk (AMT) Service through an API called Boto and they used four-way sentiment classifications (Subjective Positive, Subjective Negative, Subjective Mixed and Objective) which is very challenging and most of the researchers focus on three-way sentiment classification so their corpus is an addition.

Itani et al. [22] developed a corpus that is based on informal Arabic or dialectal Arabic, the corpus was gathered from Facebook social media platform and they classified the posts as negative, positive, neutral, dual or spam. To the author's best knowledge there is no other work considered spam posts into account. They developed two corpora with the collected posts; the news corpus (NC) from Al Arabiyya News Facebook page and the arts corpus (AC) from the Voice Facebook Page.

Abdellaoui & Zrigui [23] used a distant supervision algorithm to automatically collect and label TEAD (Dataset for Arabic Sentiment Analysis); the data was collected from twitter using the top twenty most used emojis on Twitter according to emojitracker. They classified the tweets as positive, negative and neutral and replaced dialect words with their respective synonyms in MSA using dialect lexicons.

Saleh [24] This corpus represents one of the first sentiment analysis datasets for Arabic language; it was collected manually from movie reviews collected from different web pages and classified reviews into two categories only (positive and negative).

Abdul-Mageed & Diab [25] Created AWATIF a multi-genre MSA corpus, the corpus data relied on three different resources; Wikipedia Talk Pages, Penn Arabic Treebank (PATB) and other web forums. Corpus annotation was divided into three parts; one part was annotated using crowdsourcing on Amazon Mechanical Turk (AMT), second part was annotated by students that received instructions and last part was also annotated by students but it was simpler instructions. One advantage about AWATIF corpus is the fact that its collected from many different domains, but unfortunately this corpus is not available for free use so no other researcher could benefit from.

Aly, & Atiya [26] Built large scale corpus collected from book reviews, they used 1 to 5 score sentiment classification system. And they collected the data from www.goodreaders,com from the first 2143 books in the list of Best Arabic Book. They considered positive reviews those which have ratings 4 or 5, and negative reviews those which ratings 1 or 2. Reviews which have rating 3 are considered neutral and were not included in the polarity classification. To the author's best knowledge most of the works included neutral data in polarity classification.

Elsahar, & El-Beltagy [27] collected multi-genre Arabic reviews corpus from a wide range of domains and their work consider one of the most important works in creating multi–genre corpora. Table II shows more information about the reviewed corpora.

TABLE XIII
SHOWS ARABIC SENTIMENT CORPORA AND DATASETS

Article	Corpus/Dataset	Size	Source of Data	Arabic variety
[21]	Arabic Sentiment Tweets Dataset (ASTD)	10,000 Tweets - 793 Positive - 1684 Negative - 6691 Neutral - 832 Mixed	Twitter	Dialectal Arabic
[22]	The News Corpus (NC)	1000 Posts - 230 Negative - 236 Positive - 161 Dual - 193 Spam - 180 Neutral	Facebook	Dialectal Arabic
	The Arts corpus (AC)	1000 Posts - 224 Negative - 233 Positive - 151 Dual - 197 Spam - 195 Neutral	Facebook	Dialectal Arabic
[23]	TEAD	6 Million Tweets - 3,122,615 Positive - 2,115,325 Negative - 378,003 Neutral	Twitter	Modern Standard Arabic
[24]	OCA (Opinion Corpus for Arabic)	500 Reviews - 250 Positive - 250 Negative	Movie Reviews	Dialectal Arabic
[25]	AWATIF	2,855	Wikipedia TalkPages/Forums	MSA/ Dialectal Arabic
[26]	LABR (Large Scale Arabic Book Reviews)	63,257 (1 to 5 Scale Rate) - 1 - 2 Positive - 3 Neutral - 4 - 5 Negative	GoodReads.com	MSA/ Dialectal Arabic
[27]	Hotel Reviews (HTL)	15,527	TripAdvisor.com	MSA/ Dialectal Arabic
	Restaurant Reviews (RES)	10,970	Qaym.com	MSA/ Dialectal Arabic
	Product Reviews (PROD)	4,272	Souq.com	MSA/ Dialectal Arabic
	Movie Review (MOV)	1,524	Elcinemas.com	MSA/ Dialectal Arabic

6 ARABIC SENTIMENT ANALYSIS TOOLS

There are few tools available which researchers in Arabic sentiment use, usually those tools help in preprocessing phase or data collection. This section will present the tools used or created for Arabic.

Researchers in [21] used in the first stage of data collection SocialBakers to find out the most active twitter account Egyptian Twitter accounts. They also used Amazon Mechanical Turk (AMT) service for data annotation through Boto API.

In [28] researchers built a tweet collector tool (TCT) by using Java language, the goal behind this tool is helping researches in sentiment analysis. The tool consists of two main phases; the first phase enables the user to search for any specific hashtag or word and retrieve all the related data regardless its size, the second phase consist of extracting sentiment analysis features from dataset.

Researchers in [29] presented a web-based tool for Arabic sentiment analysis using a combination of five parameters as the time of the tweets, preprocessing techniques (e.g. stemming and retweets), n-grams features, lexicon-based methods, and machine-learning methods. The tool was created using R Language by the help of several packages as RWeka, shiny and Twitter, and to the author's best knowledge this is the first web-based tool that target Arabic.

In [30] the researcher presented SentiArabic a lexicon-based sentiment analyzer for Standard Arabic, the analyzer identifies the polarity (positive or negative) of the text. The analyser is based on an extended and modified version of the SLSA a large-scale Sentiment Lexicon for

Standard Arabic, training and evaluation of SentiArabic analyser was based on a new corpus they created for modern standard Arabic and tagged each sentence for polarity.

7 CONCLUSION

It can be noticed from the previously mentioned work that there is a number of challenges researchers' face when working on Arabic Sentiment Analysis resources and tools but the amount of resources developing is increasing rapidly. Nowadays there is trend towards creating Dialectal Arabic sentiment lexicons which is also reflected in creating dialectal corpora and datasets.

The number of studies in the last five years is concerned with collecting data from social media platforms as the amount of time Arabs spend on social media and the huge content they post about their opinions is growing rapidly.

REFERENCES

[1] Hassan Yousef, Ahmed & Medhat, Walaa & Mohamed, Hoda. (2014). Sentiment Analysis Algorithms and Applications: A Survey. Ain Shams Engineering Journal. 5. 10.1016/j.asej.2014.04.011.

[2] Tsytsarau, M., Palpanas, T. Survey on mining subjective data on the web. Data Min Knowl Disc 24, 478–514 (2012). https://doi.org/10.1007/s10618-011-0238-6.

[3] Boudad, Naaima & Faizi, Rdouan & Rachid, Oulad haj thami & Chiheb, Raddouane. (2017). Sentiment analysis in Arabic: A review of the literature. Ain Shams Engineering Journal. 9. 10.1016/j.asej.2017.04.007.

[4] Facebook in the Arab Region, Arab Social Media Report: http://www.arabsocialmediareport.com/Facebook/LineChart.aspx?&PriMenuID=18&CatID=24&mnu=Cat , (accessed 6 January 2020).

[5] Twitter in the Arab Region, Arab Social Media Report: http://www.arabsocialmediareport.com/Twitter/LineChart.aspx?&PriMenuID=18&CatID=25&mnu=Cat , (accessed 6 January 2020).

[6] Hamdi, Ali & Shaban, Khaled & Zainal, Anazida. (2016). A Review on Challenging Issues in Arabic Sentiment Analysis. Journal of Computer Science. 12. 471-481. 10.3844/jcssp.2016.471.481.

[7] Chader, Asma & Dihia, Lanasri & Hamdad, Leila & Belkheir, Mohamed & Hennoune, Wassim. (2019). Sentiment Analysis for Arabizi: Application to Algerian Dialect. 475-482. 10.5220/0008353904750482.

[8] Duwairi, Rehab & Marji, Raed & Sha'ban, Narmeen & Rushaidat, Sally. (2014). Sentiment Analysis in Arabic tweets. 2014 5th International Conference on Information and Communication Systems, ICICS 2014. 1-6. 10.1109/IACS.2014.6841964.

[9] Duwairi, Rehab. (2014). A Study of the Effects of Preprocessing Strategies on Sentiment Analysis for Arabic Text. Journal of Information Science. 40. 501-513. 10.1177/0165551514534143.

[10] Assiri, Adel & Emam, Ahmed & Aldossari, Hmood. (2015). Arabic Sentiment Analysis: A Survey. International Journal of Advanced Computer Science and Applications. 6. 10.14569/IJACSA.2015.061211.

[11] Tedmori, Sara & Awajan, Arafat. (2019). Sentiment Analysis Main Tasks and Applications: A Survey. Journal of Information Processing Systems. 15. 500-519. 10.3745/JIPS.04.0120.

[12] Mohammad, Saif & Salameh, Mohammad & Kiritchenko, Svetlana. (2016). Sentiment Lexicons for Arabic Social Media.

[13] Refaee, E., & Rieser, V. (2014). An Arabic Twitter Corpus for Subjectivity and Sentiment Analysis. In N. Calzolari (Ed.), *Proceedings of the 9th International Conference on Language Resources and Evaluation* (pp. 2268-2273). European Language Resources Association.

[14] Badaro, Gilbert & Baly, Ramy & Hajj, Hazem & Habash, Nizar & El-Hajj, Wassim. (2014). A Large Scale Arabic Sentiment Lexicon for Arabic Opinion Mining. 10.3115/v1/W14-3623.

[15] Eskander, Ramy & Rambow, Owen. (2015). SLSA: A Sentiment Lexicon for Standard Arabic. 2545-2550. 10.18653/v1/D15-1304.

[16] El-Beltagy, Samhaa. (2016). NileULex: A Phrase and Word Level Sentiment Lexicon for Egyptian and Modern Standard Arabic.

[17] Youssef, Mohab & El-Beltagy, Samhaa. (2018). MoArLex: An Arabic Sentiment Lexicon Built Through Automatic Lexicon Expansion. Procedia Computer Science. 142. 94-103. 10.1016/j.procs.2018.10.464.

[18] Badaro, Gilbert & Jundi, Hussein & Hajj, Hazem & El-Hajj, Wassim & Habash, Nizar. (2018). ArSEL: A Large Scale Arabic Sentiment and Emotion Lexicon.

[19] Elshakankery, Kariman & Farouk, Mona. (2019). HILATSA: A hybrid Incremental learning approach for Arabic tweets sentiment analysis. Egyptian Informatics Journal. 20. 10.1016/j.eij.2019.03.002.

[20] Mobarz, Hanaa & Rashown, Mohsen & Farag, Ibrahim. (2014). Using Automated Lexical Resources In Arabic Sentence Subjectivity. International Journal of Artificial Intelligence & Applications. 5. 01-14. 10.5121/ijaia.2014.5601.

[21] Nabil, Mahmoud & Aly, Mohamed & Atiya, Amir. (2015). ASTD: Arabic Sentiment Tweets Dataset. 2515-2519. 10.18653/v1/D15-1299.

[22] Itani, Maher & Roast, C.R. & Al-Khayatt, Samir. (2017). Developing Resources For Sentiment Analysis Of Informal Arabic Text In Social Media. Procedia Computer Science. 117. 129-136. 10.1016/j.procs.2017.10.101.

[23] Abdellaoui, Houssem & Zrigui, Mounir. (2018). Using Tweets and Emojis to Build TEAD: an Arabic Dataset for Sentiment Analysis. Computacion y Sistemas. 22. 10.13053/CyS-22-3-3031.

[24] Saleh, Mohammed & Martín-Valdivia, Maria & López, L. & Perea-Ortega, José. (2011). OCA: Opinion corpus for Arabic. JASIST. 62. 2045-2054. 10.1002/asi.21598.

[25] Abdul-Mageed, Muhammad & Diab, Mona. (2012). AWATIF: A Multi-Genre Corpus for Modern Standard Arabic Subjectivity and Sentiment Analysis. Proceedings of LREC.

[26] Aly, Mohamed & Atiya, Amir. (2013). LABR: A Large Scale Arabic Book Reviews Dataset. 10.13140/2.1.3960.5761.

[27] Elsahar, Hady & El-Beltagy, Samhaa. (2015). Building Large Arabic Multi-domain Resources for Sentiment Analysis. Lecture Notes in Computer Science. 9042. 23-34. 10.1007/978-3-319-18117-2_2.

[28] Abdallah, Emad & Abo-Suaileek, Sarah. (2019). Feature-based Sentiment Analysis for Slang Arabic Text.
[29] El-Masri, Mazen & Altrabsheh, Nabeela & Ahmed, Hanady & Ramsay, Allan. (2017). A web-based tool for Arabic sentiment analysis.
[30] Eskander, Ramy. “SentiArabic: A Sentiment Analyzer for Standard Arabic.” LREC (2018).

I. ACKNOWLEDGMENT

First and foremost, I would like to thank God for giving me the strength, knowledge, ability and opportunity to undertake this research. I also would like to thank my family and all my professors especially Dr. Sameh Alansary for encouraging me to take this research.

BIOGRAPHIES

Miramar Etman: *Master Degree student at Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt*

She graduated from Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, (2018), graduated from Information Technology Institute (ITI), Ministry of communications and Training, Track: Open Source Applications Development, Intake 39. Her research interests include natural language and speech processing and Programming languages.

Dr. Sameh Alansary: *Director of Arabic Computational Linguistic Center at Bibliotheca Alexandrina, Alexandria, Egypt.*

He is professor of computational linguistics in the Department of Phonetics and Linguistics and the head of Phonetics and Linguistics Department, Faculty of Arts, Alexandria University. He obtained his MA in Building Arabic Lexical Databases in 1996, and his PhD from Nijmegen University, the Netherlands in building a formal grammar for parsing Arabic structures in 2002. His main areas of interest are concerned with corpus work, morphological analysis and generation, and building formal grammars. He is also the head of Arabic Computational Linguistics Center in Bibliotheca Alexandrina; supervising and managing the Universal Networking Language project in Library of Alexandria since 1-6- 2005 till now. Dr. Alansary is the co-founder of the Arabic Language Technology Center (ALTEC), an NGO aims at providing Arabic Language resources and building a road map for Arabic Language Technology in Egypt and in the Middle East. He has many scientific works in Arabic Natural Language Processing published in international conferences and periodicals, and a member in many scientific organizations: (1) Egyptian Society of Language Engineering, Cairo, (2) Arabic Linguistic Society - USA, (3) Association of Computational Linguistics - USA – Europe, (4) Universal Networking Language foundation, United Nations, Geneva, Switzerland.

مراجعة نقدية لموارد وأدوات اللغة العربية الخاصة بتحليل الآراء والمشاعر

Miramar Etman[1], Sameh Alansary[2]

Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt

[1]miramar.etman@alexu.edu.eg

[2]sameh.alansary@bibalex.org

ملخص — ينمو مجال تحليل الآراء والمشاعر بشكل سريع في الوقت الحالي، حيث يكرس الباحثون من التخصصات المختلفة وقتهم وجهودهم لاستخراج واكتشاف الآراء والمشاعر للأشخاص اتجاه موضوع معين أو منتج أو علامة تجارية أو خدمات. ومع تزايد عدد الأشخاص على الإنترنت حيث يقضي معظم الأشخاص وقتهم في التعبير عن آرائهم وردود أفعالهم على منصات التواصل الاجتماعي، وهذا جعل الباحثون يتجهون لجمع البيانات من منصات التواصل الاجتماعي. اكتسب تحليل الآراء والمشاعر في اللغة العربية الكثير من الاهتمام مؤخرًا. ومع ذلك ، لا توجد موارد وأدوات للغة متاحة بشكل كبير عند مقارنتها باللغة الإنجليزية. ستحاول هذه الورقة إلقاء الضوء على مصادر وأدوات تحليل الآراء والمشاعر المطورة للغة العربية، لذلك يمكن للمبتدئين في مجال تحليل الآراء والمشاعر في العربية استخدام هذه الورقة كدليل.

الكلمات الدالة: المشاعر العربية ، المعاجم ، مجموعات البيانات ، المدونات اللغوية ، وسائل التواصل الاجتماعي ،العامية العربية ، العربية الفصحى ، أدوات تحليل المشاعر.

A Pilot Study of Biber's Model for Language Variation Detection: A Language Engineering Approach

Maram Elsaadany

Pharos University, Alexandria, Egypt

maram.elsaadany@pua.edu.eg

Sameh Alansary

Bibliotheca Alexandrina, Alexandria, Egypt
Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt

sameh.alansary@bibalex.org

***Abstract* — this paper is primarily a translation analysis of Tramp's speech and a letter sent to President Trump regarding family separation on "The Leadership Conference on Civil and Human Rights". Google translate is the engine used to translate these two texts into Arabic. The data selected for this analysis is 1000 words in each script. Biber's model (1988) used 67 features to prove that writing is more complicated than speech. The study enforces Biber's claim that writing is more complicated than speech. The findings assure Biber's claim as there were lots of problems in the translation of the speech text into Arabic in comparison to the translated written texts. The Results clarify the fact that google translate has to be adapted and equipped with a new grammar for speech that is different than the one used for writing to achieve the best outcome for both translations in the same language. This paper is a pioneer in its application as there is no research paper adapted such claim in the field of translation.**

Key words: translation, computational, linguistic variations, speech, writing

1 INTRODUCTION

Computational linguistics (CL) combines resources from linguistics and computer science to discover how human language works. Computational linguistics is a vital field in the information age. According to [1],computational linguists create tools for important practical tasks such as machine translation, speech recognition, speech synthesis, information extraction from text, grammar checking, text mining and more. [2]has stressed the idea that contrastive Analysis (CA) is a method that is connected to Contrastive Linguistics, which is considered a branch of linguistics that focuses on illustrating the differences and similarities among two or more languages at different linguistic levels as semantics, syntax, and phonology.

According to [3] earlier programs have been criticized by the lack of a dictionary; to identify linguistic features, they relied on small lists of words that were built into the program structure itself. These lists included prepositions, conjuncts, pronominal forms, auxiliary forms. Since these word lists were relatively restricted, the grammatical category of many words in texts could not be accurately identified, and therefore these programs could not identify all of the occurrences of some linguistic features. The programs have been designed to avoid skewing the frequency counts of features in one genre or another so that the relative frequencies were accurate. The main disadvantage of this earlier approach was that certain linguistic features could not be counted at all. For example, there was no way to compute a simple frequency count for the total nouns in a text, because nouns could not be identified. For these reasons, the second set of programs has been taking place.

The second stage of program development took place during the years (1985-1986). The approach used in this stage is different from that of the first stage. As a result, a general tagging program to identify the grammatical category of each word in a text was developed. The aim is to develop a program that was general enough to be used for tagging both written and spoken texts. For example, the program could not depend on upper case letters or sentence punctuation. This goal is achieved by using a large-scale dictionary together with a number of context-dependent disambiguating algorithms. The main problem that had to be solved is that many of the common words in English are ambiguous as to their grammatical category. Words like "absent" can be either adjectives or verbs; words like "acid" can be either nouns or adjectives. All past and present participial forms can function as noun (gerund), adjective, or verb. A simple word like that can function as a demonstrative, demonstrative pronoun, relative pronoun, complementizer, or adverbial subordinator.

[4] has developed algorithms to disambiguate occurrences of certain words, depending on their surrounding contexts. For example, a participial form preceded by an article, demonstrative, quantifier, numeral, adjective, or possessive pronoun is functioning as a noun or adjective. That is to say, it is not functioning as a verb in this context; given this preceding context, if the form is followed by a noun or adjective then it will be tagged as an adjective; if it is followed by a verb or preposition, then it will be tagged as a noun. Tagged texts enable automatic identification of a broad range of linguistic features that are major for differentiating between genres in English. The tagged texts are subsequently used as input to other programs that count the frequencies of certain tagged items (e.g. nouns, adjectives, adverbs) and compute the frequencies of particular syntactic constructions (e.g. relativization on subject versus non-subject position).There has also been a debate concerning the need for a linguistic comparison of speech and writing. Historically, academics have regarded writing as the true form of language, while speech has been considered to be unstable and not worthy of study. By the early twentieth century, linguists regarded speech as primary and writing as a secondary form of language derived from speech; thus only speech was considered worth serious linguistic analysis. In fact, the historical view that written, literary language is true language continues as the dominant perception to the present time. it might well be the case that neither speech nor writing is primary; that they are rather different systems, both deserving careful analysis. This is in fact the view advocated by some researchers studying communicative competence. In this paper, the researcher is comparing the translation of google translate program in two different forms of the same language, speech and writing. In other words, are the two translations accurate? Is one of them better than the other and why?

2 RESEARCH QUESTION

1. From a language engineering point of view, can a computer program which is capable of translating written texts translate speech texts with same accuracy?

2. Is the translated text in the written form less erroneous than the text translated in the speech form?

3 BIBER'S VARIATIONS

[4] model is the main model upon which this study is based. The initial step is to collect the English and Arabic texts that are used as the data used in this study. Next, text normalization is crucial for any comparison of frequency counts across texts, because text length can vary widely. Translation studies, have only evolved during the last decades [5]. Scientific research in this area is a very recent phenomenon, as stressed by [6]. The call for research in translation is overwhelming as "a whole range of issues seemed to be waiting for examination, and inquiry is overdue". [7] Grammatical competence is concerned with the linguistic structure of 'grammatical' utterances; communicative competence is concerned with the form and use of all language - both speech and writing. Within this framework, neither speech nor writing needs to be considered primary to the exclusion of the other. Rather, both require analysis, and the linguistic comparison of the two modes becomes an important question.

4 THE SAMPLE

The sample used for this pilot study consists of two texts .The first one is the script that represents Donld Trump's speech. It was a telephone conversation with President Zelenskyy of Ukraine. The second text is extracted from a letter written to Donald Tramp regarding the family separation. The machine translation used is Google Translate that translates both texts into Arabic. In order to normalize texts, in this study, the frequency counts of all linguistic features are normalized to a text length of 1000 words so we have to delete some words to normalize the texts to have the same length.

5 Analysis

Translation is a complicated task, during which the meaning of the source-language text should be conveyed to the target-language readers. In other words, translation can be defined as encoding the meaning and form in the target language by means of the decoded meaning and form of the source language. [8] listed eight strategies, which have been used by professional translators, to cope with the problematic issues while doing a translation task as translation by a more general word, translation by a more neutral/ less expressive word, translation by cultural substitution, translation using a loan word, translation by paraphrase using a related or unrelated word, translation by omission or by illustration. Cohesion is the network of lexical, grammatical and other relations which provides links to various parts of the text. These relations organize a text and to some extent create it. One example of this is reference to other words and expressions in the surrounding sentences and paragraphs. Google translation tends to misuse references I in the speech texts. Most of the time pronouns are not with clear references. This is really clear with the pronoun (It) as there is no clear reference to whom it refers.

For example:

Much more than the European countries are doing and they should be helping you more than they are

Google Translation

أكثر بكثير مما تفعله الدول الأوروبية وينبغي أن يساعدك أكثر مما هو عليه

Here, google translation misused the pronoun as the correct translation should be

أكثر بكثير مما تفعله الدول الأوروبية وينبغي أن تساعدك أكثر مما هو عليه

E.g.

They are not working as much as they should work for Ukraine

Google Translation

انهم لا يعملون بقدر ما ينبغي أن تعمل من أجل أوكرانيا

Correct translation

انهم لا يعملون بقدر ما ينبغي أن يعملوا من أجل أوكرانيا

The problem of non-equivalence was apparent in google translation as the source language concept is not lexicalized in the target language. The source language word may express a concept which is known in the target culture but simply not lexicalized. In other words [8] stated that there is no specific term in the target language for this word that is not allocated in the target language word to express it. For example, the word (run more) this means to exceed or to go over. However, google misuse this by translating it into the literal meaning of the word which is totally incorrect .This means that the target language lacks this term.

E.g.

When I was speaking to Angela Merkel she talks Ukraine, but she doesn't do anything.

Google translation:

عندما كنت أتحدث إلى أنجيلا ميركل تتحدث إلى أوكرانيا ، لكنها لا تفعل أي شيء.

This is a literal translation that implies a vague and unclear message

Correct translation

على سبيل المثال عند التحدث إلى أنجيلا ميركل تتحدث عن اوكرانيا دون فعل شى من أجلها

E.g.

I told them that they are not doing quite as much as they need to be doing on the issues with the sanctions

Google translation

إنهم لا يقومون بالقدر الذي يجب عليهم فعله بشأن القضاياالمتعلقة بالجزاءات

Correct translation

إنهم لا يقومون بالقدر الذي يجب عليهم فعله بشأن القضايا المتعلقة بالجزاءا

E.g.

It turns out that even though logically, the European Union should be our biggest partner but technically the United States is a much bigger partner than the European Union

Google Translate:

أنه على الرغم من المنطق ، يجب أن يكون الاتحاد الأوروبي أكبر شريك لنا ولكن الولايات المتحدة من الناحية الفنية شريك أكبر بكثير من الاتحاد الأوروبي

Correct translation

علي الرغم من ان الاتحاد الاوروبي هوشريكنا الأكبر لكن فعليا الولايات المتحده هى اقوى واهم شريك لنا خاصا عند فرض العقوبات على الاتحاد الأوروبي

E.g.

We are almost ready to buy more Javelins from the United States for defense purposes.

Google translate

على وجه التحديد نحن على وشك. على استعداد لشراء المزيد من الرمح من الولايات المتحدة لأغراض الدفاع.

Correct translation

ونحن على استعداد لشراء السلاح اللازم من الولايات المتحدة لأغراض دفاعيه

E.g.

They are not working as much as they should work for Ukraine.

Google translate

انهم لا يعملون بقدر ما ينبغي أن تعمل من أجل أوكرانيا.

Correct translation

لقد قمت بالتحدث إلى انجلينا ماركل بشان الجهود الضئيله المبذولة فى قضايا العقوبات

Presupposed meaning arises from the co-occurrence restrictions on what other words or expressions that we expect to see before or after a particular lexical Item. This was clear in google 's translation we spend a lot of effort and a lot of time as collocational restrictions these are semantically arbitrary restrictions which don't follow the prepositional meaning of a word.

E.g.

Also, I think I should run more often so you can call me

Google translate

أعتقد أنني يجب أن أجري أكثر من مرة حتى تتمكن من الاتصال بي

Correct translation

واعتقد أن على بالاستمرار و مواكبه هذا التقدم لكى يستمر تواصلنا و علاقتنا ببعضنا البعض

Definitely, the implied meaning from this sentence is not "running as in a race" and that is why google misused the word as it should be المحاوله في الاستمرار

[8] referred to problems of non-equivalence above the word level. She suggested lots of strategies to solve such problem as translation by paraphrase. This is also related to google's ignorance of the fixed expressions of the target language.

For example:

We all watched from the United States and you did a terrific job. The way you came from behind

Google Translation:

شاهدنا جميعًا من الولايات المتحدة وقمت بعمل رائع. الطريقة التي أتيت بها من الخلف

Correct translation:

تهانينا على النصر العظيم الذى شهدناه جميعا من الولايات المتحده لقد قمت بعمل اكتر من رائع لقد اجتنزت الفرصه و انتصرت ببراعه

Here the words connote a different meaning rather than the expressed one. The literal translation of the words together does not deliver the right message as collocations and idiomatic expressions are really problematic in their translations. This according to Baker is referred to as translation beyond the word level. Register is a variety of language that a language user considers suitable to a specific situation. Its variation occurs from variations in the field of discourse as the linguistic choices will vary whether the speaker is taking part in a political speech or something else. Google misinterpreted this formality as in translating main words such as (congratulations) and (win big).

The problem of nonequivalence was apparent in google translation in (crowd stick) as this level of difficulty posed can vary tremendously depending on the nature of non-equivalence. Different kinds of non-equivalence require different strategies that the translator should be aware of. However, some strategies are difficult to handle.

Grammar is a set of rules which determines the way in which units such as words and phrases can be combined in a language .Google's translation totally neglected the English structure in the way it translated the sentences into Arabic .A language can of course express any kind of information its speakers need to express, but without using the same grammatical structure

For example

They are not working as much as they should work

Word by word translation is another fatal mistake that google has committed this way of translating word by word reflects the lack of terms and items in the target language which makes the final translation very poor and totally incorrect.

The tension between accuracy and naturalness is always apparent in google translation as Mona Baker stated that it is not easy for a translator to produce a collocation which is typical to the target language. This ideal cannot be achieved with the source collocation. For example, the word earlier is translated in a totally wrong way.

6 FEATURES OF SPEECH

1. Integration: It refers to the way in which a large amount of information is packed into relatively few words in typical writing, because the writer operates under few time constraints and can therefore construct a carefully packaged text. In contrast, typical speech cannot be highly integrated because it is produced and comprehended on-line. Features that are used to integrate information into a text include attributive adjectives, prepositional phrase series, phrasal coordination, and careful word choice.

2. Fragmentation: It refers to the linguistic characteristics of texts produced under severe time constraints, the case for typical speech. Under these conditions, information cannot be carefully incorporated into the text, and the resulting structure is much looser, or fragmented. Linguistic features associated with a fragmented text include clauses strung together with simple conjunctions (e.g., and) or with no connectives at all.

3. Involvement: This refers to those linguistic features which reflect the fact that speaker and listener typically interact with one another, while writer and reader typically do not. Due to this interaction, speakers often make direct reference to the listener (by use of second person pronouns, questions, imperatives, etc,), and they are typically concerned with the expression of their own thoughts and feelings (e.g., marked by use of first person pronouns, affective forms such as emphatics and amplifiers, and cognitive verbs such as think and feel). As a result of this concern, speech often has a distinctly non-informational and imprecise character (marked by hedges, pronoun it, and other forms of reduced or generalized content). These features can be considered together as the characteristics of involved text. In contrast, detachment refers to the characteristics of typical writing which result from the fact that writer and reader usually do not interact (e.g., marked by agentless passives and nominalizations).

7 FEATURES OF WRITING

1. [9] stated that writing is more structurally complex and elaborate than speech, indicated by features such as longer sentences or T- units and a greater use of subordination.
2. As for [10] more explicit than speech, in that it has complete idea units with all assumptions and logical relations encoded in the text.
3. [11] explained that it is more decontextualized, or autonomous, than speech, so that it is less dependent on shared situation or background knowledge.
4. [9] also illustrated that less personally involved than speech and more detached and abstract than speech.
5. [12] showed that it is characterized by a higher concentration of new information than speech (Brown and Yule 1983); and more deliberately organized and planned than speech. [12] noted that 'in writing we have time to mold a succession of ideas into a more complex, coherent, integrated whole', whereas speech, because it is produced on-line, is more fragmented.

8 CONCLUSIONS

There is a huge distinction between speaking and writing in their characteristics. There is also a difference in the channel. There may be many sub-channels available in speaking but only the lexical-syntactic sub-channel available in writing. Also, the opportunity for interaction with the text varies. In one hand, there are no real-time constraints in writing. On the other hand, severe real-time constraints appear in speech. Even these two differences are not absolute. Features such as underlining, bold-face, and certain punctuation marks can be used to represent prosodic or paralinguistic sub-channels in writing. Tape-recorded speech bypasses some of the real-time constraints of speech, more so in comprehension than in production. The recommendations of this study is to build a new grammar to be used in translating speech which is totally different than the grammar used in translating written texts of the same language to avoid the problems that exist due to the differences between the two forms of the same language.

REERENCES

[1] Mitkov, R. (ed.).(2015). The Oxford handbook of computational linguistics. 1st ed. Oxford University Press: Oxford

[2] Fisiak, J. (1981). Some introductory notes concerning contrastive linguistics. In J. Fisiak (Ed.), Contrastive linguistics and the Language Teacher (pp. 1-11). Oxford: Pergamon.

[3] Towell, R. & Hawkins, R. (1994). Approaches to second language acquisition. Clevedon: Multilingual Matters.

[4] Biber, D. (1988). Variation across speech and writing, Cambridge: Cambridge University Press.

[5] Broeck, R. (1986). Contrastive discourse analysis as a tool for the interpretation of shifts in translated texts. In House, J. & Blum Kualka, S. (Eds.) Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies, Tubingen: Gunter Narr P. 37-47.

[6] Gile, D. (1994) Beyond testing towards a theory of educational assessment, London: Falmer Press.

[7] Simon, S. (1996). Gender in Translation: Cultural Identity and Politics of Translation. London and New York: Routledge.

[8] Baker, M. (1992). In other words: A course book on translation. London: Routledge.

[9] CHAFE, Wallace. 1985. Linguistic Differences Produced by Differences Between Speaking and Writing. In OLSON, D., ORRANCE, N. & HILDYARD, A. Literacy, Language and Learning. Cambridge: Cambridge University Press, 1985. pp. 105-123.

[10] Coulmas, Florian. 1996. The Blackwell encyclopedia of writing systems. Oxford: Blackwell.

[11] Dewaele, J.-M. (2000). Gender, social and situational variables in the choice of speech style in native Dutch. Paper presented at the Sociolinguistics Symposium 2000, Bristol

[12] Coffin, Caroline et al. (2003). Teaching Academic Writing: A Toolkit for higher education. Routledge (London).

BIOGRAPHY

She has attained her PH-D degree in Translation studies from the Institute of Applied Linguistics and Translation (2018) , Faculty of Arts, Alexandria University. She has also attained her Masters degree in Applied Linguistics in (2014). Her main areas of interest are applied linguistics, computational linguistics and computational studies in the field of translation. Elsaadany is holding the position of a lecturer at Pharous University.

She is a former English instructor in the Arab Academy for Science and Technology. Furthermore, she is a speaking examiner for IELTS and OET exams.

Dr. Sameh Alansary: *Director of Arabic Computational Linguistic Center at Bibliotheca Alexandrina, Alexandria, Egypt.*

He is professor of computational linguistics in the Department of Phonetics and Linguistics and the head of Phonetics and Linguistics Department, Faculty of Arts, Alexandria University. He obtained his MA in Building Arabic Lexical Databases in 1996, and his PhD from Nijmegen University, the Netherlands in building a formal grammar for parsing Arabic structures in 2002. His main areas of interest are concerned with corpus work, morphological analysis and generation, and building formal grammars. He is also the head of Arabic Computational Linguistics Center in Bibliotheca Alexandrina. He is supervising and managing the Universal Networking Language project in Library of Alexandria since 1-6-2005 till now.

Dr. Alansary is the co-founder of the Arabic Language Technology Center (ALTEC), an NGO aims at providing Arabic Language resources and building a road map for Arabic Language Technology in Egypt and in the Middle East. He has many scientific works in Arabic Natural Language Processing published in international conferences and periodicals, and a member in many scientific organizations: (1) Egyptian Society of Language Engineering, Cairo, (2) Arabic Linguistic Society - USA, (3) Association of Computational Linguistics - USA – Europe, (4) Universal Networking Language foundation, United Nations, Geneva, Switzerland.

للكشف عن التباين اللغوي: منهج هندسة اللغة Biber دراسة تجريبية لنموذج

مرام السعدني

جامعه فاروس- الاسكندريه- مصر

maram.elsaadany@pua.edu.eg

سامح الانصاري

مصر-مكتبه الاسكندريه -الاسكندريه

قسم الصوتيات- كلية الاداب- جامعة الاسكندرية

sameh.alansary@bibalex.org

ملخص - هذه الورقة هي في المقام الأول تحليل ترجمة لخطاب ترامب ورسالة مرسلة إلى الرئيس ترامب بشأن انفصال الأسرة عن "مؤتمر القيادة حول الحقوق المدنية والإنسانية". ترجمة جوجل هو المحرك المستخدم لترجمة هذين النصين إلى العربية. البيانات المحددة لهذا التحليل هي ١٠٠٠ كلمة في ميزة لإثبات أن الكتابة أكثر تعقيدًا من الكلام. تفرض الدراسة إدعاء بيبر بأن الكتابة أكثر تعقيدًا من الكلام. تؤكد النتائج كل برنامج نصي. استخدم نموذج ادعاءات لأن هناك الكثير من المشاكل في ترجمة نص الكلام إلى العربية مقارنة بالنصوص المكتوبة المترجمة. توضح النتائج حقيقة أن يجب أن يتم تكييفها وتجهيزها بقواعد جديدة للتعبير تختلف عن تلك المستخدمة في الكتابة لتحقيق أفضل نتيجة لكلا الترجمتين في نفس اللغة. تعتبر هذه الورقة رائدة في تطبيقها حيث لا توجد ورقة بحثية تم تكييفها لهذه المطالبة في مجال الترجمة.

الكلمات المفتاحية: الترجمة ، الاختلافات الحسابية ، اللغوية ، الكلام ، الكتابة

Towards Building a Semantic Role Labeling System for Modern Standard Arabic: A Rule-based Approach

Amena Helmy Abd El-khaleq Muhammad Dheif[1]*, Sameh Saad Abu-Elmagd Alansary*[2]

Department of Phonetics and Linguistics, Faculty of Arts, Alexandria University.

amena.helmy@alexu.edu.eg

Sameh.Alansary@bibalex.org

Abstract— ***Semantic role labeling (SRL), the computational identification and labeling of arguments in text, has become a leading task in computational linguistics today, as it is a step towards natural language understanding (NLU). Many linguistic theories had discussed the nature of semantic roles, on the other hand, language engineers headed directly to build automatic statistical-based and rule-based SRL systems that would predict the correct semantic roles of a predicate argument structure. Many systems, corpora, and lexical resources have been developed to tackle the task of semantic role labeling in many languages, but Arabic lacks such attention. The study's main goal is to build a rule-based semantic role labeler for Modern Standard Arabic (MSA), which is the first rule-based SRL system to be developed on MSA, that works on the verbal predicates of the top frequent 40 Arabic VerbNet (AVN) classes. The study defined the input data to be gold-standard fully syntactic trees drawn from the Arabic Penn Treebank (ATB) part1 and part3. Forty knowledge bases were constructed to store the syntactic and semantic features that realize the semantic roles of each AVN class predicates. Features of each AVN class were extracted from the training data of the frequent predicate, but it was used to represent the whole class. The developed system was tested on two types of testing data, the first type is drawn from the sentences of the same frequent verb the knowledge base is built upon and the second type is drawn from the sentences of the other verbs that belong to the same AVN class. The current system achieves a final F1 score of 92.9% over the first type of testing data, F1 score of 86.0% over the second type of testing data and a 91.0% F1 score on both.***

Keywords***: Semantic role labeling, VerbNet, Predicate, Core semantic role, Adjunct semantic role.***

1 INTRODUCTION

Semantic parsing is one of the main topics of NLU that has undergone intense study after syntactic parsing was found to be insufficient for providing the layer of meaning representation needed for proper understanding of natural language. Semantic parsing has mainly two types, the first type is deep semantic parsing and it is defined as "Mapping a natural language sentence to a detailed representation of its complete meaning in a fully formal language that has a rich ontology of types, properties, and relations, and supports automated reasoning" [1]. The Abstract meaning representation (AMR) project [2] is one of the main projects that seeks to provide a fully semantic parsed corpora for English. The second type of semantic parsing is Semantic Role Labeling (SRL) or shallow semantic parsing which is the core topic of the current study. SRL is considered as a shallow semantic parsing because it represents only a subtype of all the semantic relations, which is the predicate-argument semantic relations. SRL task can be defined as the automatic analysis of an input text into the number of propositions that compose it, where a proposition consists of a predicate and its set of arguments, then map the predicate arguments to their semantic relations [3] to properly answer the question of who did what to whom and perhaps when where and why[4]. Figure (1) shows a shallow semantic representation of sentence (1) where it consists of only one proposition and maps each argument to its correct semantic relation.

١) المحققُ قبضَ على المتهمِ في مسرحِ الجريمةِ مساءً .

- al-muḥaqiqu qabaḍa ʻalá al-mutahami fī masraḥi al-jarīmati masāʼan.[4]

[4] Arabic transliteration is provided according to the ALA-LC Romanization Scheme which provides transliteration for Non-Roman languages and is approved by the Library of Congress and the American Library Association [5].

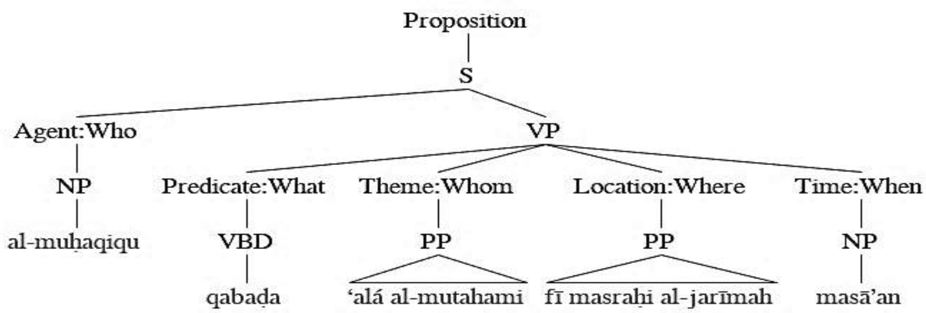


Figure 1: A shallow semantic representation of sentence (1)

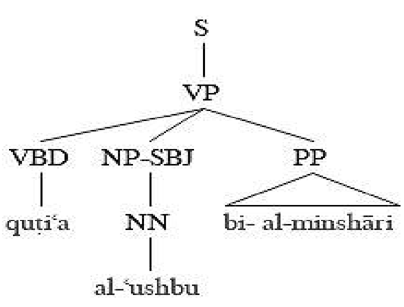
A predicate is the main unit in any proposition, Arab grammarians call predicates the “musnad ” of the proposition. Ref. [6] defines the predicate as “any word (or sequence of words) which (in a given single sense) can function as the predicator of a sentence”. In turn, the predicator is defined as “the word (sometimes a group of words) which does not belong to any of the referring expressions and which, makes the most specific contribution to the meaning of the sentence. Intuitively speaking, the predicator describes the state or process in which the referring expressions are involved.”[6].

The second main unit that makes up a proposition is the predicate arguments. Each argument plays a certain semantic role assigned by the predicate. Predicates assign their arguments two kinds of semantic roles namely core semantic roles and adjunct (or peripheral) semantic roles. On one hand, an argument would be classified as playing a core role if it is essential and central to the event or state being described by the predicate, on the other hand, an argument would be seen as peripheral or adjunct if it provides extra information about the event or the sate being described by the predicate [7].

As mentioned above, SRL is considered important as it provides a layer of shallow semantic representation that depicts the predicate-argument semantic relations which is a major step towards NLU. But what makes SRL more important is that SRL systems abstract and unifies the different argument structure realizations of a predicate which is called “diathesis alternations” [8] into a unified semantic representation [9]. This contribution takes SRL a step ahead of syntactic parsing, for example, in both sentences (2) and (3), the argument “al-‘ushb (the grass)” plays the patient semantic role, though it is realized as a subject in sentence (2) and as an object in sentence (3). The reason behind the different realizations of the same role is the voice changing of the verb in the two sentences. Another difference can be seen with the argument “al-minshār” where it’s realized as an indirect object in sentence (2) and as a subject in sentence (3) while it plays the instrument semantic role in both sentences. The reason behind the different realizations of “al-minshār” argument is the instrument alternation of predicate “qaṭa‘” [10]. Figure (4) shows the shallow semantic representation both sentences (2) and (3) map to.

٢) قُطِعَ العشبُ بالمنشارِ. (quṭi‘a al-‘ushbu bi-al-minshāri)

٣) المنشارُ قطعَ العشبَ. (al-minshāru qaṭa‘a al-‘ushba.)



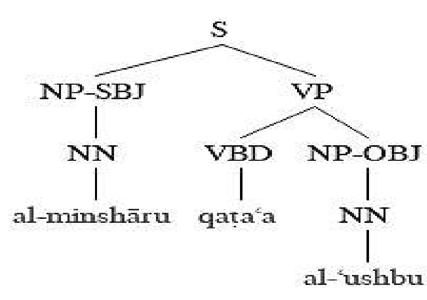


Figure 2: syntactic tree of sentence (2). **Figure 3: syntactic tree of sentence (3).**

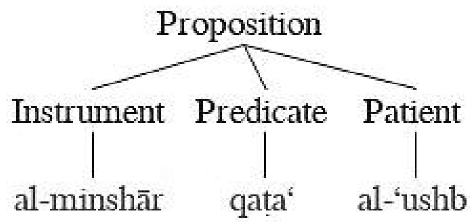


Figure 4: The shallow semantic layer both sentence (2) and sentence (3) map to.

Of course, Many higher-level natural language processing tasks like Information extraction, question answering, automatic summarization, and machine translation rely on the layer of semantic representation provided by an SRL system and consider it as an intermediate stage that need to be performed before pursuing to its main task since it is easier to work on a layer of semantic representation than working on a raw text [11].

2 LITERATURE REVIEW

The notion of semantic roles had been studied thoroughly and extensively in many linguistic theories. It has a numerous different terms across different theories, as it is termed deep semantic cases [12] Theta roles by Chomsky and his followers [13], proto-roles [14], thematic relations by [15], [16] and semantic macro roles [17]. The term "semantic roles" is the term used in the current study.

Unfortunately, these extensive studies were not found to be that fruitful since none of these theories provided a standard universal set of semantic roles that could be applied to any language [9]. The popular proposed lists range from a large set of situation-specific roles such as the Killer and the Victim roles used in the semantic frames theory [18] to a small set of general roles, such as the Agent, Theme, Location, and Goal roles used in [12] to an overgeneralized set of two proto-roles like the Proto-Agent and Proto-Patient roles used in[14].These linguistic theories also fail in providing a unified definition for the main roles like the "Agent" role that they all share. Another problem is the different rolesets designed for each predicate and the different criteria used to differentiate between core and adjunct roles of a specific predicate.

While linguists were trying to settle down their arguments around the notion of semantic roles, computer scientists started to develop and build systems that could map a sentence to its underlying semantic representation, without trying to uncover thoroughly the linguistic side of the phenomenon. Their motivation was not to answer the linguistic questions of what semantic roles are or what are the linguistic principles that govern it, but rather to try to develop systems that could with some linguistic and world knowledge provide a semantic interpretation of a text [11].

Early developed SRL systems were domain-specific like the Air Traveler Information System (ATIS) which was developed to enable users to book their flights automatically [19]. The system understands the users' queries by mapping their queries constituents to a pre-arranged set of semantic roles related to the air travel semantic frame only like destination and flight date roles [20]. SRL systems had undergone a breakthrough after the advent of large corpora annotated with semantic roles like Propbank [21] and FrameNet [7], and because of the availability of other lexical resources like the VerbNet [22] which is a verb lexicon that consists of a number of classes where verbs that behave syntactically alike are assigned the same class, and because of the availability of elaborate machine learning algorithms. This mixture motivated researchers to develop domain-independent statistical SRL systems that could train over a vast amount of annotated data to extract linguistic features that realize semantic roles and then tag unseen other data. The CoNLL, the Conference on Natural Language Learning, devoted three of their yearly meetings for the SRL task [3],[4], [23].

Most of the developed SRL systems tackled English texts till Diab [24] proposed a statistical SRL system that was trained over Modern Standard Arabic (MSA) data. The release of the Arabic version of Propbank

[25] made developing such a system feasible. This system was built following the methodology advocated for other SRL systems without considering the special morphological, syntactic, and semantic characteristics of the MSA. The system yielded an F1 score of 81.34%. Ref. [26] proposed another SRL system that considered the special linguistic characteristics of MSA which yielded an F1 score of 82.17% which is to be considered as an improvement in the state of art of Arabic statistical SRL. Although the majority of the SRL systems follow the statistical-based approach, this approach suffers from many drawbacks. To begin with, availability of a large amount of annotated corpora is a need to build a statistical SRL system. There is a direct correlation between the amount of annotated data and the accuracy of the system. But the major flaw of the statistical SRL system is that the developed systems extracted features over-fits the data it trains on, thus when trying to test the developed system on other genres, the system shows a significant performance deterioration.

Rule-based SRL systems are really few. The only well-known rule-based system was developed by [27]. Semantic and syntactic rules were extracted from the FrameNet annotations and then added to a knowledge-base. The system achieved 74.5% accuracy. Though the rule-based approach is not used widely to build SRL systems since extracting and structuring such rules is very exhausting, its bright side comes from its reliance on extracting economic general relevant features that would achieve the task in hand effectively, unlike statistical SRL systems where plenty of features are extracted.

3 METHODOLOGY

This section describes in detail the design and implementation of the MSA rule-based SRL system. It starts by giving a brief description of the resources and tools needed by the system. Then it describes in detail the procedures followed to design the training and testing data, construct the system knowledge bases, and build the rule-based SRL system.

I. Data Resources

This section provides a brief description of the data resources used to build the system and the reason behind choosing these resources.

1) *Arabic Penn Treebank (ATB):* The (ATB) project [28] is a million words treebank, built to support the development of natural language processing research on the written MSA and was accomplished by a team from the Linguistic Data Consortium (LDC) and the University of Pennsylvania. The ATB is a newswire corpus that is comprised of three parts, ATB1 which represents 166K words of 734 stories drawn from the Agence France Presse corpus, ATB2 which represents 168K words of 501 stories drawn from the Ummah Arabic News Text corpus and ATB3 which represents 350K words of 600 stories drawn from Annahar News Agency corpus. The current study selected its dataset from both ATB1 and ATB3, as these two parts are the only available parts for the researcher. The ATB corpus is provided with both morphological and syntactic annotations. The drawback ATB suffers from is the lack of word sense disambiguation (WSD) for the token's lemmas.

2) Arabic VerbNet (AVN): the AVN [8] is a lexical resource that provides a classification of the Arabic verbs according to Levin's verbs classification structure [10] where verbs that behave syntactically alike are assigned to the same class, as they share the same semantic meaning according to Levin's hypothesis. The current study uses the AVN to assign the verbal predicates drawn from the ATB to their corresponding AVN classes. Of course, polysemous and homonymous verbs would be found to belong to multiple AVN classes which is the second advantage AVN provides, as it shows the different senses of these verbs by showing the different AVN classes they belong to.

3) SemLink: Since the current study needs to define a roleset for each verbal predicate, the three resources namely the FrameNet, the VerbNet and the Propbank were the researcher's destination as each of them specify a roleset for each predicate. Unfortunately, these three resources follow different criteria while defining a predicate roleset which yields different rolesets for the same predicate. SemLink is an ongoing project which maps these three resources rolesets of each predicate, which is what the current study needs [29].

4) Arabic And English WordNet: WordNet [30], [31] in general is a large electronic thesaurus-like relational database that groups synonymous nouns, verbs, adjectives, and adverbs that represent a unique lexicalized concept together into sets called synonym-sets. The synonym-sets are interlinked together with a number of lexical relations such as antonymy, hyponymy, meronymy, and entailment, which results in a semantic network that can be navigated. The Arabic WordNet (AWN) database current version AWN.2 covers around 11,269 synonym-sets [32] and the English WordNet Current version EWN.3 covers around 117,597 synonym-sets [30].

J. Tools

This section provides a brief description of the tools used to build the SRL system and the reason behind choosing these tools.

5) *Python*: Python [33] is an interpreted powerful high-level, object-oriented programming language created by Guido van Rossum. It is not really a tool, but it represents the programming language used to build the system. The reasons behind choosing Python as the language to work with are, the availability of many APIs written in the Python language, and the availability of the natural language toolkit (NLTK) which is the main reason behind using Python. Python 3.6 is the version used for building the study's SRL system.

6) Extensible Markup Language (XML): XML is a markup language designed to store and transport data. Again, XML is not a tool, but it represents the language chosen to build the knowledge bases.

7) *Natural language toolkit (NLTK):* NLTK is a famous open source Python library that contains corpora and modules which supports the development of NLP applications. Using NLTK in the current study facilitated working with and traversing through the complicated syntactic tree data structure to extract the needed information and allowed the direct use of the EWN as it is available within its corpora.

8) *Google translate API:* Google Translate API [34] is a free Python library that translates texts from a source language specified by the user to a target language. It can also detect the language of an input text. The reasons behind using the Google translate API are its high speed and the reliability and consistency of its translation. A study made by [35] proved that Google Translate outputs are slightly better than the outputs of Bing, Babylon and Systranet machine translation Systems.

K. Procedure

This section provides in detail the procedure followed to, firstly choose the study's list of semantic roles, secondly select the AVN classes along with their verbal predicates that this study will focus on, thirdly design the core roleset of each AVN class, fourthly design the study's adjunct semantic roles, fifthly design and annotation of data, sixthly design the linguistic features that will be extracted to be stored in the knowledge base of each AVN class, seventhly construct the SRL system and finally design the testing data and how the performance of the system was calculated.

1) Choosing the semantic roles list: Proposing the study's semantic roles list was a troublesome challenge since a standard agreed-upon list of semantic roles does not exist. Many semantic roles lists were proposed by many theories, corpora, and projects, each designed based on different criteria. The current study adopted a different number of criteria that defines the characteristics of a proper list of semantic roles. These criteria are adopted from [36], [37] The criteria state that:

I. The set of semantic roles should be small in size [36].
II. "Semantic roles are neither syntactic nor lexical structures but are semantic categories" [37].
III. Semantic roles list must not contain roles defined for specific verb or classes of verbs" [37].
IV. Semantic roles list must be universal, it could describe the semantic relations within any language.

Regarding the adopted criteria, the study mainly adopted its list of semantic roles from the LIRCIS Project [37] since most of the adopted criteria are drawn from that project besides the project's proposed list was tested for completeness on many languages and yielded successful results. Some other roles were drawn from The UNL project set [38] since the UNL set was meant to be a universal set to suit any language. These roles are not found in the LIRICS project but were found in Arabic. Other roles were drawn from the Propbank set of semantic roles [39] since the Propbank set could be seen universal as it was used to annotate corpora from eight languages one of them is the Arabic language. The adopted semantic roles list consists of 32 roles.

2) Assigning the PATB verb lemmas to their AVN class: The second step considered assigning the PATB verb lemmas to their AVN classes. The verb lemmas extracted comprises 2,355 verb lemmas. The AVN is comprised of 336 verb classes populated with around 7,774 verb entries [8]. The output of this step was 366 files, with each file representing one of the AVN classes, populated with the AVN class verb lemmas along with their frequencies in the ATB.

One major problem encountered while extracting the frequencies of the verb lemmas was the need to disambiguate the polysemous and homonymous verbs senses like "أصاب" (aṣāba) polysemous verb which is assigned to the two AVN classes "jaraḥ" class and "ishtará" and has a frequency of 116 sentences that needs to be disambiguated according to these two meanings to correctly assign the frequency of each sense. A layer of verb sense disambiguation was required to tackle this problem, unfortunately, this layer was not available, thus the researcher carried on the disambiguation phase manually.

3) Designing the core roleset of the selected AVN classes: Designing the core roleset of the selected AVN classes considers the part where the number and types of the AVN classes core semantic roles are specified. The three popular resources namely the FrameNet, the VerbNet and the Propbank designed their core semantic roles of each class or frame depending on different criteria, which yielded different rolesets for the same class across the three resources. Thus, the study needed to define its own criteria upon which the rolesets of each class could be selected. The basic criterion the study adopted states that the obligatory participants (core roles) are implied in the semantics of each verb whether or not they are syntactically instantiated [40]. When these semantically obligatory arguments are not explicitly syntactically expressed, it is termed implicit arguments[40]. FrameNet design of the core frame elements depends on many criteria, one of them is the previous criterion, which is stated as follows "A frame element which, when omitted, receives a definite interpretation, is also core" [7]. For that reason, the current study adopted the FrameNet design of the core roleset which yielded a need to map the study's AVN classes to their corresponding FrameNet frames. Mapping each AVN to their corresponding FrameNet frame was accomplished using the SemLink resource. After this mapping step, another step was needed which concerns mapping the situation-specific frame elements of the chosen frame to their corresponding semantic role tags from the studies proposed list.

4) Designing the study's adjunct roleset: The current study adopted the Propbank criteria, where the same set of adjuncts is used across the chosen classes unless one of the adjuncts is specified as a core role in a class, it is then transferred from the adjunct set to the core set as the case with "ḥaṣala" AVN class where both the "Time" and the "Location" roles are specified as core roles of that class where these two roles are in most case adjunctive roles.

5) *Annotating data with semantic roles:* the researcher manually annotated the trees of each AVN class with their semantic roles, to be used afterwards to extract the relevant linguistic features with which each semantic role is realized, then properly structure and organize these features in the class knowledge base to be used by the system eventually.

6) Design of the Linguistic features to be extracted: One of the current study's goals is to design an economic number of linguistic features that MSA uses to realize the semantic roles of any predicate.

The study designed a number of features related to the predicate which were found to play a significant role in the assignment of the predicate semantic roles to its arguments, these features are:

- Voice: the voice feature was extracted directly from the label of the predicate in the syntactic tree, which has the value of either active or passive voice.
- The syntactic frame: the syntactic frame represents the sister nodes of the verbal predicate in the syntactic tree and the sisters of the verb phrase (VP) that contains that predicate. This feature collects all the predicate arguments that represent the candidates that would realize the predicate semantic roles.
- The predicate sense: the predicate sense is inferred from the AVN class name the predicate belongs to.
- The roleset of the predicate: both the core and adjunct roles of the predicate are retrieved from the XML lexicon that is built for each class, which will be discussed later.

Another list of features was extracted, which is related to each argument which are:

- Phrase type: phrase type feature represents the syntactic category (NP, PP, etc.) of the argument. It is drawn from the label of the argument in the syntactic tree.
- Function tag: This feature indicates the grammatical function of the NPs whether it is subject or object, and the other function tags the ATB provides are discarded as the TMP, LOC. etc.
- Status: the status feature shows if the argument is dropped, topicalized, sister to the predicate or sister to the VP.
- Distance: the distance feature indicates how far the argument is from its predicate. This feature is calculated only for the arguments that are sisters to the predicate.
- Headword: The headword feature was firstly introduced in [20] and It is used to show the selectional restrictions of the lexical items that might fill a semantic role. The headword extraction process follows the headword rules defined by Collins [41].
- The headword lemma: the lemma of the headword is extracted.
- The POS tag of the headword: the POS tag of the headword is extracted from the label of the headword in the syntactic tree. The reason behind extracting the head lemma POS is the fact that it replaces to some extent the name entity feature, as proper nouns lemmas which map mostly to a person, organization, or location name entities.
- Semantic features of the headword: this feature considers extracting the semantic features of the headword lemma. To achieve this task, the same procedure followed by [20] of getting semantic features by traversing the WordNet hypernyms hierarchy of the headword lemma was followed. The list of hypernyms of that lemma is considered as the semantic features of the headword lemma. AWN version 2 was used. AWN was found to be suffering from poor coverage since many of the searched lemmas weren't found which affected the results of the system dramatically. To overcome this problem, lemmas that are not found in the AWN were translated using the Google translator API and then their hypernyms were drawn from the EWN version 3 which is available among the NLTK corpora.
- Content word: this feature was introduced by [42], where it selects a content word of the argument. Content word feature was extracted for PP arguments only where the right most child is selected. Content word feature was introduced to give better information when the headword feature is less informative, like the case with the PP arguments.
- Lemma of the content word.
- POS tag of the content word.
- Semantic features of the content word.

7) Knowledge base construction: The knowledge base construction phase considers the part where the extracted features are properly structured and stored in a knowledge-base to be used later by the system. For each AVN class, a knowledge base was created with the XML language. Each AVN class knowledge base contains the AVN class name, its mapped FrameNet frame, and the Arabic Propbank (APB) frame-file name of the frequent verb. The knowledge base is then divided into two parts, the core semantic roles part and the adjunct semantic roles part. For each role either core or adjunct, the role tag name is provided according to the study's proposed list of semantic roles, along with its APB correspondent tag. The features of each role are then split according to the predicate voice since it was found that the features of the same role change according to the predicate voice.

For each role and within each voice, the different phrase types that were found to realize that role in the annotated data were extracted and stored in the knowledge base along with their frequencies. Each phrase type is then constrained with other features which are the function tags that cooccur with the phrase type if are, the different statuses of the same phrase type and the different number of distances the same phrase type may occur away from its predicate. The semantic selectional restrictions of the headword lemmas and the content word lemmas of that phrase type is also stored but according to an order of priorities which goes as follows: if the list of the semantic features of the headword lemma is retrieved successfully, then the most proper semantic feature that can be seen as the selectional restriction that made the headword fit for the role it realizes is extracted and stored, on the other hand if the system fails to retrieve such list, the POS tag of the headword is stored but only if the POS tag again plays the role of the selectional restriction, otherwise the headword lemma is stored. The content words of the phrase type are stored in the lexicon following the same procedure of storing the headword. Each phrase type of each role has a minimum score attribute that represents the minimum number of features an argument needs to be marked with, in order to be classified with that role. Also, each phrase type has a score attribute that represents the ideal number of features an argument needs to be marked with, in order to be classified with that role.

8) Building the Rule-based SRL system: Building a rule-based semantic role labeling system for MSA is the ultimate goal of the current study. The input data that shall be fed to the system must be a fully syntactic parsed sentence according to the ATB parsing convention. The strategy of choosing to work with syntactic hand-annotated data was justified by the researcher need to avoid the system performance errors that may result from feeding erratic input data to it.

The first step the system performs considers detecting the verbal predicate in the syntactic tree and then extracting its/their lemma(s). The lemma extracted was then searched and if it was found to be one of the 372 verbs the current study focuses on, the predicate features explained earlier were extracted from the syntactic tree.

The second step, then extracts from the fed syntactic tree the syntactic nodes that are highly probable to be assigned a semantic role. The current study followed the same hypothesis of [43] where it was claimed that the candidates that highly represent the predicate semantic roles are located in the predicate immediate clause which represents the clause that contains the predicate. Thus, the study focused on retrieving the syntactic arguments that are either sisters to the predicate or sisters to the VP that contains the predicate. Retrieving the candidate arguments of the extracted predicate was not a challenge-free task, two major challenges were encountered. The first challenge encountered pertains to retrieving empty arguments, and the second challenge pertains to retrieving arguments that have an explicit pronoun "dhamīr ẓāhir" headword. These two problems needed to be solved with anaphora resolution layer the system lacked, thus, the problem was left unhandled.

The last step is the critical step in the whole system, as it considers the part where the system predicts and assigns the correct roles to the retrieved arguments. This step starts with retrieving the AVN class knowledge-base, the verb belongs to. The system firstly extracts the core and adjunct roles features from the knowledge base according to the predicate voice and then calculates the probability of each retrieved syntactic argument to be assigned any of the class core or adjunct roles. This calculation is done firstly by comparing the phrase type of each syntactic argument to the phrase types of each role (PTs), and if the argument PT was found among any role PTs the probability score of that argument to be assigned that role becomes one, and consequently the other argument's extracted features get compared to that role PT constriction rules which includes:

- Searching the argument headword lemma among that role headword lemmas, if found the argument probability score is incremented by one, if not found, the argument headword lemma's semantic features are searched among that role selectional semantic features and if one of the argument features was found within the role features, the argument probability score gets incremented by one, if not found the headword lemma POS tag is searched among that role POS tags and if found the probability score gets incremented by one, otherwise the probability score is left unchanged.
- Searching the argument content word if it is different from the headword follows the same strategy explained above.
- Searching the argument function tag FT among that role FTs, if found the probability score gets incremented by one, otherwise the probability score is left unchanged.
- Searching the argument status among that role statues, if found the probability score gets incremented by one, otherwise the probability score is left unchanged.
- Searching the argument distance among that role distances, if found the probability score is incremented by one, otherwise the probability score is left unchanged.

After the calculation of the argument probability score for each role, the argument is added as a potential realization of any role if its final score equals or exceeds the minimum score specified for that role.

After getting the potential arguments of each role, a filtration process takes place where each role keeps only the argument with the highest score and the other potential arguments of that role are erased even if they have the minimum role score.

After the filtration process is done, another filtration process takes place where the scores of the arguments that are assigned multiple roles are compared to assign that argument the role at which it gets the highest score.

In some cases, it was found that the same argument may get the optimum score of two different roles. At this condition, the system assigns that argument the more frequent role of the two role candidates.

In some other cases, some arguments are left without being assigned any semantic role as their features suffice none of the roles specified in the knowledge base, at this time the argument gets assigned with a "None" role. The system specifies for each sentence its syntactically unexpressed obligatory semantic roles. The full system flow diagram is shown in figure (5).

9) *Testing phase: T*he testing phase considers first the strategy followed to design the testing data, and second, the measurement calculated to evaluate the SRL built system.

Designing the testing data followed two different paths which resulted in two types of testing data. The first type represents unseen testing sentences drawn from the same frequent verb, to find out the system accuracy when implemented over unseen data which is called testing data type 1. The second type represents the testing sentences drawn from the other verbs that belong to the same class to find out if the features drawn from the annotated data of the most frequent verb would apply correctly to the other class predicates which is called testing data type 2. The researcher claims that the system would perform accurately when assigning core roles of any AVN class with both types of testing data, and it would perform moderately when assigning adjunct roles. Figure (6), shows the system output for one of the testing sentences.

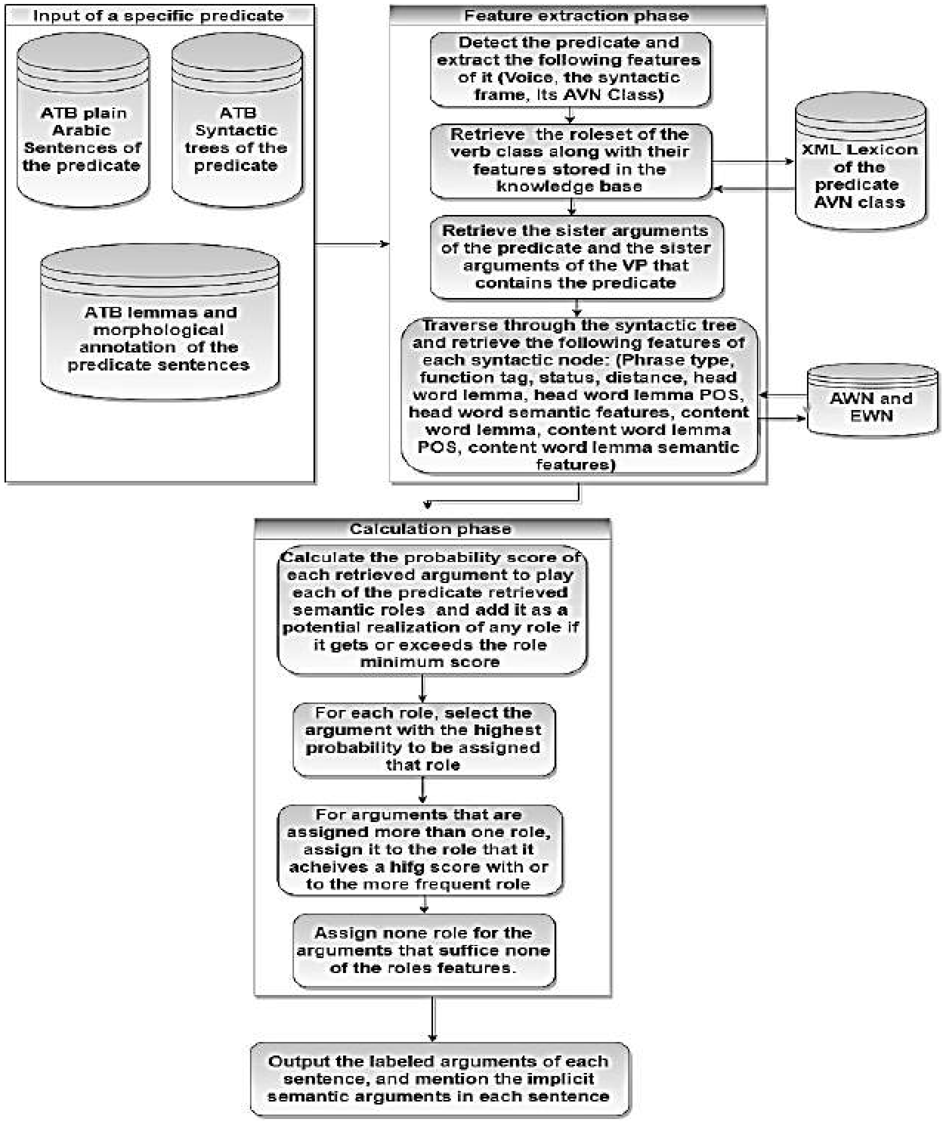


Figure 5: The full system flow diagram

وقد قتل اكثر من 90 شخصا منذ بداية تشرين الاول/اكتوبر في مجازر واعتداءات نسبت الى الجماعات الاسلامية المسلحة المعادية لسياسة الوفاق الوطني التي ينتهجها الرئيس عبد العزيز بوتفليقة وفق حصيلة اعدت استنادا الى الصحف .

- Predicate Voice: Passive
- Type: Core
  Patient: اكثر من 90 شخصا
  Optimal score: 5 arg_score: 5
- Type: Adjunct
  Time: منذ بداية تشرين الاول/اكتوبر
  Optimal score: 4 arg_score: 4
- Type: Adjunct
  Setting: في مجازر واعتداءات نسبت الى الجماعات الاسلامية المسلحة المعادية لسياسة الوفاق الوطني التي ينتهجها الرئيس عبد العزيز بوتفليقة
  Optimal score: 4 arg_score: 4
- Type: Adjunct
  Adverbial:وفق حصيلة اعدت استنادا الى الصحف.
  Optimalscore:4 arg_score:4

- **Not_instintiated: Agent, Cause, Instrument, Means

Figure 6: The system output for an input sentence.

The system is evaluated with respect to precision, recall and the F1 (F-measure) of the predicted arguments. To calculate these measurements, the study must first define when an argument is correctly labeled. The study adopted the definition of a correct assigned argument from [44] which states, for an argument to be recognized as labeled correctly, the words spanning the argument as well as its semantic role tag must be correct. When one of these two criteria drops the assigned argument is considered wrong. The second two criteria are adopted from [45]. These two criteria state that:

1. "If an argument is a reference (R-arg) to some other argument arg, then this referenced argument must exist in the sentence.
2. If there is a C-arg argument, then there has to be an arg argument; in addition, the C-arg argument must occur after arg." [45].

The current system allows the duplicate roles since it was found that Arabic allows such repetition of roles, especially with the adjunct roles.

The measurements calculated are the Precision, recall and F1 each calculated for the core roles, adjunct roles, and all roles. Precision is defined as "the proportion of arguments predicted by a system which are correct" [3]. It is calculated with the following formula: Precision = roles correctly assigned / roles assigned relevant. Recall is defined as "the proportion of correct arguments which are predicted by a system [3]. It is calculated with the following formula: Recall = roles correctly assigned / total of roles. Finally, the F1 measure "computes the harmonic mean of precision and recall and is the final measure to compare the performance of systems" [3], It is calculated with the following formula: F1 = (2 * Precision * Recall) / (Precision + Recall).

4 RESULTS

This section shows the results of the built rule-based SRL system over the manually annotated data and testing data. Table (I) shows the number of sentences of the training and testing data.

Table I

THE NUMBER OF SENTENCES OF THE TRAINING AND TESTING DATA.

Manually annotated data sentences of all classes	Testing data of type 1 sentences	Testing data of type 2 Sentences	Total of Testing data sentences	Total
2786	1439	821	2260	5046

The manually annotated data represents 55.2 % of the whole sentences fed to the system, the other 44.8% of sentences represent the testing data fed to the system which is comprised of 63.7% sentences that represent testing data type 1and 36.3 % of the sentences represent testing data type 2. Running the system over the manually annotated data was done firstly to find out if the linguistic features extracted from the manually annotated data was structured properly and secondly to find out if the calculation procedure adopted by the system to predict the propositions' semantic roles is accurate. Table (II) depicts the precision (P), recall (R) and F1 of the system results over the manually annotated data, testing data of type 1 and type 2.

TABLE II

PRECISION (P), RECALL (R) AND F1 OF THE SYSTEM RESULTS OVER THE TRAINING DATA, TESTING DATA OF THE SAME FREQUENT VERB AND OF THE OTHER AVN CLASS VERBS TESTING DATA.

Data	Core P	Core R	Core F_1	Adjunct P	Adjunct R	Adjunct F_1	All roles P	All roles R	All roles F_1
Manually annotated data	97.6	٩٦,٤	96.9	٩٤,٥	٩٢,١	93.3	٩٦	٩٥,٢	95.6
Testing data type 1	97.1	96.7	96.9	91.1	82	86.3	93.1	92.8	92.9
Testing data type 2	94.5	90.3	92.4	82	69	75	87	85.1	86

Figure (7) shows the overall system performance over the three datasets according to the core, adjuncts, and all roles.

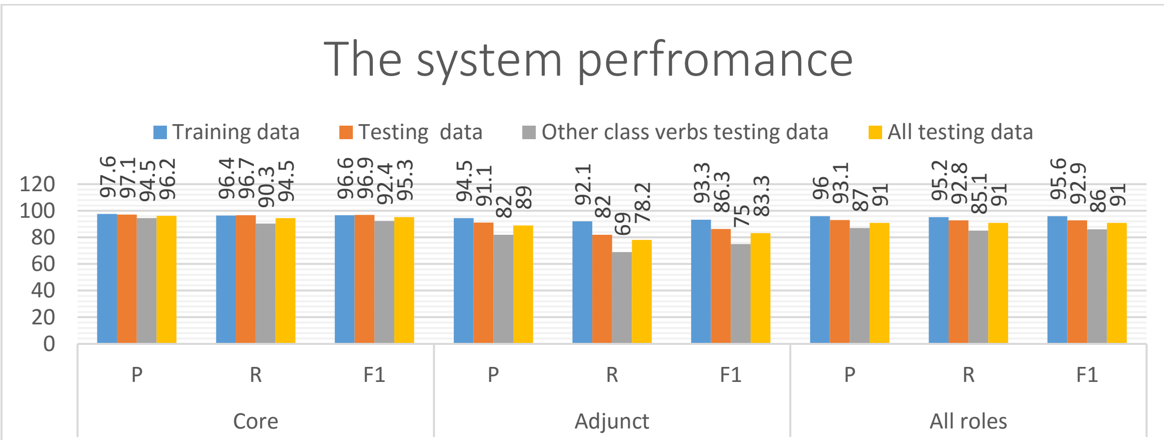


Figure 6: the overall system performance over the three datasets according to the core, adjunct and all roles.

5 DISCUSSION

The results achieved by the system shall be discussed in this section to give a linguistic insight of the results that goes beyond just the numbers.

L. Core results Vs. adjunct results.

The system achieved nearly consistent F1 scores while predicting the core roles across the manually annotated data, the first type and the second type of testing data as the core F1 scores of the three data sets are 96.4%, 96.7%, and 90.3% respectively. On the other hand, the system showed a significant drop while predicting the adjunct roles across the three sets of data as the adjunct F1 scores of the three data sets are 93.3%, 86.3%, and 75% respectively. The core roles which are described in [14] as more important than the adjunct ones in human life since who did what to whom is more important than where or when it happened. For that reason, core roles show less variable behavior when realized to be easily understood. On the other hand, the adjunct roles, gets its high variability not only because of its less importance in a sentence if compared to the core role, but also because they occur with most of the predicates which suggest that the better procedure that needs to be followed while working with adjunct roles is to extract their feature cumulatively form the different AVN classes and the final lexicon of the whole adjunct roles shall be used by all the AVN classes.

M. Wrong trees

The researcher found out that for achieving high accuracy with syntactic parsing, one needs to have a sufficient knowledge of the predicates semantic roles especially the core ones since the reasons behind many of the erratic trees found in the ATB are explained by the lack of availability of such knowledge for the annotators, which resulted in many confusions while parsing the trees.

N. Metaphoric roles

Some of the encountered arguments were metaphorically expressed. For example, the “aṣ-ṣaḥīfah” argument was used numerously to express the “Agent” of the “qāla” predicate. Using that argument refers actually to the journalist who wrote or said a message via that journal. The dilemma here was how should the researcher annotate the semantic role of such argument, based on its literal meaning or on the meaning it stands for. The researcher chose not to work on the argument literal meaning and to assign the semantic role of these constituents based on the intended meaning.

O. The completeness of the study’s proposed list of roles

The currents study proposed a set of semantic roles for Arabic predicates with the goal that this list covers all the semantic relations the Arabic predicates assign. The defined list proved its success since it was easy to map the FrameNet specific-frame elements to their correspondent general semantic roles the study proposed. One problem was only encountered while mapping the “Topic” FrameNet role found in most communication-related frames, which represents the topic about which a speaker is talking. Since the utterance communicated by the speaker was mapped to the “Theme1” semantic role, the “Topic” FrameNet role was assigned to “Theme2” semantic role, as each Theme represents a different semantic relation.

P. Pipeline nature of SRL system

The proposed SRL system follows a pipeline of tasks that need to be accomplished correctly first for the system to yield a good performance. The main prerequisites any SRL system is built upon is the availability of a rich morphological and syntactic analysis of the input text, the availability of word sense and anaphora coreference assignment and the availability of different types of lexicons that maps the Arabic predicates to their rolesets, provided the ontological categories of the Arabic lemmas with high coverage. The pipeline nature of the developed system is a major drawback of the current system since any missing layer of these layers would turn the SRL system work incomplete and the erratic output of any of these layers would affect the performance of the currents system.

6 Conclusion and Future work

Rule-based SRL systems are not frequent in NLU domain. The Current study is considered as an attempt to build a rule-based semantic role labeler for Modern Standard Arabic (MSA). The uses the general roles list of Agent, Theme unlike most of the Arabic SRL systems which only the Propbank tags. The developed system achieved a final F1 score of a 91.0% F1 score on the testing data.

The researcher looks up to build a robust system that accepts an input syntactic trees parsed automatically, rather than working on gold-standard syntactic trees. The researcher also aims at covering the 337 AVN classes verbs in the new system. Not only the verbal predicates shall be the point of research, but also other kinds of predicates like the nominal, and adjectival predicates. The researcher needs to build a system with higher accuracy, and this system could be integrated with other systems.

II. References

[1] R. J. Mooney, "Semantic Parsing: Past, Present, and Future," Austin: University of Texas, 2018.

[2] L. Banarescu, C. Bonial, S. Cai, M. Georgescu, K. Griffitt, U. Hermjakob and N. Schneider, "Abstract Meaning Representation for Sembanking, "in *Proc. of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pp.178-186, Sofia, Bulgaria, 2013.

[3] X. Carreras and L. M`arquez, "Introduction to the CoNLL-2005 Shared Task: Semantic Role Labeling," in *Proc. of the Ninth Conference on Computational Natural Language Learning*, pp.152-164, Ann Arbor, Michigan, 2005.
[4] L. M`arquez, X. Carreras, K. C. Litkowski and S. Stevenson, "Semantic Role Labeling. An Introduction to a special issue," *Journal of Computational Linguistics,* vol.34, no.2, pp.145-159, 2008.
[5] *ALA-LC romanization tables: transliteration schemes for non-Roman scripts*. Washington: Cataloging Distribution Service, Library of Congress, 1997
[6] J. Hurford, B. Heasley and M. Smith, *Semantics: A Coursebook*, Cambridge University Press, 2007.
[7] J. Ruppenhofer, M. Ellsworth,M. R. Petruck, C. R. Johnson and J. Scheffczyk, "FrameNet II: Extended Theory and Practice," Berkeley, California: International Computer Science Institute, 2016
[8] J. Mousser, "A Large Coverage Verb Lexicon for Arabic," Doctoral dissertation, University of Konstanz, 2013.
[9] M. Palmer, D. Gildea and N. Xue, *Semantic Role Labeling*, Morgan and Claypool Publishers, 2010.
[10] B. Levin, *English verb classes and alternations : a preliminary investigation*, University of Chicago, 1993.
[11] S. S. Pradhan, "Robust semantic role labeling," Doctoral dissertation, University of Colorado, 2006.
[12] C. J. Fillmore, "The case for case," *In collection. Bach, & H. R., Universals in Linguistic Theory*, pp. 1-88, 1968.
[13] L. Haegeman, *Introduction to government and binding theory*, 2nd ed, Oxford : Blackwell, 1999.
[14] D. Dowty, "Thematic Proto-Roles and Argument Selection," *journal of Language,* Vol. 67, No. 3, pp.547-619, 1991.
[15] J. Gruber, "Studies in Lexical Relations," Doctoral dissertation, Massachusetts Institute of Technology, 1965.
[16] R. Jackendoff, "The Status of Thematic Relations in Linguistic Theory," *journal of Linguistic Inquiry*, vol. 18, No. 3, pp. 369-411. 1987
[17] R.Valin, "Semantic macroroles in role and reference grammar," 2004.
[18] C. J. Fillmore and C. Baker, "A Frames Approach to Semantic Analysis," *In B. Heine, and H. Narrog, The Oxford Handbook of Linguistic Analysis*, 2012.
[19] S. Miller, "A Fully Statistical Approach to Natural Language Interfaces," in *Proc. of the 34th Annual Meeting on Association for Computational Linguistics*, pp.55-61, 1996.
[20] D. Gildea and D. Jurafsky, "Automatic Labeling of Semantic Roles," *journal of Computational Linguistics*, vol.28, no.3, pp. 245–288, 2002.
[21] M. Palmer, D. Gildea and P. Kingsbury, "The Proposition Bank: An Annotated Corpus of Semantic Roles," *journal of Computational Linguistics*, vol.31, no.1, pp. 71-106, 2005.
[22] K. K. Schuler, "Verbnet: A Broad-coverage, Comprehensive Verb Lexicon," Doctoral dissertation, University of Pennsylvania, 2005.
[23] M. Surdeanu, R. Johansson, A. Meyers, L. M`arquez and J. Nivre, "The CoNLL-2008 shared task on joint parsing of syntactic and semantic dependencies," in P*roc. of the Twelfth Conference on Computational Natural Language Learning*, pp.159-17, Manchester, England, 2008.
[24] M. Diab, A. Moschitti, and D. Pighin, CUNIT: "A Semantic Role Labeling System for Modern Standard Arabic," *in Proc. of the Fourth International Workshop on Semantic Evaluations*, (SemEval-2007), pp.133-136, Prague, Czech Republic, 2007.
[25] M. Palmer, O. BabkoMalaya, A. Bies, M. Diab, M. Maamouri, A. Mansouri, and W. Zaghouani, "A Pilot Arabic Propbank," *in Proc.of the Sixth International Conference on Language Resources and Evaluation*, pp.28-30, Marrakech, Morocco, 2008.
[26] M. Diab, A. Moschitti and D. Pighin, "Semantic Role Labeling Systems for Arabic using Kernel Methods," *in Proc. of ACL-08: HLT*, pp. 798-806, Columbus, Ohio, 2008.
[27] L. Shi and R. Mihalcea, "An Algorithm for Open Text Semantic Parsing," *in Proc. of the 3rd Workshop on Robust Methods in Analysis of Natural Language Data*, pp. 59-67, Stroudsburg, PA, USA, 2004.
[28] M. Maamouri, A. Bies, T. Buckwalter and W. Mekki," The penn Arabic treebank: Building a large-scale annotated Arabic corpus," *NEMLAR Conference on Arabic Language Resources and Tools*, 2004.
[29] M. Plamer, "SemLink-Linking PropBank, VerbNet, FrameNet," in *Proc. of the Generative Lexicon ConferenceGenLex-09*, Pisa, Italy, 2009.
[30] Princeton University (2010), About WordNet,Available from: https://wordnet.princeton.edu/, (accessed 28 August 2020).

[31] G. A. Miller, "WordNet: A Lexical Database for English," *journal of Commun. ACM*, vol. 38, no.11, pp.39-41, 1995.
[32] Y. Regragui, L. Abouenour, F. Krieche, K. Bouzoubaa and P. Rosso, "Applications, Arabic WordNet: New Content and New applications," in *Proc. of Global Wordnet Conference*, Bucharest, Romania, 2016.
[33] Learn Python Programming Web Site, https://www.programiz.com/python-programming, (accessed 28 August 2020).
[34] Googletrans Web Site, https://pypi.org/project/googletrans/,(accessed 28 August 2020).
[35] H. Al-Sarhan, A. Darabsh, I. Al-Husban and R. AlShalabi, "Evaluating Machine Translations from Arabic Into English and Vice Versa,"*in Proc. of ICETMS*, vol.8, Dubai, UAE, 2017.
[36] W. Croft, *Syntactic Categories and Grammatical Relations: The cognitive organization of information*, University of Chicago Press, 1991
[37] A. Schiffrin and H. Bunt(2007), "LIRICS Deliverable.D4.3. Documented compilation of semantic data categories". Available from http://lirics.loria.fr, (accessed 28 August 2020).
[38] R.Martins, "Le Petit Prince in UNL," in *Proc. of the Eighth International Conference on Language Resources and Evaluation*, (LREC'12), pp. 3201–3204, Istanbul, Turkey, 2012.
[39] A. Mansouri, M. Foster, J. D. Hwang, O. Babko-Malaya and M. Palmer, *ARABIC PropBank ANNOTATION GUIDELINES*, 2013.
[40] R. Jackendoff and P.W. Culicover, *Simpler Syntax*, Oxford University Press, 2005.
[41] M.J. Collins, "Head-driven Statistical Models for Natural Language Parsing", Doctoral Dissertation , University of Pennsylvania,1999
[42] M. Surdeanu, S. M. Harabagiu, J. Williams, and P. Aarseth, "Using Predicate-Argument Structures for Information Extraction," *in Proc. of the 41st Annual Meeting of the Association for Computational Linguistics*, vol.1, pp. 8-15, USA, 2003.
[43] J. Lim, Y. HwangY, S. Park and & H. Rim, "Semantic Role Labeling using Maximum Entropy Model," *in Proc. of the Eighth Conference on Computational Natural Language Learning (CoNLL-2004) at HLT-NAACL Boston*, pp. 122–125, Massachusetts, USA, 2004.
[44] X. Carreras, Màrquez, and Lluís, "Introduction to the CoNLL-2004 Shared Task: Semantic Role Labeling," *in Proc. of the Eighth Conference on Computational Natural Language Learning*, (CoNLL), pp. 89–97, Boston, Massachusetts, USA, 2004.
[45] V. Punyakanok, D. Roth, W. Yih and D. Zimak, "Semantic Role Labeling via Integer Linear Programming Inference," *in Proc. of the 20th International Conference on Computational Linguistics*, pp. 1346–es, Geneva, Switzerland, 2004.

III. BIOGRAPHY

Dr. Sameh Alansary: *Director of Arabic Computational Linguistic Center at Bibliotheca Alexandrina, Alexandria, Egypt.*

He is professor of computational linguistics in the Department of Phonetics and Linguistics and the head of Phonetics and Linguistics Department, Faculty of Arts, Alexandria University. He obtained his MA in Building Arabic Lexical Databases in 1996, and his PhD from Nijmegen University, the Netherlands in building a formal grammar for parsing Arabic structures in 2002. His main areas of interest are concerned with corpus work, morphological analysis and generation, and building formal grammars. He is also the head of Arabic Computational Linguistics Center in Bibliotheca Alexandrina. He is supervising and managing the Universal Networking Language project in Library of Alexandria since 1-6-2005 till now. Dr. Alansary is the co-founder of the Arabic Language Technology Center (ALTEC), an NGO aims at providing Arabic Language resources and building a road map for Arabic Language Technology in Egypt and in the Middle East. He has many scientific works in Arabic Natural Language Processing published in international conferences and periodicals, and a member in many scientific organizations: (1) Egyptian Society of Language Engineering, Cairo, (2) Arabic Linguistic Society - USA, (3) Association of Computational Linguistics - USA – Europe, (4) Universal Networking Language foundation, United Nations, Geneva, Switzerland.

Amna Helmy: *Assistant teacher at the Phonetics and Linguistics department, Faculty of Arts, Alexandria University*

She got a Bachelor of Arts in 2013 (Phonetics and Linguistics department – Faculty of Arts - Alexandria university – Alexandria - Egypt).She has graduated from Information technology institute (ITI) in 2014 intake 34.She worked as a demonstrator at Phonetics and Linguistics department , Faculty of Arts ,Alexandria university form 2013 till 2019 .She has discussed her Master thesis in 2019 about semantic role labeling. She had participated in three published papers.

نحو بناء نظام لتعيين الأدوار الدلالية للغة العربية المعاصرة: اتجاه مبني على قواعد لغوية

Amena Helmy Abd El-khaleq Muhammad Dheif[1], Sameh Saad Abu-Elmagd Alansary[2]

Department of Phonetics and Linguistics, Faculty of Arts, Alexandria University.

amena.helmy@alexu.edu.eg

Sameh.Alansary@bibalex.org

ملخص— تعيين الأدوار الدلالية هو التحليل الأوتوماتيكي للنص المدخل إلي عدد العناصر المكونة له، حيث يتكون الاقتراح من المسند ومجموعة العناصر الخاصة به، ثم يربط متعلقات الفعل بعلاقاتها الدلالية؛ أو بعبارة أخرى فإنه يسعى إلى الإجابة بشكل صحيح على سؤال من فعل ماذا لمن ولماذا ومتى وأين ولماذا .

يتمثل الهدف الرئيسي للدراسة الحالية في بناء معيّن للأدوار الدلالية يقوم على قواعد الفصحى العربية المعاصرة، وهو أول نظام لتعيين الأدوار الدلالية معتمد على قواعد لغوية يتم تطويره على الفصحى العربية المعاصرة، ويشكل الأربعون فعل الأكثر شيوعا من مجموعة شبكة الأفعال العربية (AVN) اهتمام الدراسة حاليا. وتستخدم الدراسة قائمة أدوار دلالية عامة مثل المنفذ والضحية، وتعد الأولى من حيث الاستخدام في تصنيف الأدوار العربية حيث أن الأدوار الوحيدة المستخدمة في الأنظمة السابقة هي تلك المستخدمة في مدونة PropBank. وقد حقق النظام المطور نسبة نجاح على بيانات الاختبار بمتوسط ٩١,٠٪.

الكلمات الرئيسية في ورقة البحث : ***تعيين الأدوار الدلالية، شبكة الأفعال، المسند، الأدوار الدلالية الرئيسية، الأدوار الدلالية الثانوية.***

دور القواعد اللغوية في التمييز الآلي بين معاني الحروف

خالد مصطفى أبو شبانة*[1]، أحمد عبد الغني**[2]

*1 قسم اللغة العربية، كلية الآداب، جامعة الإسكندرية، مصر

**2 قسم الصوتيات، كلية الآداب، جامعة الإسكندرية، مصر

*1kh33shabana@gmail.com

**2hmd.abdelghany@gmail.com

ملخص **— تتناول هذه الورقة الدور الذي تلعبه القواعد اللغوية للغة العربية في بناء التطبيقات الحاسوبية التي تقوم بمعالجة اللغة الطبيعية (NLP)، مثل تطبيقات الترجمة الآلية والتلخيص الآلي للنصوص، والتصحيح النحوي والإملائي، وإحصاء المفردات، فمن المشكلات التي تواجه تلك التطبيقات وهي تخص جانب الدلالة مشكلة المعاني المختلفة التي يحتملها اللفظ الواحد والتي يصعب على الحاسب التمييز بينها، والبحث يركز على ما يخص حروف المعاني العربية (Particles) وذلك لما تلعبه من دور محوري في الربط بين الجمل، فهي في اللغة بمثابة المفاصل في الجسد، وأيضًا لتعدد المعاني التي تكون لكثير من هذه الحروف مع لزومها هيئة واحدة لأنها مبنية دائمًا، وذلك يزيد الأمر تعقيدًا حيث لا يوجد شيء في هيئتها يتغير بتغير معناها، كما أن الحرف نفسه يستخدم مع الأفعال بمعانٍ مختلفة، فالأمر إذًا يحتاج إلى تحليل أعمق من الشكل، والبحث يحاول رصد القواعد اللغوية والتي قد تكون تركيبية أو سياقية والتي من الممكن أن يتم حوسبتها بحيث تمكن التطبيق من تحديد معنى الحرف في الجملة التي يقوم بمعالجتها، فقد قام الباحث بضبط بعض القواعد اللغوية والتي تساعد على تحديد معاني الحرف في السياق الوارد فيه وقام بصياغة هذه القواعد صياغة حاسوبية، ثم قام بتجربتها عمليًا من خلال تطبيق حاسوبي، يعد بمثابة أداة (SDK) يمكن دمجها في تطبيقات أخرى، وقد استمد الباحث النصوص التي تم تطبيق القواعد عليها من عدة مصادر، أولها: القرآن الكريم، والثاني: إحدى المدونات اللغوية وهي المدونة اللغوية العربية الدولية التابعة لمكتبة الإسكندرية (ICA)، والثالث: المقالات المنشورة على مواقع الإنترنت.**

الكلمات المفتاحية: **القواعد اللغوية – اللبس الدلالي - حروف المعاني – المعالجة الآلية – المدونات اللغوية – التطبيقات الحاسوبية**

١- مقدمة

المقصود بالحروف هنا حروف المعاني، وهي أحد أقسام الكلام، وسميت بذلك للتمييز بينها وبين حروف المباني وهي حروف الهجاء التي يبنى بها الكلم، وأما حروف المعاني فهي الحروف التي تدل على معانٍ، وتؤدي وظائف في الجملة، والحرف هو أحد أقسام الكلام، وقد عرفه النحاة القدماء بأنه "ما دل على معنى في غيره"[1]، أي أن معناه لا يتضح بنفسه ولكنه يحتاج إلى غيره، وللحروف دور كبير في اللغة وتأليف الكلام، فهي تقوم بوظيفة الربط بين أجزاء الجمل بل وبين الجملة وأختها حتى يتماسك النص وتلتحم أجزاؤه، كما أن وظيفة التعليق التي يقوم بها الحرف تعد من الوظائف التي تبنى عليها كثير من الجمل العربية، والتعليق بالأداة أشهر أنواع التعليق في اللغة العربية الفصحى, فإذا استثنينا جملتي الإثبات والأمر بالصيغة "قام زيد، وزيد قام، وقم", وكذلك بعض جمل الإفصاح, فإننا سنجد كل جملة في اللغة الفصحى على الإطلاق تتكل في تلخيص العلاقة بين أجزائها على الأداة[٥][7]، وإننا في كثير من الأحيان لا نستطيع فهم النص ولا معرفة المراد منه من غير تحديد وظيفة الحرف ومعرفة معناه بدقة، من أجل ذلك وقع اختيار الباحث على هذا الموضوع وهو ضبط القواعد التي تعين على فك لبس معاني الحروف من أجل المعالجة الآلية للنصوص العربية.

٢- مفهوم الحرف

اختلف مفهوم الحرف عند اللغويين القدماء والمحدثين، وذلك يرجع إلى اختلافهم في أقسام الكلام (POS)، فأقسام الكلام عند القدماء ثلاثة: اسم وفعل وحرف[8]، فالحرف هو القسم الثالث عندهم وهو ما دل على معنى في غيره، والحرف عندهم مبني لا محل له من الأعراب، ومنه ما يعمل ومنه غير العامل ولكن لا يعمل فيه غيره، أما عند

---

[٥] اللغة العربية معناها ومبناها، تمام حسان، ١٢٣.

المحدثين فقد حاول بعضهم إعادة النظر في أقسام الكلام ووضع تقسيمًا جديدًا مثل تمام حسان الذي جعل اقسام الكلام سبعة، وهي: (الاسم والصفة والفعل والضمير والخالفة والظرف والأداة)[6][7]، وقد اختلفت نظرته للحرف فأدخل فيه بعض الأسماء والأفعال التي تؤدي وظيفة الحرف النحوية وهي الربط والتعليق، وأطلق عليه اسم الأداة وعرفها بأنها: "مبنى تقسيمي يؤدي معنى التعليق، والعلاقة التي تعبر عنها الأداة إنما تكون بالضرورة بين الأجزاء المختلفة من الجملة"[7][7]، وتنقسم الأداة عند تمام حسان إلى قسمين: الأداة الأصلية: وهي الحروف ذات المعاني كحروف الجر والنسخ والعطف، وهذا ينطبق على الحرف عند القدماء، والأداة المحولة: وهي ما كان يؤدي وظيفة الحرف من الأسماء والأفعال والظروف والضمائر[7]، والباحث يميل إلى هذا التقسيم؛ لأن الكلمات عندما تؤدي وظيفة واحدة يجب وضعها تحت قسم واحد ولا داعٍ لتشتيتها بين الأقسام.

٣- القواعد اللغوية ودورها في تحديد معاني الحروف في السياقات المختلفة

إن القواعد اللغوية لها دور في تحديد المعنى السياقي للحرف، ولا يشترط أن تكون القواعد نحوية، بل قد ترجع هذه القواعد إلى ما يصاحب الحرف من كلمات ذات خصائص معينة، أو تنتمي إلى قسم معين من أقسام الكلم، فعند تحقق ذلك نستطيع أن نحدد معنى الحرف في هذا السياق أو نقوي احتمالًا معينًا ونستبعد احتمالًا آخر، فعلى سبيل المثال إذا جاء بعد "ما" "إلا" تكون "ما" في هذا السياق هي النافية، ويكون هذا الأسلوب للحصر وليس للاستثناء، وأيضًا إذا جاء بعد "إن" فعل ناسخ ثم تلا ذلك اللام التي للتوكيد فإن ذلك يقوي احتمال كون "إن" في هذا السياق المخففة من الثقيلة والتي تستخدم لتوكيد الجملة الفعلية. والباحث يحاول استخراج مجموعة من القواعد التي تساعد على تحديد معنى الحرف في سياقه، ولا شك إن هذه الخطوة من المعالجة يجب أن تسبقها خطوات أخر، فيجب أولًا تقطيع النص (Tokenization)وذلك بفصل الكلمات عما يسبقها أو يلحق بها من السوابق واللواحق، ثم وسم النص (Tagging) والمقصود به تحديد قسم الكلام (POS) الذي تنتمي إليه كل كلمة في النص، فهذه الخطوات يجب أن تتم قبل تطبيق القواعد؛ لأن هذه القواعد تعتمد على نوع الكلمات المصاحبة للحرف، فيجب أن يمر النص أولًا على أحد التطبيقات التي تقوم بتقطيع النص وتحديد أقسام الكلام ثم بعد ذلك يتم تطبيق القواعد من أجل فك اللبس عن معاني الحروف ، وقد تمكن الباحث من ضبط جملة من القواعد وتجربتها، وهذه القواعد لا تحصر جميع الحروف بكل ما تحتمله من معاني إنما نستطيع من خلالها تحديد بعض المعاني لبعض الحروف، والمجال مازال مفتوحًا لضبط قواعد أخر.

٤- قواعد تمييز معاني الهمزة

أ- ضبط همزة الاستفهام

قد تختلط همزة الاستفهام بهمزة النداء نظرًا لاتحاد الصورة "أ"، فهل نستطيع التمييز بينهما عن طريق القواعد؟ هناك مواضع انفردت بها همزة الاستفهام يمكن أن تضبطها القاعدة، فإذا وجدنا الهمزة في أحد تلك المواضع فنستطيع الجزم بأنها همزة الاستفهام ونقوم باستبعاد احتمال أن تكون همزة النداء، وهذه المواضع هي:

- إذا جاء بعد الهمزة فعل، فهي همزة الاستفهام لأن همزة النداء يجب أن يتلوها اسم، فإنك حين تقول: "أذهب أحمد إلى العمل؟" تكون الهمزة هنا للاستفهام بلا شك، ولا يمكن أن تأتي همزة النداء أبدًا في هذا التركيب؛ لأنك لا تنادي الفعل. ويمكن ضبط هذه القاعدة حاسوبيًا على النحو التالي:

"أ" + فعل = استفهام

- إذا جاء بعد الهمزة حرف عطف (الواو- الفاء – "ثم") أو حرف نفي ("لا"- "ما"- "لم") أو كلاهما معًا ("أولم"- "أوما"- "أفلا") فإنها أيضًا همزة الاستفهام، فقد قرر النحاة أن همزة الاستفهام وحدها لها الصدارة وأنها تتقدم حتى على حروف العطف وحروف النفي، كما في {أَلَمْ يَجِدْكَ يَتِيمًا فَآوَى}[8]، وفي {أَثُمَّ إِذَا مَا وَقَعَ آمَنتُم بِهِ}[9]، وفي {أَفَلَا تَعْقِلُونَ}[10]. قال المرادي في أثناء كلامه عن الهمزة: "وهي أصل أدوات الاستفهام، ولأصالتها استأثرت بأمور، منها، تمام التصدير بتقديمها على الفاء والواو و"ثم"، وكان الأصل في ذلك تقديم حرف العطف على الهمزة، لأنها من الجملة المعطوفة. لكن راعوا أصالة الهمزة، في استحقاق التصدير، فقدموها "[11][1]، وهذا يمكن ضبطه حاسوبيًا كما يلي:

---

[6] اللغة العربية معناها ومبناها، تمام حسان، ١٢٢- ١٢٤.
[7] المرجع السابق، ١٢٣.
[8] الضحى، ٦.
[9] يونس،٥١.
[10] البقرة، ٤٤.
[11] الجنى الداني، المرادي، ٣١.

"أ" + (نفي / عطف) = استفهام

- إذا جاء بعد الهمزة حرف جر ("في"- "على"- "من") أو ظرف ("عند"- "مع") مثل {أَلَكُمُ الذَّكَرُ وَلَهُ الأُنْثَى}[١٢]، {أَعِندَهُ عِلْمُ الْغَيْبِ فَهُوَ يَرَىٰ}[١٣]، "أفيكم محمد؟"، "أمنكم العقلاء؟"، ويكون الضبط الحاسوبي في هذه الحالة على النحو التالي:

"أ"+ (حرف جر/ ظرف) = استفهام

- وجود "أم" في الجملة بعد الهمزة يدل على أنها همزة الاستفهام، كما تقول: "أحضر عليٌّ أم سعيد؟". ولا تجيء "أم" مع همزة النداء، ويمكن صياغة ذلك كما يلي:

"أ" + "أم" = استفهام.

ب- ضبط همزة التسوية

قد تخرج الهمزة عن الاستفهام إلى معانٍ أخر تتحدد من خلال السياق، من هذه المعاني التسوية، وتعني أنه سيان عند المتكلم وقوع الفعل من عدمه، كما تقول: "سواء عليَّ أحضر محمد أم لم يحضر"، وكما في قوله تعالى: {سَوَاءٌ عَلَيْنَا أَجَزِعْنَا أَمْ صَبَرْنَا مَا لَنَا مِن مَّحِيصٍ}[١٤]، فكي تكون الهمزة للتسوية لابد أن يتقدمها في الجملة كلمة "سواء" وأن يتأخر عنها في الجملة "أم"، وذهب بعض النحاة إلى أنه قد يتقدمها كلمة تدل على هذا المعنى مثل ("ما أدري"- "ما أبالي") وما في معناها، ولكن رجح ابن هشام أن هذه الكلمات لا تنافي معنى الاستفهام فلا تدل على التسوية[١٥][10]، ويمكن ضبط هذه القاعدة حاسوبيًا على النحو التالي:

"سواء" + "أ" + "أم" = تسوية

٥- قواعد تمييز معاني "إن"

الحرف "إن" له عدة أنواع، فهناك "إنَّ" المشدَّدة الثقيلة، وهناك "إن" المخففة من الثقيلة، وهناك "إن" النافية، و"إن" الشرطية، وهناك "إن" الزائدة، وكلها تأخذ هيئة واحدة ، فهل من سبيل للتمييز بينها عن طريق القواعد؟ صحيح تتميز "إنَّ" الثقيلة بالتشديد والفتح، لكن إذا غاب التشكيل تصبح الصورة واحدة، ولذلك لابد من البحث عن قواعد تمكن من التمييز بين تلك الصور أو على الأقل ضبط بعضها.

أ- تمييز "إنْ" المخففة من الثقيلة

تدخل "إن" (بكسر الهمزة) المخففة من الثقيلة على الجملتين الاسمية والفعلية، ولكنها تأتي مع الفعلية أكثر، ودائمًا يصحبها اللام التي للتوكيد، وحيث وُجدت "إن" وبعدها اللام المفتوحة فيُحْكَم عليها بأن أصلها التشديد"[١٦][10]، ومن ذلك قوله تعالى: {وَإِن كَانَتْ لَكَبِيرَةً إِلَّا عَلَى الَّذِينَ هَدَى اللَّهُ}[١٧]، وقوله: {إِنْ كَادَ لَيُضِلُّنَا عَنْ آلِهَتِنَا لَوْلَا أَنْ صَبَرْنَا عَلَيْهَا}[١٨]، وكذلك قوله عز وجل : {وَإِنْ يَكَادُ الَّذِينَ كَفَرُوا لَيُزْلِقُونَكَ بِأَبْصَارِهِمْ لَمَّا سَمِعُوا الذِّكْرَ}[١٩]، وقوله: {وَإِنْ كُنْتَ مِنْ قَبْلِهِ لَمِنَ الْغَافِلِينَ}[٢٠]، ومن ذلك أيضًا: {وَإِنْ وَجَدْنَا أَكْثَرَهُمْ لَفَاسِقِينَ}[٢١]، وأيضًا قوله: {وَإِنْ كُنَّا لَمُبْتَلِينَ}[٢٢]، فنلاحظ في الشواهد السابقة أنّ "إن" تدخل على الجملة الفعلية، وأنه في الغالب يتبعها فعل ناسخ ماض، كما في ("كانت"،

---

[١٢] النجم،٢١.
[١٣] النجم، ٣٥
[١٤] إبراهيم، ٢١.
[١٥] مغني اللبيب، ابن هشام، ٦٢.
[١٦] مغني اللبيب، ٣٧.
[١٧] البقرة، ١٤٣
[١٨] الفرقان، ٤٢.
[١٩] القلم، ٥١.
[٢٠] يوسف، ٣.
[٢١] الأعراف، ١٠٢.
[٢٢] المؤمنون، ٣٠.

"كنت"، "كنا"، "كاد")، وقد يأتي مضارعًا، كما في "يكاد"، وقد يأتي غير ناسخ، كما في "وجدنا"، وقد رأينا أن اللام التي للتوكيد لم تفارقها أبدًا في كافة النماذج المتقدمة، وهذا يجعلنا نقول: إذا جاء بعد "إن" فعل، وجاءت بعدها اللام فهي "إن" المخففة من الثقيلة، والتي تأتي لتوكيد الجملة الفعلية، وأما القاعدة فتصاغ حاسوبيًا كما يلي:

إن + فعل + لام التوكيد = إن المخففة من الثقيلة المؤكدة للجملة الفعلية

ب- تمييز "إنْ" النافية

وأما "إن" النافية، فتدخل على الأسماء وعلى الأفعال، وتفيد النفي مثل "ما" النافية، وفي كثير من الأحيان يأتي بعدها "إلا" أو "لمّا" ، وهذا لا يعني أنها لا تجيء إلا هكذا، فقد تأتي بدونهما، ولكن هذا نادر، "فقد وردت "إن" النافية في القرآن في عشرة ومئة موضع كلها مقترنة بـ"إلا" أو "لما" عدا سبع آيات[٢٣][9]، ومن شواهد ذلك قوله تعالى: {إِنْ أَنْتَ إِلَّا نَذِيرٌ}[٢٤]، وقوله عز وجل: {وَإِنْ مِنْ أُمَّةٍ إِلَّا خَلَا فِيهَا نَذِيرٌ}[٢٥]، وقوله سبحانه: {إِنْ عَلَيْكَ إِلَّا الْبَلَاغُ}[٢٦]، وقد يأتي بعد "إن" النافية فعل، كما في قوله: {إِنْ أَرَدْنَا إِلَّا الْحُسْنَى}[٢٧]، وكذلك {إِنْ يُرِيدُونَ إِلَّا فِرَارًا}[٢٨]، فنستطيع القول إن "إن" إذا جاء بعدها "إلا" فهي النافية، ونصوغ القاعدة كما يلي:

"إن" + "إلا" = "إن" النافية

ولكن قد رأينا في أحد الشواهد أن "إن" قد جاءت بعدها "إلا" وحكمنا عليها بأنها "إن" المخففة من الثقيلة، وهو قوله تعالى: {وَإِن كَانَتْ لَكَبِيرَةً إِلَّا عَلَى الَّذِينَ هَدَى اللَّهُ}[٢٩]، والسبب الذي جعلنا نقول إنها المخففة من الثقيلة هو وجود لام التوكيد بعدها ("لكبيرة")، وهذا يحتم علينا إذا وجدنا "إن" وبعدها فعل أن نبحث أولًا عن اللام فإذا وجدناها فهي المخففة من الثقيلة، وإذا لم نجدها نبحث عن "إلا" فإذا وجدناها فهي النافية، وهذا الأمر يمكن برمجته حاسوبيًا.

وأما عن "لمّا" فقد ورد في التنزيل {وإن كلٌ لمّا جميعٌ لدينا محضرون}[٣٠]، {وإن كل ذلك لمّا متاع الحياة الدنيا}[٣١]، {إن كل نفس لمّا عليها حافظ}[٣٢]، والضابط الحاسوبي هنا:

"إن" + "كل" + "لمّا" = نفي

ولابد هنا أن تكون "كل" مرفوعة، أما إذا كانت منصوبة فتكون "إن" في هذا التركيب هي الثقيلة أو المخففة منها، وتنوين النصب يمثَّل بالألف التي تكون في نهاية "كل"، ويمكن تمثيل ذلك حاسوبيًا كما يلي:

"إن" + "كلا" + "لمّا" = "إنَّ" / "إنْ" المخففة منها

٦- قواعد تمييز معاني "إذا"

"إذا" ظرف للزمان المستقبل مضمَّن معنى الشرط[10]، ولها ثلاث أحوال: "إذا" الفجائية، و"إذا" الشرطية، و"إذا" الظرفية، فأما الفجائية فهي التي تدل على المفاجأة، وتدخل على الجملة الاسمية، كما في الآية الكريمة: {فَأَلْقَاهَا فَإِذَا هِيَ حَيَّةٌ تَسْعَى}[٣٣]، وأما الشرطية فتدخل على الجملة الفعلية غالبًا، وهي تربط بين جملتين الأولى جملة الشرط، والثانية جواب الشرط، فيجب أن يأتي بعد الفعل جملة فعلية أو اسمية مقترنة بالفاء، ومن ذلك قوله عز وجل: {وَإِذَا قِيلَ لَهُمْ لَا تُفْسِدُوا فِي الْأَرْضِ قَالُوا إِنَّمَا نَحْنُ مُصْلِحُونَ }[٣٤]، أما الظرفية فتشير إلى زمان وقوع الحدث، ويأتي بعدها فعل دائمًا ويكثر مجيئها بعد القَسَم. وعند تتبع الشواهد في المدونة العربية العالمية (ICA) تبين للباحث أن الاستخدام المعاصر لـ"إذا" الشرطية يأتي أحيانًا على الترتيب المعتاد: "إذا" ثم فعل الشرط ثم جواب الشرط، كما في الأمثلة :

"فإذا رفض المجلس جاء الرئيس بوزارة أخرى"

"فإذا سأل سائل فردا من هو ؟ أجاب إجابة في صيغة جمعية"

وأحيانًا يتقدم الجواب على "إذا" وفعل الشرط كما في الأمثلة التالية:

---

[٢٣] معاني النحو: فاضل السمرائي، ٣٥/٤.
[٢٤] فاطر، ٢٣.
[٢٥] فاطر، ٢٤.
[٢٦] الشورى، ٤٨.
[٢٧] التوبة، ١٠٧.
[٢٨] الأحزاب، ١٣.
[٢٩] البقرة، ١٤٣
[٣٠] يس، ٣٢
[٣١] الزخرف، ٣٥
[٣٢] الطارق، ٤
[٣٣] طه، ٢٠.
[٣٤] البقرة، ١١.

"عليه أن يتوسع إلى مناطق أخرى إذا نجحت التجربة"

"لم يستطيعوا أن يقرؤوا ، أو يكتبوا شيئا ، إذا لم تكن أمامهم أية حروف سوى حرف واحد"

أما الفجائية فلابد أن يتلوها اسم[10]، وفي الغالب إذا كان هذا الاسم ضميرًا فإنه يكون مقترنًا بباء الجر، كما نرى في تلك الشواهد:

"فإذا بها تكتشف وفاته ، فتضيق بها الدنيا"

"انتصبت واقفة ، لأسرع لنجدته .. فإذا بي أرى رجلا ينبثق من بين البيوت"

وفي بعض الأحيان يليها اسم ظاهر كما في:

"بدأ يرفس الهواء ، ثم يهبط ويرمح مرة ، مرتين .. فإذا المحراث يتفكك ويتطاير قطعا رغم أنه من الخشب القوي"

وفي كل الأمثلة السابقة نلاحظ مجيء فعل مضارع بعد الاسم أو الضمير بعدها ("تكتشف"، "أرى"، "يتفكك")، وقد يأتي بعدها ما يقوم مقام الفعل المضارع كاسم الفاعل أو الصفة المشبهة، كما في قوله تعلى: {حَتَّى إِذَا فَتَحْنَا عَلَيْهِمْ بَابًا ذَا عَذَابٍ شَدِيدٍ إِذَا هُمْ فِيهِ مُبْلِسُونَ}[35]، وكما في قول الشاعر: "قد غَفَونا وَانتَبَهنا ....فَإذا نَحنُ غَرقى وَإذا المَوتُ أَمَم"[2]. وأحيانًا يجيء بعد الاسم فعل ماض مسبوق بـ"قد" كما في المثال: "خرجنا لملاقاته على الطريق فإذا به قد أقبل والأنوار تسبقه". وهذا يعني أن أهم ما يميز "إذا" الفجائية دخولها على الجملة الاسمية، وغالبًا ما يكون المبتدأ ضميرًا وفي كثير من الأحيان في الاستعمال المعاصر يتصل بباء الجر، وأما عن الخبر ففي الغالب يكون جملة فعلية فعلها مضارع، وقد يكون الفعل ماضيًا مسبوقًا بـ"قد"، وقد يكون الخبر مفردًا.

وأما "إذا" الظرفية فتتشابه كثيرًا "مع" الشرطية، فيأتي بعدها فعل قد يكون ماضيًا كما في قوله عز وجل: {وَالنَّجْمِ إِذَا هَوَى}[36]، وقد يكون مضارعًا كما في قوله تعالى: {وَاللَّيْلِ إِذَا يَغْشَى}[37]، وقد يأتي بعدها اسم، فقد وقع الخلاف بين النحاة في "إذا" التي يتلوها اسم كما في الآية: {فَإِذَا النُّجُومُ طُمِسَتْ}[38]، والآية: { إِذَا السَّمَاءُ انْشَقَّتْ}[39].

ومن ذلك كله نستطيع أن نقول إنّ "إذا" حين يتلوها فعل تنحصر بين الشرطية والظرفية، ونستبعد كونها فجائية، فتصبح القاعدة:

"إذا" + فعل = الشرطية/الظرفية

أما إذا تلاها اسم فإننا ننظر إلى ما بعده فإن كان فعلًا ماضيًا غير مسبوق بـ"قد" فهي شرطية أو ظرفية، وإن كان غير ذلك فهي فجائية، ونصوغ القواعد كما يلي:

"إذا" + اسم + فعل ماض ليس مسبوق بـ"قد" = شرطية/ظرفية

"إذا" + اسم + (صفة/فعل مضارع/فعل ماض مسبوق بـ"قد") = فجائية

٧- قواعد تمييز معاني "ما"

إنّ "ما" من الكلمات التي تتعدد أنواعها ومعانيها، فهناك "ما" النافية، و"ما" الموصولة، و"ما" الاستفهامية، و"ما" الشرطية، وغير ذلك، وبغض النظر عن كونها اسمًا أو حرفًا فسوف ننظر إليها على أنها تؤدي وظيفة الحرف في الجملة طبقًا لوجهة النظر التي تضع كل الكلمات التي تؤدي معنى الحرف في قسم واحد وتسمّيه الأداة، ويحاول الباحث التمييز بين بعض أنواع "ما" عن طريق القواعد.

أ- ما الاستفهامية

من القواعد الظاهرة التي تميز "ما" الاستفهامية عن غيرها وجوب حذف الألف منها إذا سبقها حرف جار، قال ابن هشام: "وَيجب حذف ألف مَا الاستفهامية إِذا جُرّت وإبقاء الفتحة دَلِيلا عَلَيْهَا نَحْو "فيمَ" و"إلام" و"علام""[40][10]، ومن نماذج ذلك قوله تعالى: {فَنَاظِرَةٌ بِمَ يَرْجِعُ الْمُرْسَلُونَ}[41]، وقوله: {لِمَ تَقُولُونَ مَا لَا تَفْعَلُونَ}[42]. لكن قد تشتبه "ما" الاستفهامية إذا لحقت ببعض حروف الجر ببعض الكلمات مثل "لَمْ"، وأيضًا "عَمَّ" قد تشتبه بـ "عَمَ"، و"عَلَامَ" قد تشتبه

---

[35] المؤمنون، ٧٧.

[36] النجم، ١.

[37] الليل، ١.

[38] المرسلات، ٨.

[39] الانشقاق، ١.

[40] مغني اللبيب، ٣٩٣.

[41] النمل، ٣٥.

[42] الصف، ٢.

بـ "عَلَّام"، وهذا ليس داخلًا في نطاق هذا البحث، أما عند اتصالها بحروف الجر الأخرى كـ "في" و"إلى" والباء فيمكن تمييزها بحذف الألف منها.

ب- "ما" المصدرية أو الموصول الحرفي:

تأتي "ما" مصدرية فتدخل على الفعل فتكون هي والفعل في تأويل مصدر، مثل: {فَاتَّقُوا اللَّهَ مَا اسْتَطَعْتُمْ}[٤٣]، والمعنى: "فاتقوا الله استطاعتكم"، وأيضًا مثل قوله عز وجل: {وَدُّوا مَا عَنِتُّمْ}[٤٤]، أي: "ودوا عنتكم"، وهناك حالة واحدة يمكن تمييز "ما" المصدرية فيها آليًا، وهي عندما تقع بعد كاف التشبيه بين فعلين متماثلين[٤٥][10]، كما في قوله تعالى: {وَإِذَا قِيلَ لَهُمْ آمِنُوا كَمَا آمَنَ النَّاسُ قَالُوا أَنُؤْمِنُ كَمَا آمَنَ السُّفَهَاءُ}[٤٦]، والمراد: "إذا قيل لهؤلاء آمنوا مثل إيمان الناس قالوا أنؤمن مثل إيمان السفهاء"، ومنه أيضًا قوله تعالى: {إِنْ تَكُونُوا تَأْلَمُونَ فَإِنَّهُمْ يَأْلَمُونَ كَمَا تَأْلَمُونَ}[٤٧]، أي: "يألمون مثل ألمكم"، ومن أمثلة ذلك أيضًا، قوله تعالى: {إِنَّا أَرْسَلْنَا إِلَيْكُمْ رَسُولًا شَاهِدًا عَلَيْكُمْ كَمَا أَرْسَلْنَا إِلَى فِرْعَوْنَ رَسُولًا}[٤٨]، والفعلان السابق لـ"كما" والتالي لها لا يشترط أن يتطابقا، فقد يجيء أحدهما ماضٍ والآخر مضارع، وقد يجيء أحدهما فعل أمر والآخر ماضٍ كما في "آمنوا كما آمن"، فلابد أن يكون البحث هنا عن مادة الفعل وليس الفعل نفسه، ثم نصوغ القاعدة :

فعل +كاف التشبيه + "ما" + فعل من نفس مادة الأول = "ما" المصدرية

ت- ما النافية:

تعد "ما" النافية من حروف النفي التي تدخل على الأسماء وعلى الأفعال، ويمكن تمييزها آليًا في الأحوال الآتية:

- إذا جاء بعد "ما" "كان" وبعدها جار ومجرور ثم أتبع بمضارع مسبوق بـ"أن" فهي نافية، كما في قوله عز وجل: {مَا يَكُونُ لِي أَنْ أَقُولَ مَا لَيْسَ لِي بِحَقٍّ}[٤٩]، وقوله: {وَمَا يَكُونُ لَنَا أَنْ نَعُودَ فِيهَا إِلَّا أَنْ يَشَاءَ اللَّهُ رَبُّنَا}[٥٠]، وفي قوله تعالى: {ما كَانَ لِلَّهِ أَنْ يَتَّخِذَ مِنْ وَلَدٍ}[٥١]، وقوله: {مَا كَانَ لَكُمْ أَنْ تُنْبِتُوا شَجَرَهَا}[٥٢]، وكذلك قوله عز وجل: {وَمَا كَانَ لِبَشَرٍ أَنْ يُكَلِّمَهُ اللَّهُ إِلَّا وَحْيًا}[٥٣]، ومن ذلك أيضًا "وضع النقاط فوق حروف ما كان لها أن تكتب بدونه"، وكذلك "تتساوى في لحظات الاختيار الأصعب أحزاب ما كان لها أن توضع في سلة واحدة"، فكل هذه المواضع وغيرها قد جاءت فيها "ما" وبعدها "كان" في الماضي أو المضارع، ثم بعد ذلك "أن" مع الفعل المضارع، وهي نافية في كل ذلك بلا استثناء، فعلى هذا يمكن ضبط القاعدة:

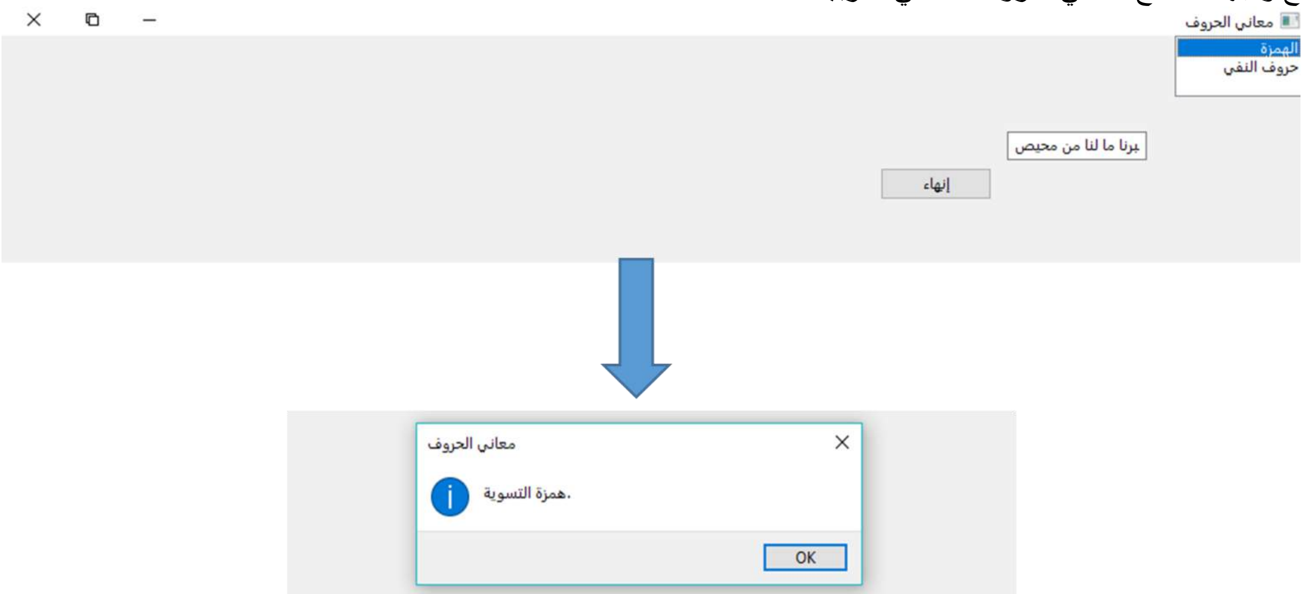
  "ما" + "كان/يكون" + "أن" + فعل مضارع = "ما" النافية
- إذا جاءت "ما" ثم تلاها "إلا" أو "غير" أو "سوى" فهي النافية، وهنا يكون الأسلوب أسلوب حصر، وهناك نماذج كثيرة لذلك منها قوله تعالى: {وَمَا مِنْ إِلَهٍ إِلَّا اللَّهُ} ٥٤، وقوله عز وجل: {وَمَا مُحَمَّدٌ إِلَّا رَسُولٌ} ٥٥، وقوله سبحانه: {وَمَا أَرْسَلْنَاكَ إِلَّا رَحْمَةً لِلْعَالَمِينَ}[٥٦]، وكذلك قوله تعالى: {مَا لَكُمْ مِنْ إِلَهٍ غَيْرُهُ}[٥٧]، وقول الشاعر: "نعيب زماننا والعيب فينا...وما لزماننا عيب سوانا". ومن أمثلة ذلك أيضًا "ما كان منه إلا أن يسلم ساقيه للريح"[٥٨]، وكذلك "ليكون مثلا يقتدي به رفاقه الذين ما كان يلقى منهم سوى الخشونة"[٥٩]، فتكون القاعدة كما يلي:

  "ما" + "إلا/غير/سوى" = "ما" النافية في أسلوب الحصر

---

[٤٣] التغابن، ١٦.
[٤٤] آل عمران، ١١٨.
[٤٥] انظر مغني اللبيب، ٤٠٠.
[٤٦] البقرة، ١٣.
[٤٧] النساء، ١٠٤.
[٤٨] المزمل، ١٥.
[٤٩] المائدة، ١١٦.
[٥٠] الأعراف، ٨٨.
[٥١] مريم، ٣٥.
[٥٢] النمل، ٥٩.
[٥٣] الشورى، ٥١.
[٥٤] آل عمران، ٦٢.
[٥٥] آل عمران، ١٤٤.
[٥٦] الأنبياء، ١٠٧.
[٥٧] الأعراف، ٥٩،٦٥،٧٣،٨٥.
[٥٨] المدونة اللغوية العربية الدولية، www.bibalex.org/ica/ar
[٥٩] المدونة اللغوية العربية الدولية، www.bibalex.org/ica/ar

وقد قام الباحث بتصميم تطبيق حاسوبي بغرض اختبار صحة القواعد التي توصل إليها في تحديد معاني الحروف آليًا، بحيث يتم إدخال جملة تحتوي على الحرف المراد تحديد معناه، ثم الضغط على الحرف بحيث يقوم الحاسوب بتطبيق القواعد المتعلقة بهذا الحرف، ثم بعد ذلك تظهر رسالة تحتوي على معنى الحرف الذي توصل إليه الحاسوب بعد تطبيق القواعد، وفي حالة عدم توافر القواعد في النص المدخل فإن التطبيق يظهر رسالة تفيد عدم إمكانية تحديد معنى الحرف، فهذه القواعد لا تحصر جميع معاني الحروف، وإنما هي تضبط جملة من تلك المعاني، والشكل التالي يوضح واجهة المعالج الدلالي لحروف المعاني العربية.



شكل (١): واجهة المعالج الدلالي لحروف المعاني العربية

٨- النتائج والتوصيات

قد أوضحت هذه الورقة إمكانية استخراج قواعد لغوية يمكن تغذية التطبيقات الحاسوبية بها لترفع قدرتها على فك اللبس عن معاني الحروف وتحديد دورها الدلالي أو على الأقل ترجيح بعض المعاني المحتملة للحرف في النص الوارد به، وقد تمكن الباحث من استخلاص بعض هذه القواعد وصياغتها صياغة ملائمة للحاسوب، كما قام بتجربة هذه القواعد من خلال تطبيق حاسوبي بسيط للتحقق من صحتها، ويوصي الباحث بضرورة مواصلة البحث في مجال استخلاص القواعد اللغوية التي يمكن أن تدعم تطبيقات المعالجة الآلية للغة العربية مما يساعد على تحسين تلك التطبيقات بشكل يخدم اللغة العربية والناطقين بها ومحبيها.

المراجع العربية

[1] ابن قاسم المرادي، الجنى الداني في حروف المعاني، تحقيق: فخر الدين قباوة، محمد نديم فاضل، دار الكتب العلمية، بيروت، ط: ١، ١٩٩٢م.
[2] ديوان حافظ إبراهيم.
[3] ديوان الشافعي
[4] أحمد بن عبد النور المالقي، رصف المباني في شرح حروف المعاني، تحقيق: أحمد الخراط، دار البشير، جدة، ط: ٣، ٢٠٠٢م.
[5] ابن يعيش، شرح المفصل، دار الكتب العلمية، بيروت – لبنان، ط: ١، ٢٠٠١م.
[6] سيبويه، الكتاب، تحقيق: عبد السلام محمد هارون، مكتبة الخانجي، القاهرة، ط: ٣، ١٩٨٨م.
[7] تمام حسان، اللغة العربية معناها ومبناها، عالم الكتب، ط: ٥، ٢٠٠٦م.
[8] ابن جني، اللمع في العربية، تحقيق: فائز فارس، دار الكتب الثقافية - الكويت.
[9] فاضل صالح السمرائي، معاني النحو، دار الفكر للطباعة والنشر والتوزيع، الأردن، ط: ١، ٢٠٠٠ م.
[10] ابن هشام، مغني اللبيب عن كتب الأعاريب، تحقيق: مازن المبارك، محمد علي حمد الله، دار الفكر، دمشق ط: ٦، ١٩٨٥م.
[11] المبرد، المقتضب، تحقيق: محمد عبد الخالق عظيمة، عالم الكتب، بيروت.
[12] المدونة اللغوية العربية الدولية، www.bibalex.org/ica/ar
[13] مجلة فكر الثقافية، https://www.fikrmag.com/article_details.php?article_id=368

السيرة الذاتية

خالد مصطفى أبو شبانة، باحث دكتوراه قسم اللغة العربية كلية الآداب جامعة الإسكندرية، موضوع رسالة الدكتوراه يدور حول حوسبة حروف المعاني العربية حصلت على الماجستير في تخصص اللغة من جامعة دمنهور سنة ٢٠١٥، أعمل بوزارة التربية والتعليم بالإسكندرية، كما أعمل مدربًا للحاسب الآلي للمكفوفين وضعاف البصر، وكذلك معلمًا للغة العربية عن بعد.

The Role of Linguistic Rules In automatic disambiguation of Particles Meanings

Khaled Mustafa Abou-shabana

PhD researcher, Arabic language department, faculty of Arts, Alexandria University

kh33shabana@gmail.com

Abstract— This paper deals with the role played by Arabic linguistic rules in designing natural language processing NLP applications Such as machine translation applications, automatic text summarization, grammar and spelling correction, and vocabulary statistics. One of the problems that face these applications, which is related to the semantic side, is the problem of the various meanings that one word can tolerate and that it is difficult for the computer to distinguish between them, and the research focuses on what is related to the Arabic particles. This is because of the central role it plays in linking the sentences. As well as particles always keep its form and don't change when its meaning changes. Also when used with a verb it give the verb more than one meaning So it needs deep analysis in order to catch its meaning. The research attempts to monitor linguistic rules that may be syntactic or contextual and that can be computerized so that the application enables to determine the meaning of the particle in the sentence that it processes, the researcher has set some linguistic rules that help to define the meanings of the particle in the context in which it was formulated and formulated these Grammar Computer Formulation, then tried it in practice through a computer application, which serves as an SDK tool that can be integrated into other applications.

Keywords: **Linguistic Rules -Semantic Ambiguity - Particles - NLP – Linguistic Corpora - Computer Applications**

Speech processing

قياس انفعالات الممثل الصوتية باستخدام تقنيات هندسة اللغة

صديقة عبد المنعم عبد الغنى أحمد لاشين

أستاذ مساعد

*قسم الدراسات المسرحية-كلية الآداب-جامعة الإسكندرية
4عمارات الطابية-السرايا-الإسكندرية
siddeeka_lasheen@yahoo.com

ملخص البحث:

عنوان البحث" *قياس انفعالات الممثل الصوتية باستخدام تقنيات هندسة اللغة* "،هدف البحث إلى تقييم أسس العملية الإبداعية وإخضاعها للبحث العلمى تماشيا مع التكنولوجيا الحديثة. فعن طريق التحليل العلمى للصوت باستخدام البرامج الحديثة يمكننا أن نحكم على مدى تحكم الممثل بصوته فى التعبير عن الانفعالات المختلفة كخطوة أساسية فى الوصول إلى حالة الإبداع. مشكلة الدراسة هى ما مدى تحقق فرضية قياس الانفعالات التى يعبر عنها الممثل عن طريق تحليل الصوت مع عدم المساس بالحالات الإبداعية . لذا وجب تقسيم الدراسة إلى ثلاثة أقسام، القسم الأول وهو بعنوان "فن الآداء التمثيلى وعلاقته بالصوت" وتتناول الباحثة فيه الصوت وعلاقته بفن التمثل، وماهية قواعد الإلقاء التى تستخدم فى مجال الأداء التمثيلى، والقسم الثانى بعنوان: الصوت وعلم هندسة اللغة، وتتناول الباحثة أهم المصطلحات المستخدمة فى مجال هندسة اللغة، أما القسم الثالث فهو الجانب التطبيقى ووقع اختيار الباحثة على نموذج أداء صوتى تمثيلى من مسرحية هاملت تأليف ويليم شكسبير، وأداء عدد من الممثلين المختلفين مكانيا وزمانيا. وتم تحليل الأصوات فيها باستخدام برنامج برات PRAAT . وينتهى البحث بعرض أهم النتائج ومناقشتها ثم التوصيات، وأخيرا قائمة المصادر والمراجع.

كلمات مفتاحية: *Keywords:* الانفعال الصوتى-هندسة اللغة- الأداء التمثيلى-قياس الصوت- .PRAAT-

مقدمة

مما لا شك فيه أن صوت الممثل هو آداة من أدواته الرئيسة للتعبير، فهو عنصر أساسى لا يقل أهمية عن الجسد فى عملية التعبير؛ حيث إنه لابد أن تتضح فيه صفات الشخصية المؤداة، وكامل تفاصيلها من سن وجنس وصفات نفسية واجتماعية وغيرها مما ذكرها المنظرون لفن المسرح، بدءأ من أرسطو حتى النظريات الحديثة، وما بعدها. فالأبعاد الأساسية التى لابد أن يعبر عنها المثل، أولها البعد الطبيعى: ويعنى بها طبيعة الشخصية من الناحية التشريحية والوظيفية. ثم البعد الاجتماعى : ويخص صفات الشخصية من الناحية البيئية والعلاقات الاجتماعية والوضع الاقتصادى، وأخيرا البعد النفسى: ويخص صفات الشخصية المؤداة النفسية"[١]

وتكمن القدرة الإبداعية للمثل فى كيفية توظيف أدواته الرئيسة (صوت- جسد) فى خلق الشخصية المسرحية. ومن هنا ظهرت مدارس تدريب وإعداد الممثل بشكل عام من ناحية، وإعداده لأداء الدور المسرحى من ناحية أخرى وذلك للوصول إلى حالة الإبداع الخلاقة. ففن الأداء التمثيلى "فن خلاّق، كفن العازف، يستخدم الممثل من خلاله كل مفردات جسمه وصوته لنفس الغرض.. لينقل خيال الكاتب بصدق وبراعة.. إنه إنسان خلاق مبدع من خلال أدواته ولغته الخاصة إذا عرف كيف يستخدمها."[٢]

ولكن هل تعتبر حالة الإبداع الخلاقة هذه حالة غير قابلة للقياس؟ حالة تحكمها ذات الممثل، أو المؤدى دون الاستناد على أسس علمية لقياسها، ومن ثم الحكم عليها؟

فمن المتعارف عليه أن الأداء التعبيرى لصوت الممثل وجسده فن إبداعى، فكم من الممثلين برعوا كل منهم فى أداء الدور نفسه، حيث يُظهر ذلك كم التدريب الذى تعرض له الممثل، ومخزون الخبرة لديه ومن ثم هذه العملية ذاتية إبداعية. ولكن كيف يمكن إخضاع هذه العملية الإبداعية للقياس؟ وهنا تظهر مشكلة الدراسة، وقد اختارت الباحثة أحد أدوات التعبير للممثل والذى يتمثل فى صوته . وذلك للأسباب الآتية:

١- أن الصوت بشكله المجرد- بعيدا عن الشخصية المؤداه- ، هو عنصر قابل للقياس. خاصة مع الثورة التكنولوجية المعاصرة.
٢- هناك بعض القواعد الحاكمة لأداء الممثل الصوتى، تلك القواعد تتمثل فى فن الإلقاء بقواعده الرئيسة من مخارج حروف، وتفخيم وترقيق، ونبر، تنغيم ...إلخ والتى يتم قياسها فى الأداء التمثيلى من خلال (عملية السمع) الأذن. وهى عملية ذاتية أيضا، تخضع لمدى إلمام مدرب الأداء التمثيلى بقواعد اللغة، فضلا عن إلمامه بقواعد فن الإلقاء، ولكن ما هو معيار القياس فى حال عدم إلمام المدرب بتلك القواعد؟ مع اطلاع الباحثة علي بعض برامج تحليل الصوت وجدت أن هذه العناصر أيضا قابلة للقياس من

خلال البرامج الإليكترونية التى تحدد شده الصوت، ونوعه، ودرجته، وكذلك قواعد النبر المتحكمة فيه. هذه البرامج تسهل على المدرب وضع معايير للحكم على مدى إتقان الممثل وتمكنه من اللغة وقواعدها، دون الاعتماد على المعايير الذاتية فى عملية التقييم.

٣- المحرك الثالث لهذه الدراسة يتمثل فى اهتمام الباحثة بالدراسات البينية التى تجمع أكثر من مجال مما يفيد ويثرى البحث العلمى.

٤- اذا كان عنصر الصوت هو عنصر قابل للقياس والحكم عليه، فإن ذلك يفيد مدربي الأداء التمثيلى للوقوف على أهم المشاكل التى يعانى منها صوت الممثل، ومن ثم إيجاد الطرق العلمية الفعالة لحل تلك المشاكل.

٥- التقييم الذاتى هو عملية تعتمد على خبرة المدرب، وحالته النفسية، وتركيزه فى تفاصيل الشخصية محل التقييم.

والتساؤل الذى يطرح نفسه هنا، ويمثل مشكلة الدراسة هو ما مدى تحقق فرضية قياس الانفعالات التى يعبر عنها الممثل عن طريق تحليل الصوت مع عدم المساس بالحالات الإبداعية، وأقصد هنا بأنه -على سبيل المثال- هناك اثنان أو أكثر من الممثلين قد برعوا فى أداء دور ماكبث ، كل بطريقته، وقد نرى أنهم على نفس القوة فى تحقيق حالة الإبداع، هذا التقييم ذاتى، ماذا لو تم إخضاع أصواتهم إلى القواعد العلمية لتحليل أصواتهم ؟ هل نظل على الرأى نفسه؟ أم يتغير؟

ليس الهدف فى هذا البحث انتفاء الدور الإبداعى للممثل، ولكن الهدف هو تقييم أسس العملية الإبداعية وإخضاعها للبحث العلمى تماشيا مع التكنولوجيا الحديثة. فعن طريق التحليل العلمى للصوت باستخدام البرامج الحديثة يمكننا أن نحكم على مدى تحكم الممثل بصوته فى التعبير عن الانفعالات المختلفة كخطوة أساسية فى الوصول إلى حالة الإبداع.

ولمحاولة الإجابة عن هذه التساؤلات، اختارت الباحثة عنوان: "قياس انفعالات الممثل الصوتية باستخدام تقنيات هندسة اللغة" ، لذا وجب تقسيم الدراسة إلى ثلاثة أقسام ، القسم الأول وهو بعنوان "فن الآداء التمثيلى وعلاقته بالصوت" وتتناول الباحثة فيه الصوت وعلاقته بفن التمثيل، وماهية قواعد الإلقاء التى تستخدم فى مجال الأداء التمثيلى، والقسم الثانى بعنوان: الصوت وعلم هندسة اللغة، وتتناول الباحثة أهم المصطلحات المستخدمة فى مجال هندسة اللغة، أما القسم الثالث فهو الجانب التطبيقى، ووقع اختيار الباحثة على نموذج أداء صوتى تمثيلى من مسرحية هاملت تأليف ويليم شكسبير، وأداء عدد من الممثلين المختلفين مكانيا وزمانيا. وسيتم تحليل الأصوات فيها باستخدام برنامج برات وهو تطبيق تحليل الإشارات الصوتية ومعالجتها PRAAT . وينتهى البحث بعرض أهم النتائج ومناقشتها ثم التوصيات، وأخيرا قائمة المصادر والمراجع.

أولا: الصوت وعلاقته بفن الأداء التمثيلى:

يقسم الآداء التمثيلى وفق ثلاث أنواع رئيسة، تتمثل فى الآتى:[٣]

١- نوع ملفوظ يعتمد علي اللفظ الصوتي المعبر مثل التمثيل الإذاعي، يدرك بالأذن
٢- نوع حركي يعتمد علي الحركة المعبرة، مثل التمثيل الصامت،والرقص يدرك بالعين
٣- نوع ملفوظ وحركي، مثل التمثيل علي خشبة المسرح والتلفزيون والسينما.

وفى هذا التصنيف، نجد أن الصوت بشكله المجرد يعد أحد دعائم فن الممثل الرئيسة ، حتى فى الفنون التى تعتمد على الحركة، فهناك بعض الهمهمات أو الأنات التى قد يعبر بها الممثل عن انفعالاته المختلفة، وينسحب الأمر على لحظات الصمت المعبرة ، واستخدام المؤثرات الصوتيه أو الموسيقى التى تهدف إلى نقل الحالات الشعورية المختلفة.

- فن الإلقاء والانفعال:

فن الإلقاء:[٤] :(Didacticism) "كلمة مأخوذة من اللاتينية وتعني الكلام، وتستعمل للدلالة على فن اللفظ طريقة الكلام أو طريقة إلقاء الشعر أو النثر."

أما الإلقاء في المسرح:

"هو فن لفظ النص المسرحي ، ومقوماته مهارة نطق مخارج الحروف وحسن استخدام نبرة الصوت ونغمته وشدته وسرعة الكلام وإيقاعه. وتتفاوت طبيعة الإلقاء في المسرح ما بين إبراز قيمة الصوت كعنصر سمعي في لفظ أقرب إلى الترتيل وهذا هو التنغيم ، وبين إبراز المعنى من خلال إلقاء طبيعي يشبه لهجة الحديث العادي ، وذلك تبعا لاختلاف مدارس التمثيل"

يقول الفنان المصري الراحل عبد الوارث عسر عن فن الإلقاء وعلاقته بالأداء التمثيلى: "إنه فن النطق بالكلام على صورة توضح ألفاظه ومعانيه، وتوضيح اللفظ يعني دراسة الممثل للحروف الأبجدية وصفات كل حرف، ليخرج من الفم سليما معافى، كاملا، واضحا، بلا عيوب لفظية أو نطقية، أما توضيح المعنى فيأتي من خلال دراسة وافية للصوت الإنساني: طبقاته، حدته، وأيضا دراسة موسيقية للأصوات من حيث النغمات التي تتناسب والمعاني (الحوار الرومانسي مثلا يتطلب صوتا ناعما هادئا).. لتأتي مطابقة لدرجة الصوت وتحدث وقعا لطيفا على أذن السامع.. وفي مجال التدريب نطلب من الممثل مثلا أن يؤدي مقطعا من مسرحية هاملت لشكسبير وبالأصوات الأربعة المعروفة: السوبرانو، الألتو، الباص، التينور.. وعلينا كممثلين هنا أن نتقيد بقواعد النطق المهمة وهي: مخارج الحروف والسكتات أو (التمبو) وهي من أبرز العناصر في أداء الجمل المسرحية الطويلة، إذ إنها تمثل الفواصل الزمانية في الحوار المسرحي أو أداء المونولوج" [٥]، [٦]، [٧]

وقد حصر فرحان بلبل فن الإلقاء فى غايات ثلاث:

إيصال المعانى التى يقصدها المتكلم، والتعبير عن المشاعر والعواطف التى يتضمنها النص

وأخيرا، كشف جماليات الأسلوب الأدبى للكلام .[٨]

ولكن قواعد النطق السليمة، وكشف جماليات الأسلوب الأدبى ليست معيارا على مدى نجاح الممثل فى التعبير عن الانفعالات المختلفة، فهذه القواعد عنصرا ضروريا للممثل، ولكنها ليست الأساس فى توصيل التعبير، وخير مثال على ذلك بعض المسرحيات أو التمثيليات الإذاعية، التى يبرع الممثلون فى استعراض مهاراتهم الصوتية بعيدا عن نقل الصوت للانفعال، فعلى سبيل المثال مسرحية هاملت الإذاعية [٩] فالأداء التمثيلى فيها لا ينقل الأنفعالات المختلفة، تحول الأداء الصوتى فيها وكأن الممثلون يقرؤن النص، ولا يؤدونه انفعاليا بالرغم من تمكنهم من قواعد فن الإلقاء.

وهنا تبرز أهمية الإيقاع وارتباطه بتدريب الصوت ، فالإيقاع (التمبو Tempo) يتكون - في المسرحية- من خلال انفعالات للممثلين، وهو سرعة الأداء أثناء الإلقاء أو الغناء أو الحوار أو الموسيقى، ويعرفه الكسندر دين على انه: "سرعة النمط الإيقاعي ويمكن وصفه بأنه سريع وبطيء أو متوسط السرعة، والتغير في التمبو لا يغير بحال في النمط الإيقاعي والأساسي، والمسرحية التي تسير على نبرة واحدة هي مسرحية مملة، رتيبة، ومع ذلك ينبغي ألا تؤدي التنويهات إلى كسر النبرة الأساسية" [١٠]

وليس المقصود بالإيقاع فى الأداء التمثيلى سرعة الممثل أو بطئه، ولكن إيقاع الشخصية الذى يتناغم مع الهدف الأعلى للمسرحية المقدمة. لذا اهتمت مدارس التمثيل الحديثة والمعاصرة بتدريبات الصوت ومدى ملاءمته للشخصية المؤداه، فضلا عن الاهتمام بخصائص الشخصية الصوتية. فيحدد الممثلون الميزات المرغوبة في الصوت ويوائمون

أصواتهم بناء على هذه الميزات. ووفق ذلك يتحتم على الممثل أن يدرس متطلبات كل مشهد على حدة.وفق الهدف العام من النص المسرحى المقدم.

ويلجأ الممثل فى تحديده للخصائص الصوتية للشخصية إلى الذاكرة الانفعالية، تلك التى تحدد الانفعالات المراد التعبير عنها بواسطة الصوت والجسد.

وتقوم الذاكرة الانفعالية غير الذاكرة الحسية على إحياء المشاعر والانفعالات السابقة، وبعثها من جديد في مواقف درامية حسب سياقها الذهني والنصي، وإعادة المشاعر كما وقعت في الماضي لإسقاطها على مواقف درامية في الحاضر فوق منصة المسرح مشابهة لتلك التي وقعت في الماضي. وفي هذا، يقول ستانسلافسكي:" إن اكتمال تجاربنا الخلاقة كلها وقوتها يتناسبان مع قوة ذاكرتنا الانفعالية، ودقتها، ومضائها تناسبا طرديا.أما إذا كانت ذاكرتنا ضعيفة، فإن المشاعر التي تثيرها تكون باهتة وهزيلة، ولا قوام لها، وتنعدم قيمتها على المنصة؛ لأنها لا تستطيع التأثير على الجماهير التي تجلس فيما وراء الأضواء الأرضية. وثمة درجات كثيرة لقوة الذاكرة الانفعالية؛ وأن كلا من مؤثراتها ومركباتها يختلف اختلافا كبيرا .[١١]

ويعني هذا أن الممثل لكي يؤدي دوره بصدق عليه أن يستذكر تجاربه الشخصية التي تعرض لها في الواقع، ويعيشها مرة أخرى، ويتذكر تفاصيلها معايشة ومعاناة. ومن ثم، يعيد تشغيل الذاكرة الانفعالية فوق الخشبة وفق متطلبات الشخصية المسرحية ويعد ذلك من أهم الآليات الفعالة لربط الدور المسرحي بالحياة، وربطه أيضا بالصدق النابع من الذات الحية. فالممثل هنا هو الوسيط بين الشخصية المؤداه، وحياته الواقعية.

وعلى ذلك فالذاكرة الانفعالية مرتبطة ارتباطا بمخزون الخبرات عند الممثل، وهذا ما يجعل الإبداع متفردا، ولكن الممثل فى تلك العملية الإبداعية، يستخدم عناصر خاضعة للمعايير الموضوعية، وقابلة للقياس، مثل قواعد فن الإلقاء، وكذلك شدة الصوت، ودرجته، وحدته...إلخ فإلى أى مدى نستطيع الحكم على مدى صدق الأداء الصوتى للممثل؟

ثانيا ـ الصوت وعلم هندسة اللغة:

تتناول الباحثة فى هذا القسم الدراسة بعض المفاهيم الأساسية لعلم هندسة اللغة، والتى تعد بالنسبة لعلماء اللغة ومتخصصى علم الصوتيات أمرا بديهيا، وأسسا متعارفا عليها، ولكن ترى الباحثة ضرورة إدراجها هنا بسبب عدم وجود مراجع فى الأداء التمثيلى تحتوى على تلك الموضوعات.

إن اللغة كائن حى يعيش ويتعايش، يؤثر ويتأثر، ويتطور كذلك، وتنقسم الدراسة الصوتية ومناهجها في العصر الحديث إلى عدة فروع، أهمها:

أولا: علم الأصوات phonetic، وهو علم خاص بالدراسة الصوتية البحتة سواء كانت فيزيائية أو أكوستيكية، من صفات الحروف ومخارجها وتشريح جهاز النطق الإنساني وتطور نطق الحروف عبر الزمن في اللغات المختلفة والدراسة الصوتية المقارنة للأصوات والحروف بين لغتين أو أكثر....إلخ، وكل هذه الدراسات يلزمها مناهج بحثية متنوعة ومختلفة، كالمنهج الوصفي، أو التاريخي، أو المقارن، أو التجريبي...

ثانيا: علم وظائف الأصوات phonology وهو علم يختص بوظيفة الأصوات داخل الكلمة ووظيفتها الدلالية، وهو أحد مستويات اللغة وفرع مكمل للنظام الصرفي ونظام تركيب الجملة في اللغة، ومحور دراسته يدور حول الفونيمphoneme وعُرِّف الفونيم: بأنه أصغر وحدة لغوية صوتية مجردة تفرق بين كلمة وأخرى، مثل الفرق بين الفعل (ضَرَبَ) فعل مبني للمعلوم، و(ضُرِبَ) فعل مبني للمجهول، عن طريق الضمة والفتحة كفونيمين يحددان معنى الفعل الثاني ويفرقانه عن نظيره.[١٢] و[١٣]

وتجمع الدراسة الحالية بين علم الأصوات، وعلم وظائفها، وربط هذه الفروع بالأداء التمثيلى.

- الهندسة اللغوية: "هى العلم الذى يبحث فى اللغة البشرية كأداة طيعة لمعالجتها فى الآلة" [١٤] و [١٥] ويهدف علم الهندسة اللغوية إلى معالجة اللغة حاسوبيا بغرض الوصول إلى برامج تطبيقية تقوم على معالجة جوانب اللغة كافة: مثل تحليل نحوى، تحليل صرفى، تحليل دلالى، ترجمة آلية، تعليم إليكترونى.... وغيرها. ويستفيد المهندسون من الطاقة الصوتيّة بتحويل الصوت إلى إشارات كهربائيّة، وإدخالها إلى أجهزة الكمبيوتر، ثم تحويلها مرة أخرى إلى طاقة صوتيّة يتم الاستماع إليها، كما يمكن استعمال مجموعة من التقنيات تختصّ بتضخيم الصوت، وتعديله، وإضافة أصوات أخرى إليه، وإعادة إنتاجه مرة أخرى.[١٦]
- تطبيقات هندسة اللغة فى الحياة الواقعية.

امتد مجال هندسة اللغة، وتحليل الصوت باستخدام البرامج الحديثة إلى عدد من المجالات الحياتية، فهناك بعض الشركات التى تستعين بتلك البرامج لاختيار موظفيها وترقيتهم؛ حيث يتعرض الموظف أو المتقدم للوظيفة إلى اختبارات لصوته، وتقوم هذه الاختبارات على تحليل تسجيل صوتي لإجابته مدته ١٥ دقيقه عن طريق الكمبيوتر. ويتضمن ذلك تحليل نغمة الصوت، واختيار الألفاظ، وتركيب الجملة، للتعرف على السمات الشخصية، مثل الاستعداد للتغيير، والحماسة، وتفهّم مشاعر الآخرين، ويلخص برنامج الكمبيوتر المستخدم في ذلك، وفي جزء من الثانية، سماتك الشخصية، من خلال مخططات، ورسوم بيانية، تكشف عن مدى دماثة خلقك، وطموحك الشخصي لتحقيق مكانة أفضل، وقدرتك على التنظيم، مقارنة بالنموذج المثالي الذي وضعته الشركة مسبقا. [١٧]

كما استخدمت تكنولوجيا تحليل الصوت فى المجال الطبى، فقد تمكن فريق من الباحثين من تطوير برنامج كومبيوتر يعتمد على الذكاء الإصطناعى (AI) قادر على المساعدة فى تشخيص إضطراب ما بعد الصدمة (PTSD) لدى قدامى المحاربين من خلال تحليل أصواتهم.

وتوصلت الدراسة - التى نشرت فى عدد إبريل من مجلة (الإكتئاب والقلق) - إلى أن أداة الذكاء الاصطناعى يمكن أن تميز بدقة تصل إلى ٨٩% بين أصوات من يعانون اضطراب ما بعد الصدمة أو كانوا من الأصحاء". وقد شارك في الدراسة نحو ٥٣ مشاركا يعانون من اضطراب ما بعد الصدمة، ونحو ٧٨ من قدامى المحاربين من الأصحاء، حيث تم ربط برامج الذكاء الإصطناعى بأنماط ميزات صوتية محددة مع اضطراب ما بعد الصدمة، بما فى ذلك الكلام الأقل وضوحا ونبرة هامدة. وفي الوقت الذى لم تكتشف فيه الدراسة الحالية آليات مرض اضطراب ما بعد الصدمة، إلا أن النظرية هى أن الأحداث الصادمة تغير إشارات المخ التى تعالج الإنفعال ونبرة العضلات، والتى تؤثر على صوت الشخص. [١٨]

كما تم استخدام برامج تحليل الصوت فى مجال حل الجرائم والقضاء، حيث يتم عرض تلك التسجيلات على خبراء الأصوات، الذين يقومون بأخذ بصمة صوت المتهمين، ومن ثم يضاهون تلك الأصوات بالأصوات الواردة بالتسجيلات، وثبت من خلال عملية المضاهاة تطابق الأصوات، وعليه يتم إعداد تقرير وافٍ بتلك التفاصيل، ويقدم لجهات التحقيق التى تستند إليه فى توجيه الاتهامات أو نفيها. [١٩]

وفى المجال الفنى، فقد استخدمت الأجهزة الحديثة فى تحليل صوت المطرب عبد الحليم حافظ، ولقد كشفت عمليات التحليل والقياس ايضا عن مرحلتين متميزتين في حياة صوت العندليب والحد الفاصل بين تلك المرحلتين هي اغنيه (نار يا حبيبي نار)قبلها لم تكن اثار المرض تبدو في التحليلات ثم بدأت هذه الاثار تتزايد ويقل معها انتاج عبد الحليم الغنائي وهذا يكشف عن المعاناة الانسانيه التي كان يمر بها ليبقى مستمرا في طريق الكفاح الفني. [٢٠]

تعددت التطبيقات الحياتية المرتكزة على مجال التحليل العلمى للصوت، وذلك فى عدة مجالات حيوية ودقيقة ولكن لم تجد الباحثة أثرا لاستخدامه فى مجال المسرح، بصفة عامة، والأداء التمثيلى بشكل خاص. وهذا ما يعنى به موضوع الدراسة. وقبل الوصول إلى تحديد برنامج التحليل الصوتى المستخدم فى العينة المختارة فى البحث، لابد لنا من التعرض إلى تعريف الصوت بالإضافة إلى أهم الخصائص الفيزيائية له.

تعريف الصوت:

الصوت هو موجة ميكانيكية تنتقل خلال وسط ما (غاز، سائلا، صلبا)، حيث يصدر الصوت عن جسم معيّن وينتقل عن طريق مجموعة من التضاغطات والتخلخلات إلى المستقبِل، ويستطيع الإنسان أن يميّز هذه الأصوات من خلال عضو السمع لديه وهو الأذن، ويوجد العديد من العوامل التي تؤثّر في قوّة انتقال الصوت، مثل طبيعة الوسط الذي ينتقل فيه وكثافته. [٢١] و[٢٢]

الخصائص الفيزيائية للصوت:

تتحدد خصائص الصوت من خلال موجته، ويتحكم بها:

أ- طول الموجة wavelength : وهى المسافة بين أية نقطة من الموجة، ونظيرتها فى الطور الذى يليها.

ب- سعة الموجة Amplitude : وهى شدة إشارة الموجة الصوتية، ويستدل عليها فى المنحنى الموجى بارتفاع الموجة، فكلما علت كلما كان الصوت أعلى.

ت- التردد أو التواتر Frequency : هو عدد الموجات التى تتجاوز نقطة معينة خلال فترة زمنية محددة، ووحدة قياسها Hz (موجة فى الثانية). ويتعلق التردد بسرعة اهتزاز مصدر الصوت، فعند زيادته يزداد تردد الصوت الصادر عنه، وكلما زاد تردد الصوت كلما كان أكثر حدة، والعكس صحيح.

أما فيما يتصل بالصوت اللغوى، فإنه يجرى التركيز على عدة أمور، أهمها المدة، التردد، السعة، والبوانى الصوتية.

أ- المدة (s): تعكس الحجم الزمنى الذى يشغله صوت معين حين نطقه.

ب- التردد: وللصوت اللغوى نوعان من التردد؛ تردد البوانى الصوتية الذى يتعلق بتكوين الجهاز الصوتى، والتردد الأساس(F0) المتعلق بالنبضات الفردية الناتجة عن اهتزاز الوترين الصوتيين خلال وحدة زمنية، وهو يعكس النبرة PITCH

ت- السعة db: ويتم تحديدها بقتامة الأشرطة، فكلما زادت شدة طاقة الصوت المعطى فى وقت وتردد معينين، كلما ازدادت قتامته.

ث- البانية الصوتية (النطاق الرنيني) : هى تركيز الطاقة الأكويستيكية حول تردد معين فى موجة الكلام. [٢٣]

تنطلق هذه الدراسة مما يطلق عليه التحليل الدلالى للصوت،

التحليل الدلالى للصوت:

إن الدلالة الصوتية تتحقق عن طريق أحد طريقين:

١- *طريق الأصوات وصفاتها التي تتألف منها الكلمة:*

وعلى سبيل المثال، الفرق بين كلمتى نضح ونضخ، فالنضح للقليل والنضخ لتدفق السائل بقوة وكثرة ومنه قوله تعالى: "فيهما عينان نضاختان"، ومن هنا اقترنت قوة المعنى والدلالة (تدفق الماء بقوة وكثرة) بصوت الخاء على حين اقترنت الدلالة الخفيفة والمعنى الأضعف (تسرب الماء بضعف) بصوت الحاء، والخاء أقوى في النطق من الحاء.

٢- *عن طريق الأداء، Intonation :*

للأداء وظائف دلالية متنوعة في كل لغة بشرية حية، وهناك كثير من الدلالات التي نستمدها من طريقة نطق الكلمة أو الجملة حسب السياق الواردة فيه، وهذه الدلالات ليس لها مقابل على مستوى المفردات والجمل، والذي حمل إلينا هذه الدلالات هو الأداء Intonation،وهو رفع الصوت وخفضه أثناء الكلام للدلالة على المعاني المختلفة للجملة أو للكلمة الواحدة، وطريقة نطق الكلمات والجمل يعطينا كثيرا من الدلالات المتنوعة للكلمة أو الجملة الواحدة.

هذا وقد اهتم علم اللغة الحديث بالتغيرات داخل الكلمات نفسها ، وشكلت موضوع علم الصرفMorphology ، والعلم الذي يختص بدراسة تنظيم الكلمات في نسق معين يشكل موضوع علم النحوSyntax، وكلا العلمين يعدان من أهم موضوعات علم اللغة linguistics، الذي يركز على اللغة نفسها، ومؤخرا تم دمج المستوى الصرفي مع المستوى النحوي لتحليل التراكيب اللغوية. [٢٤] و[٢٥] و[٢٦]

تناولت الباحثة فى هذا المبحث الإطار النظرى التأسيسي للبحث، وسوف تنتقل إلى المبحث الثانى الذى يتناول التجربة العملية تفصيلا.

المبحث الثانى: النموذج التطبيقى:

(قياس الانفعال الصوتى لشخصية هاملت باستخدام برنامج برات)

ونعرض فى هذا المبحث خطوات إجراء التجربة العملية بدءًا من أولى خطوات اختيار العينة، والبرنامج المستخدم فى القياس، مرورا بمراحل التجربة كاملة.

أولا: البرنامج المستخدم فى القياس:

يوجد عدد لا نهائي من البرامج التى يتم من خلالها تحليل الصوت، ويطالعنا العلم بصفة مستمرة على إضافة المزيد من الخصائص على البرامج التى تتعامل مع تحليل الصوت، ومن أشهر هذه البرامج برنامج PRAAT الذى يعمل على تعديل الصوت، وفصله، مع تحرير وتعديل على ملفات الصوت وتحليلها. وهذا التطبيق يستخدم فى مجال الصوتيات والفونولوجيا، ولكنه مستخدم على نطاق واسع فى مجالات أخرى تتعلق باللسانيات، مثل علم النفس، والموسيقى، والأنثروبولوجيا.[٢٧]

وباستشارة المتخصصين فى مجال الصوتيات اقترحوا على الباحثة العمل بالبرنامج نفسه ، حيث وجدوا أنه يحقق هدف دراستى، فضلا عن إنه من البرامج الدقيقة فى عملية القياس.

برات وتعنى بالهولندية "الكلام"، وهو برنامج مجانى لتحليل ومعالجة الموجات الصوتية، كتبه ويشرف عليه بول بورسما، وديفيد وينك Paul Boersma and David Weenink

من معهد علوم الصوتيات،بجامعة أمستردام- هولندا[٢٨]

ثانيا: عينة البحث:

قامت الباحثة باختيار نموذج مسرحى يتمثل فى شخصية هاملت [٢٩]من عدد من العروض المختلفة لممثلين مختلفين .

- تمثل النموذج الأساسي فى أداء الممثل "لورانس أوليفيه Laurence Kerr Olivier" [٣٠] (١٩٠٧-١٩٨٩) لمونولوج الكينونة باعتباره نموذجا للقياس. وذلك للأسباب الأتية:

أولا: ليس خافيا على المسرحيين بشكل عام، ودارسي فن التمثيل أن الممثل لورانس أوليفيه هو أفضل الممثلين على الإطلاق، وقيل عنه أنه أفضل ممثل فى العالم الانجليزى فى القرن العشرين".

ثانيا: حصل على جائزة الأوسكار كأفضل ممثل عن دوره فى فيلم هاملت –(١٩٤٨)، عينة البحث.

ثالثا: كان وراء إحياء أعمال شكسبير التاريخيه، وبادر بتقديمها على شاشات السينما.

- تنوع الاختيار بين المسرح والسينما، حيث إن الهدف هنا هو قياس الانفعال فحسب، وقد تأكدت الباحثة أن كل ما تم تقديمه لم يكنب معالجات أو رؤي درامية مختلفة عن رؤية شكسبير نفسه.

أما النماذج الأخرى التى سيتم تقييم أدائها الصوتى فهى كالتالى:

- أدريان ليستر[٣١] Adrian Lester هاملت (٢٠١٦)
- ميل جيبسون[٣٢] Mel Gebson هاملت (١٩٩٠)
- دافيد تينانت [٣٣] Daived tenant هاملت (٢٠١٠)
- أندرو سكوت [٣٤] Andrew Scotte هاملت (٢٠١٨)
- كينيث براناه [٣٥] Kenneth Branagh هاملت (١٩٩٦)

تم اختيار الجملة الأساسية من مونولوج (الكينونة) -هكذا يطلق المسرحيون عليه-. فهو المونولوج الذى يستهله هاملت بجملة (أكون أو لا أكون...تلك هى المشكلة (To be or not to be..that is the question)، وفى هذه الجملة تتجسد أعلى مستويات الصراع الداخلى للشخصية، ويتبدى القلق والحيرة، كما يظهر كامل ضعف شخصية هاملت. تم تقسيم عينة البحث وفق الآتى:

أ- لورانس أوليفيه: وهو الأساس الذى يتم القياس عليه وسوف يرمز له بحرف R Reference . بالإضافة إلى خمسة نماذج أخرى أربع منها مسرحية و الخامس سنيمائية. وهم:

١- كينيث براناه وسوف نرمز له بحرف K
٢- أندرو سكوت ونرمز له بحرف S
٣- أدريان ليستر ونرمز له بحرف L
٤- دافيد تانينت ونرمز له بحرف D
٥- ميل جيبسون ونرمز له بحرف M

ب - سوف يتم قياس أربعة عناصر رئيسة باستخدام برنامج برات وهى:

أولا: مدة الصوت Duration

ثانيا: الدرجة الصوتية pitch
ثالثا: شدة الصوت intensity
رابعا: النبر intonation

SIGNAL	MEAN PITCH	MINIMUM	MAXIMUM
Reference signal	99.96204072865116Hz	75.61388686444576 Hz	277.6546165681976 Hz
K	141.303843207913 Hz	96.37671077085244 Hz	415.5544620362446 Hz
S	136.49406993824252 Hz	91.13895729860967 Hz	490.8729153747852 Hz
L	92.79184993398123 Hz	75.412602429521 Hz	400.3779727431688 Hz
D	114.58316061493304 Hz	91.69784472234956 Hz	146.88185632767784 Hz
M	109.91871532544411 Hz	74.9046198220042 Hz	100.6300021207116 Hz

ثانيا: خطوات إجراء القياس:

أولا : مدة الصوت:

تم تقسيم الجملة إلى ثلاث مقاطع رئيسة ، تتخللهم فترتى صمت وذلك وفق طريقة تقسيم الجملة عند لورانس أوليفيه (أساس القياس)

1- To be
2- Silence
3- Or not to be
4- Silence
5- That is the question

ثانيا : الدرجة الصوتية:

وتم قياس الذبذبات الصوتية للنماذج كافة، وقد تم قياس الفرق بين أعلى تردد وأقل تردد لكل نموذج. ، فمن المتعارف عليه أن الذبذبة الصوتية تتوقف على طول وحجم الطيتين الصوتيتين، وبعد الأذن عن مصدر الصوت، وكذلك على جنس المتكلم سواء رجل أم امرأة . وكلما قلت الاهتزازات يعنى ذلك أن الصوت أكثر غلظة. والعكس صحيح. أى أنها تتناسب تناسبا عكسيا مع الصوت.

ويظهر فى الجدول رقم (١) قياس الذبذبات الصوتية لكل النماذج، وكذلك الفرق بين أقل ذبذبة وأعلى ذبذبة:

جدول رقم (١)

قياس الفرق بين أقل ذبذبة وأعلى ذبذبة

يتضح من الجدول أن أصوات كل من كينيث براناه K وكان قياسه ١٤١ هرتز، وأندرو سكوت S ١٣٦ هرتز، ودايفيد تانينت D ١١٤ هرتز وميل جيبسون M ١٠٩ هرتز وهو يعتبر وفق القياس من الأصوات المتوسطة فى الرجال والتى تتراوح بين (١٠٩، ١٦٣) هرتز عند الرجال ، وتمثلت الأصوات الغليظة فى صوت كل من أدريان ليستر L وكان قياسه ٩٢ هرتز ، ولورانس أوليفيه R ٩٩,٩ هرتز ويفسر ذلك بكونه من الطبقات الصوتية الغليظة. وبهذا تراوحت الأصوات بين المتوسطة والغليظة.

ثالثا: قياس شدة الصوت :

ويعنى بها شدة إشارة الموجة الصوتية، فكلما ازدادت القيمة كلما كان الصوت أعلى. أى أن شدة الصوت تتناسب طرديا مع الصوت. ويجمل الجدول الآتى (جدول رقم ٢) القياسات الثلاث السابقة (الفترة الزمنية، ودرجة اهتزاز الصوت، وشدة الصوت:

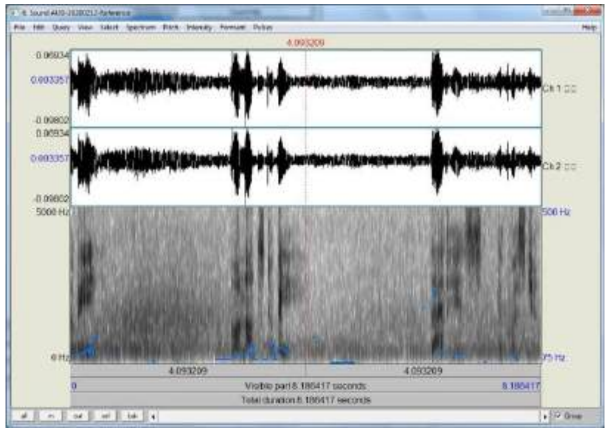
Reference	To be	Silence	Or not to be	**Silence**	That is the question
Duration	0.6266 sec	2.1576 sec	1.2338 sec	**2.3255**	2.0219
Pitch	108.42067787776764 Hz		97.31674818860985 Hz		96.52632499138801
intensity	57.201977896702516 dB		55.40933660833833 dB		52.63798422597694 dB
K	To be	Silence	Or not to be	**Silence 3**	That is the question
Duration	0.76 sec	1.79 sec	1.67 sec	**2.24 sec**	1.25 sec
Pitch	141.22745914808908 Hz		128.70262570556983 Hz		404.4665324679772 Hz
intensity	62.229461836531314 dB		51.950588301240984 dB		46.801405836345864 dB
S	To be	Silence	Or not to be	**Silence 3**	That is the question
Duration	0.855 sec	4.8 sec	2.25	**3.36 sec**	0.984
Pitch	133.66330449110515 Hz		122.8685620201669 Hz		162.5718191367407 Hz
intensity	58.992758753470234 dB		53.91157454139703 dB		53.093383257231906 dB
L	To be	Silence	Or not to be	**Silence 3**	That is the question
Duration	1.3		1.49	**2.2**	1.2
Pitch	87.13514771064078 Hz		83.71691000877189 Hz		89.28264695435283 Hz
Intensity	62.15333684856982 dB		59.330048871770245 dB		58.574580803824425 dB
D	To be Or not to be			**Silence 2**	That is the question
Duration	1.6 sec			**1.84**	1.05
Pitch	109.91871532544411 Hz				122.73726514654592 Hz
intensity	65.67681964448114 dB				65.8575646610382 dB
M	**To be Or not to be That is the question**				
Duration	2.11 sec				
Pitch	87.06416826156647 Hz				
intensity	54.00598655525319 dB				

جدول رقم (٢)

يوضح الفترة الزمنية، ودرجة الصوت وشدته

ومن المتعارف عليه أن شدة الصوت (القوة أو الضعف) يقصد بها مقدار ضخامة الصوت، واندفاعه خارج الفم بغض النظر عن طبقته متوسطة أو غليظة أو حادة. و بذلك فهى تنشأ من مقدار الرنين وقوة دفع الصوت إلى الخارج. وفيما يأتى توضح الأشكال الأتية قياس شدة الصوت لكل من النماذج المختارة.

أولا: النموذج الأساسى: R

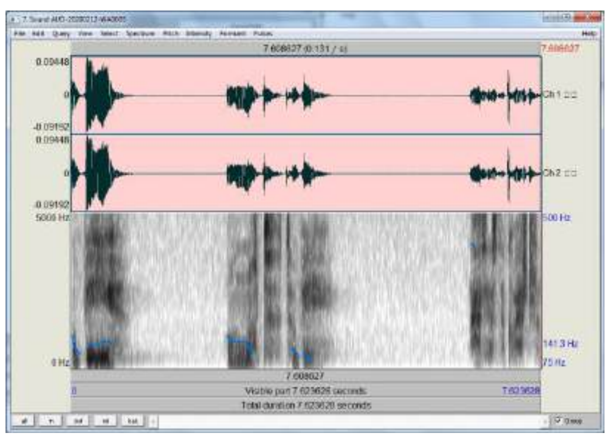


شكل رقم (١)

قياس شدة الصوت عند لورانس أوليفيه

يتضح أنه: فى نموذج القياسR لورانس أوليفيه قد تراوحت شدة الصوت بين ٥٧ ديسيبل فى المقطع الأول، فقد بلغ أعلى درجاته، وانتهى بـ ٥٢ ديسيبل فى المقطع الأخير . وهو ما يتناسب مع طريقة الأداء الصوتى لتساؤل هاملت (To be or not to be that is the question) (to be) هى التى لابد أن تكون الشدة فى هذه الجملة من حيث أنه يريد أن يكون، وقد اتخذ فترة صمت قيمتها ٢,١ ثانية لينتقل إلى المقطع الثانى من الجملة، أما كلمة (or not to be) فأتت شدتها أقل،وقد بلغت ٥٥ ديسيبل ولزم له فترة صمت أخرى ٢,٣ ثانية تمهيدا للانخفاض التدريجى فى المقطع الأخير (that is the question) ، ومن ثم تأتى أقل شدة من سابقتها من الجمل. وتعبر الشدة فى كلمة دون غيرها إحدى وسائل الوصول إلى إبراز المعنى المراد توصيله، كما أنها – وفق قواعد فن الإلقاء- تمكن الممثل من إبراز الحالة الفكرية والنفسية لشخصية المتكلم. كما أن لفترات الصمت أهميتها فى إبراز المعنى المراد توصيله، والتركيز عليه دون غيره بشكل يمنع الالتباس، ويحدد إيقاع الجملة. وبناء على ذلك فلقد جاء نموذج القياس R مستخدما لقواعد الإلقاء كافة (الارتفاع والانخفاض، فتراة الصمت، الإيقاع)

٢ - *كينيث براناه K*

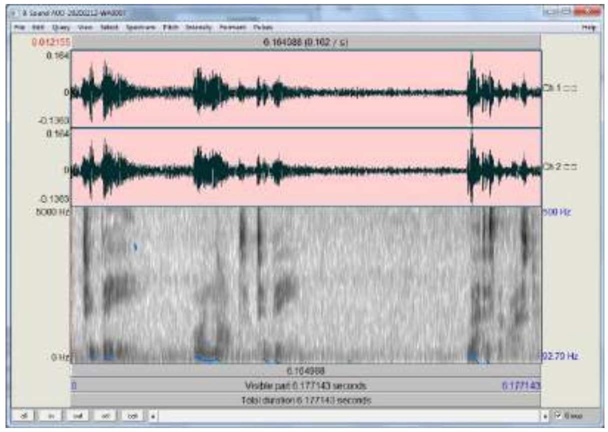


شكل رقم(٢)

شدة الصوت عند كينيث براناه

فى الشكل السابق، عند k، قد تراوحت شدة الصوت بين ٦٢ ديسيبل فى المقطع الأول، فقد بلغ أعلى درجاته، وانتهى بـ ٤٦,٨ ديسيبل فى المقطع الأخير . وقد اتخذ فترة صمت قيمتها ١,٧ ثانية لينتقل إلى المقطع الثانى من الجملة، أما كلمة (or not to be) فأتت شدتها أقل،وقد بلغت ٥١,٩ ديسيبل ولزم له فترة صمت أخرى ٢,٢ ثانية تمهيدا للانخفاض التدريجى فى المقطع الأخير (that is the question) ، ومن ثم تأتى أقل شدة من سابقتها من الجمل.

٣- أدريان ليستر L



شكل رقم(٣)

شدة الصوت عند أدريان ليستر

فى الشكل السابق، عند L، قد تراوحت شدة الصوت بين ٦٢ ديسيبل فى المقطع الأول، فقد بلغ أعلى درجاته، مثل كينيث براناه تماما وانتهى بـ ٥٨,٥ ديسيبل فى المقطع الأخير . وقد اتخذ فترة صمت قيمتها ثانية واحدة لينتقل إلى المقطع الثانى من الجملة، أما كلمة (or not to be) فأتت شدتها أقل،وقد بلغت ٥٩,١ ديسيبل ولزم له فترة صمت أخرى ٢,٢ ثانية تمهيدا للانخفاض التدريجى فى المقطع الأخير (that is the question) ، ومن ثم تأتى أقل شدة من سابقتها من الجمل.

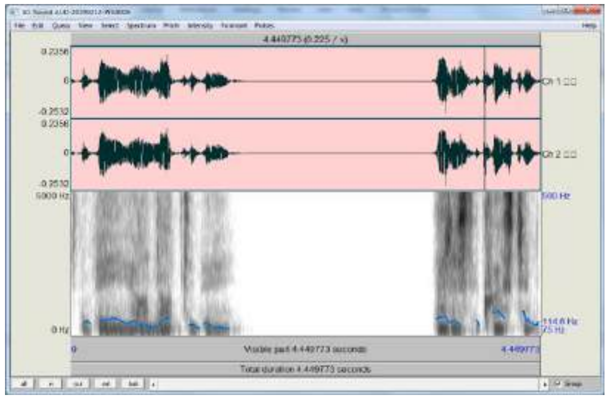
٤- أندرو سكوت S

شكل رقم(٤)

قياس شدة الصوت عند أندرو سكوت

فى الشكل السابق، عند S، قد تراوحت شدة الصوت بين ٥٨,٩ ديسيبل فى المقطع الأول، فقد بلغ أعلى درجاته، . وقد اتخذ فترة صمت قيمتها ٤,٨ ثانية لينتقل إلى المقطع الثانى من الجملة فى كلمة (or not to be) فأتت شدتها أقل،وقد بلغت ٥٣,٩ ديسيبل ولزم له فترة صمت أخرى ٣,٣ ثانية ، أما شدة المقطع الأخير (that is the question) لم تختلف عن شده المقطع الثانى حيث بلغت ٥٣ ديسيبل. لقد سجلت فترات الصمت أعلى قياسا لها فقد تعدت الفترة الزمنية للجملة المنطوقة، مما قد يصيب المشاهد بالملل، خاصة وأن شدة المقطعين الأخيرين لم يحدث اختلاف بيهما، فلا يوجد تنويع فيما بينهما.

٥-دايفيد تانينت D

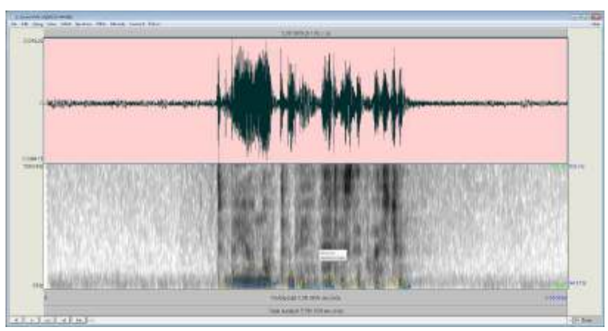


شكل رقم (٥)

شدة الصوت عند دايفيد تانينت

قسم تانينت الجملة إلى مقطعين صوتيين فقط الأول (To be or not to be) ثم المقطع الثانى (that is the question) وجاءت شدة الصوت فى كلا المقطعين واحدا بلغ ٦٥ ديسيبل. مما يدل على تساوى شده كل الجملتين، وهو ما يتنافى مع التنويع الصوتى المطلوب للأداء. كما أنه فصل بين المقطعين بفاصل زمنى مدته ١,٨٤ ثانية، إلا أنه –كما ذكرنا- لم ينوع فى طبقة الصوت بين كل من المقطعين.

٦ - *ميل جيبسون M*



شكل رقم(٦)

شدة الصوت عند ميل جيبسون

أما فيما يتعلق بشدة الصوت عند ميل جيبسون فلم يقم بتقسيم الجملة، وإنما قام بأداءها جملة واحدة، وجاء قياسها ٥٤ ديسيبل، وهو كذلك يتنافى مع المعنى التنويع المراد توصيله، فكل مقطع من الجملة لابد أن يتم التركيز فيه على حرف، أو تمييز كلمة عن أخرى.

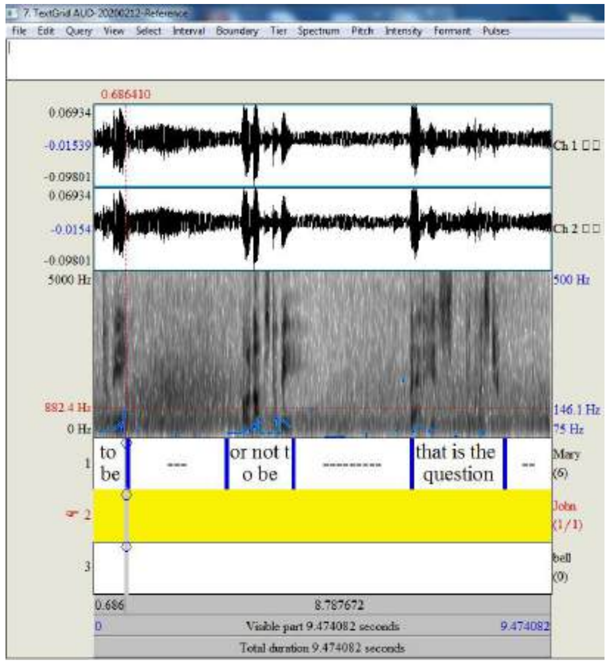
ويظل مقياس شدة الصوت ليس معيارا للحكم على طريقة الأداء ولكنه مرحلة أولية لابد من القيام بها، ومن ثم وجب القيام بقياس النبر ارتكازا على قياس شدة الصوت.

رابعا: قياس النبر:

وتبنى هذه المرحلة على المراحل السابقة، ويتحدد عن طريقه التنغيم intonation ، "فالتنغيم هو من الفونيمات فوق التركيبية أو الإضافية التى تصاحب نطقنا للكلمات والجمل، ويرتبط الارتفاع والانخفاض بتذبذب الوترين الصوتيين الذان يحدثان النغمة الموسيقية" ، وفى اللغة إجمالا يمكن تقسيم وظائف التنغيم إلى ثلاث وظائف : أولهما وظيفة نحوية تحدد الإثبات أو النفى فى الجملة، ووظيفة صوتية ويعنى بها خلق نسق صوتى مميز، تشترك فيه النغمة والمسار اللحنى فى تحديد موسيقى

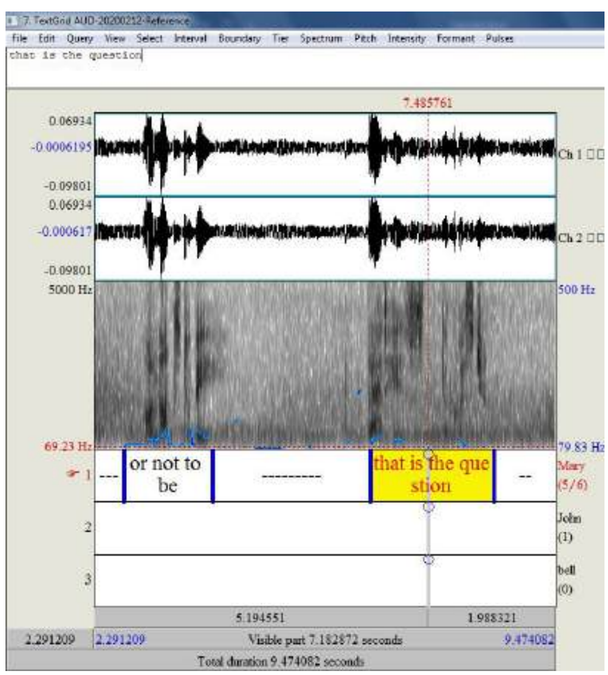
الكلام، وثالثا: وظيفة دلالية، أى الكشف عن المعنى والمشاعر من خلال علو الصوت وانخفاضه." [٣٧] وينتج عنه بالضرورة- فى الأداء التمثيلى- إيصال المعنى للسامع، وذلك عن طريق النبر على كلمات محددة دون غيرها فى المقطع الصوتى وفيها يتم تحديد أعلى درجة للنبر، وأقل درجة له فى كل مقطع صوتى، وذلك لمعرفة مناطق التنغيم فى المقطع، ومدى توافقه مع النموذج القياسى من عدمه. وسوف تعرض الباحثة قياسات لكل نموذج من النماذج المختارة.

أولا: نموذج القياس R لورانس أوليفييه:



شكل (٧)

قياس أعلى نبر فى نموذج القياس



شكل (٨)

قياس أدنى نبر فى نموذج القياس

يظهر الشكل أنه على مدار الجملة كلها فإن أعلى قمة للنبر peak of intonation كانت في مقطع "to be" حيث يمكن قياس head of intonationعند ٨٨٢,٤ HZ

كما أن فإن أدنى قيمة للنبرفي الشكل الثاني لنفس الجملة كانت في مقطع "that is the question" حيث يمكن قياس intonationعند 69.23HZ

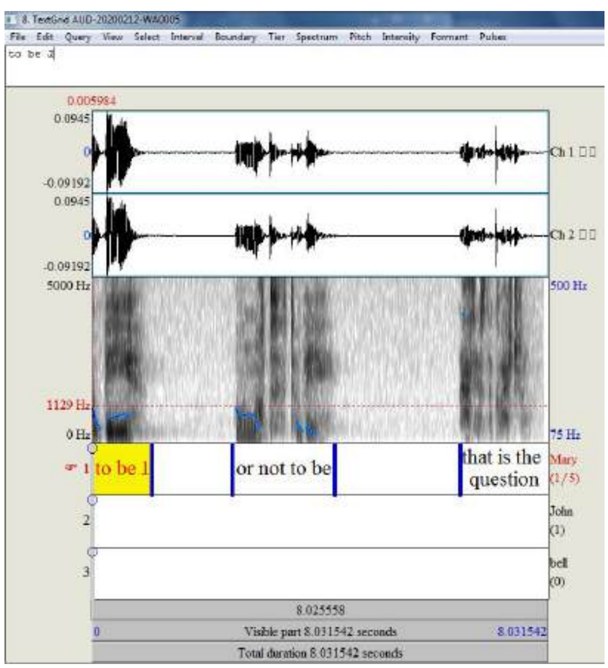
	To be	Or not to be	That is the question
أعلى نبر	HZ 882.4	660.7 Hz	414.2 HZ
أقل نبر	192.4 HZ	93.87 HZ	69.23HZ

جدول رقم(٣) قياس النبر عند لورانس أوليفييه

وبقياس منحنى النبر للجملة فى إجمالى الفترة الزمنية وهى ٩,٤ ثانية، وجدنا أن أعلى نبر جاء فى كلمة(to be) حيث بلغ ٨٨٢,٤ هرتز، وجاء أقل نبر فى مقطع (that is the question) حيث سجل ٦٩,٢ هرتز. ومن ثم فإن التنغيم هنا إذا عبرنا عنه بمنحنى، فيظهر أعلى درجة له فى المقطع الأول، وأدنى درجة فى المقطع الأخير. وهذا يعبر بصفة أساسية على أن الكلام انتهى؛ فالانعطاف الصوتى يقصد به الاتمرار أو عدم الاستمرار بالصوت حتى نهاية الكلمة أو الجملة، فإذا كان الاستمرار بالصوت بالاتجاه الصاعد، يعنى أن للكلام بقية حتى لو توقفنا بعد ذلك، وإذا اتجهنا بالصوت إلى النزول فذلك يعنى أن ليس للكلام بقية، وأن الفكرة قد انتهت.[٣٨] ويتمثل منحنى النبر هنا فى (صاعد-هابط-هابط) مما يدل على اكتمال الفكرة.

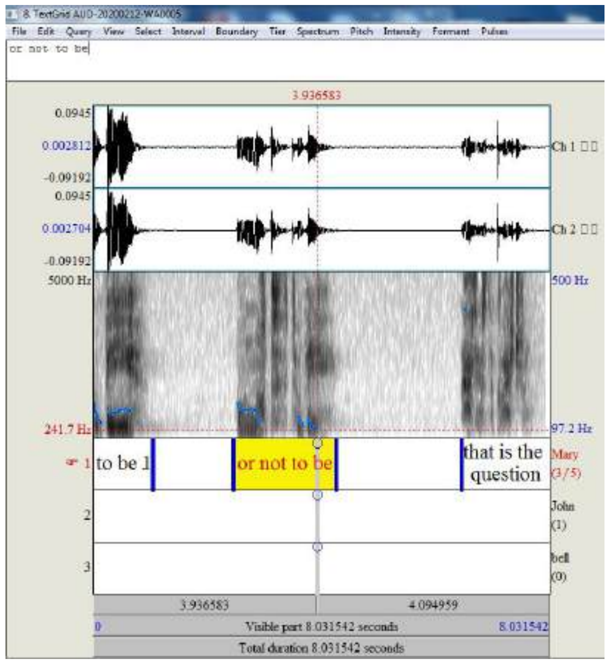
وهو ما يتناسب مع طريقة الأداء الصوتى لتساؤل هاملت (To be or not to be that is the question) ففى هذه الجملة لابد أن تتضح ملامح القلق والحيرة، والغوص فى النفس، والتساؤلات الحائرة، ففكرة كينونته فى كلمة (to be) هى التى لابد أن يكون التركيز علي هذه الجملة من حيث أنه يريد أن يكون وأن يعيش، وقد اتخذ فترة صمت قيمتها ٢,١ ثانية لينتقل إلى المقطع الثانى من الجملة، أما كلمة (or not to be) فأتت شدتها أقل، وهو ما يتسق مع حيرته، وخوفه من كونه (لا يكون) ولزم له فترة صمت أخرى ٢,٣ ثانية تمهيدا للانخفاض التدريجى فى المقطع الأخير (that is the question) ليجسد المشكلة الحقيقية التى يعانى منها هاملت، وخوفه الشديد من تلك الحياة، ومن أفعاله التى يقدم عليها ، ومن ثم فإنه يغوص أكثر داخل نفسه وكذلك يشعر بالرعب الداخلى لما هو مقدم عليه، ومن ثم تأتى أقل شدة من سابقتها من الجمل. ووفقا للشكلين السابقين، فإن التنغيم يتكون فى الأساس من: شدة الصوت (التى انخفضت تدريجيا) ، ثم اختلاف درجة الصوت (التى تباينت بين المقاطع الثلاث من الجملة) هذا التنغيم قد أعطى مضمونا للمعنى المراد التعبير عنه.

٢ - *كينيث براناه:*



شكل رقم (٩)

قياس أعلى نبر عند كينيث براناه



شكل رقم (١٠)

قياس أدنى نبر عند كينيث براناه

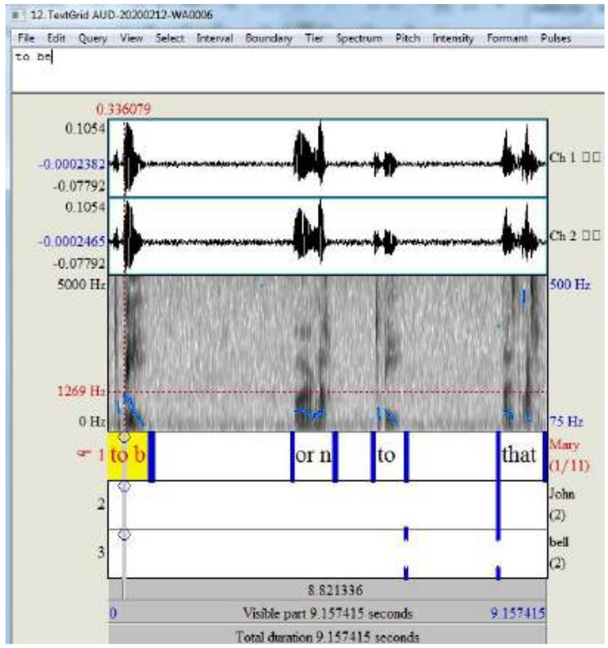
يظهر الشكل أنه على مدار الجملة كلها فإن أعلى قمة للنبر peak of intonation كانت في مقطع "to be" حيث يمكن قياس head of intonationعند 1129HZولكن في بداية المقطع كما أنه، فإن أدنى قيمة للنبر في الشكل الثاني لنفس الجملة كانت في مقطع "or not to be" حيث يمكن قياس النبر عند 241.7HZ. بينما القطع الأخير لا يكاد يظهر له نبر من ضعفه . فمنحنى النبر هنا (صاعد –هابط- لا قياس)

	To be	Or not to be	That is the question
أعلى نبر	1129HZ	241.7HZ	---
أقل نبر	HZ 398.6	HZ 266.4	----

جدول رقم(٤) قياس النبر عند كينيث براناه

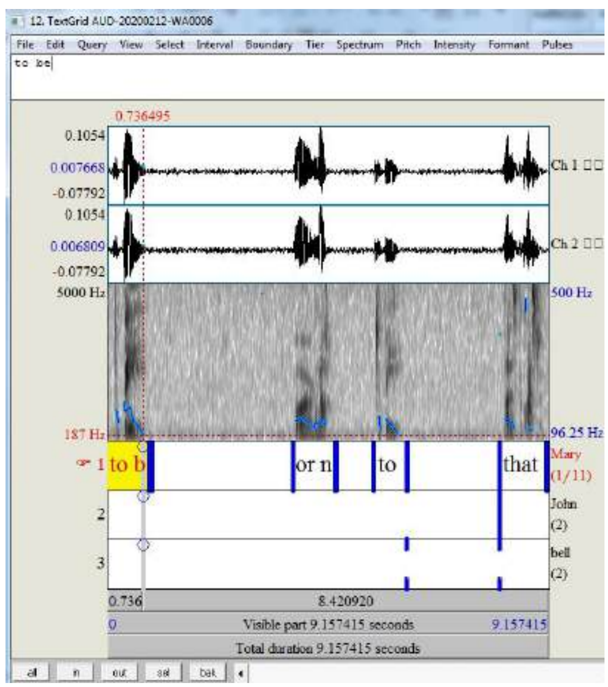
وبناء على ذلك ففى اجمالى الفترة الزمنية ٨,٠٣ ثانية نجد أن كينيث براناه قد اتفق مع لورانس أوليفيه فى تسجيل أعلى نبر فى المقطع الأول، حيث بلغ ١١٢٩ هرتز. ولكنه اختلف معه فى باقى مقاطع الجملة. فالتساؤل الحائر فى نهاية الجملة لم يظهر، وهذا يعنى عدم اكتمال إيصال معنى هذا التساؤل للمستمع.

٣- أندرو سكوت:



شكل رقم (١١)

أعلى نبر عند أندرو سكوت



شكل رقم (١٢)

أقل نبر عند أندور سكوت

يظهر في الشكل أنه على مدار الجملة كلها فإن أعلى قمة للنبر peak of intonation كانت في مقطع "to be" حيث يمكن قياس head of intonation عند 1269HZ

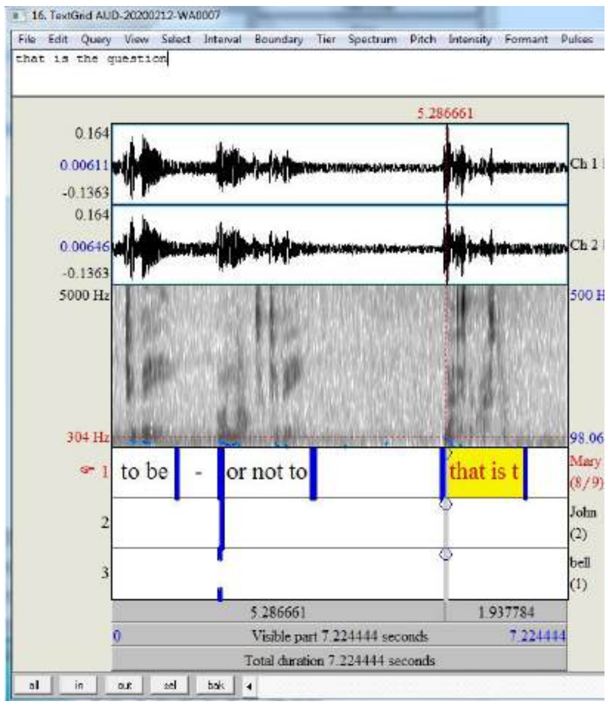
كما أن فإن أدنى قيمة للنبرفي الشكل الثاني لنفس الجملة كانت في مقطع "to be" حيث يمكن قياس intonation عند 187HZ

	To be	Or not	to be	That is the question
أعلى نبر	1269 HZ	742.6 Hz	1014 Hz	760.7 HZ
أقل نبر	187 HZ	420.9 HZ	279.8	298.7HZ

جدول رقم(٥) قياس النبر عند أندرو سكوت

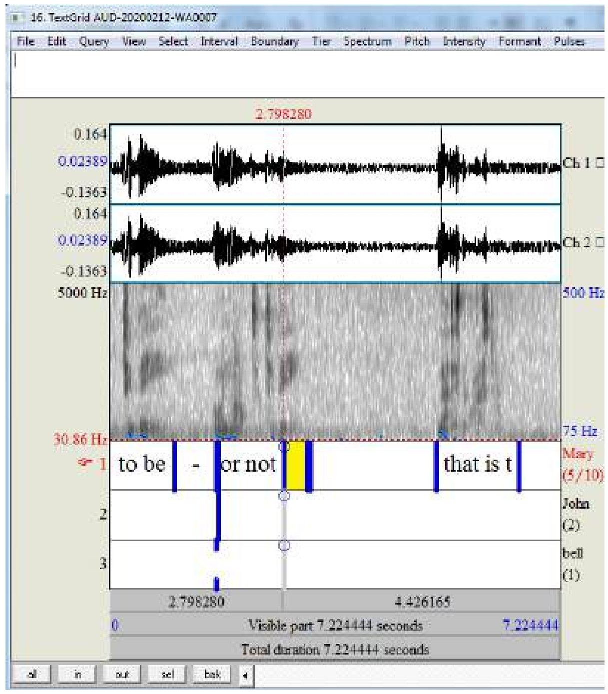
ومن خلال ذلك ففى اجمالى فترة زمنية ٩,١ ثانية، نجد أن أندرو سكوت قد اتفق مع نموذج القياس فى تسجيل أعلى نبر فى المقطع الأول حيث بلغ ١٢٦٩ هرتز ولكنه اختلف عنه فى نهاية المقطع نفسه حيث سجل ١٨٧ هرتز، وبتتبع باقى الجملة نجد أن منحنى التنغيم يتركز فى كلمة to be، سواء فى المقطع الأول أو المقطع الثانى. مما يعطيهما القدر نفسه من الأهمية، ومنحنى التنغيم هنا (صاعد-هابط-صاعد-هابط) كما كانت فترات الصمت مبالغ فيها، خاصة بعد المقطع الأول دون مبرر واضح.

٤- أدريان ليستر:



شكل رقم (١٣)

قياس أعلى نبر عند أدريان ليستر



شكل رقم (١٤)

قياس أقل نبر عند أدريان ليستر

يظهر الشكل أنه على مدار الجملة كلها فإن أعلى قمة للنبر peak of intonation كانت في مقطع "that is the question" حيث يمكن قياس head of intonation عند 304HZ

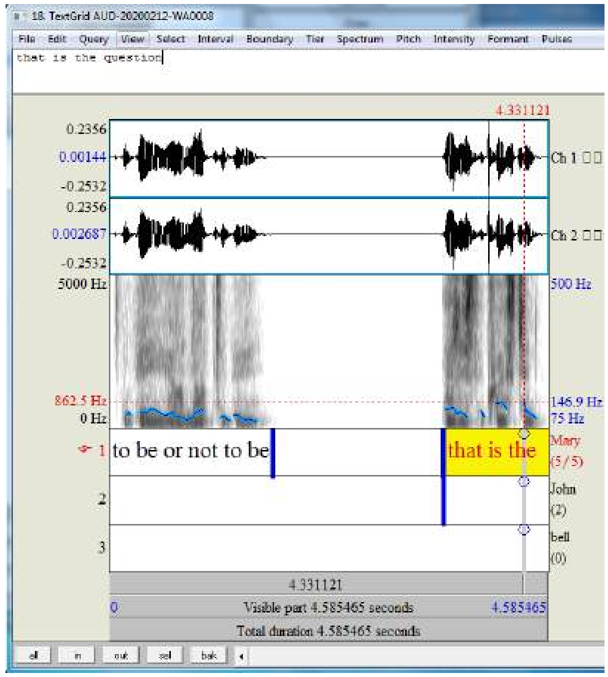
كما أن فإن أدنى قيمة للنبرفي الشكل الثاني لنفس الجملة كانت في مقطع "or not to be" حيث يمكن قياس intonation عند 30.86HZ. وجاء منحنى النبر (صاعد- هابط- أكثر صعودا)

	To be	Or not to be	That is the question
أعلى نبر	298.2	209.1	304
أقل نبر	90.26	30.86	60.56

جدول رقم(٦) قياس النبر عند أدريان ليستر

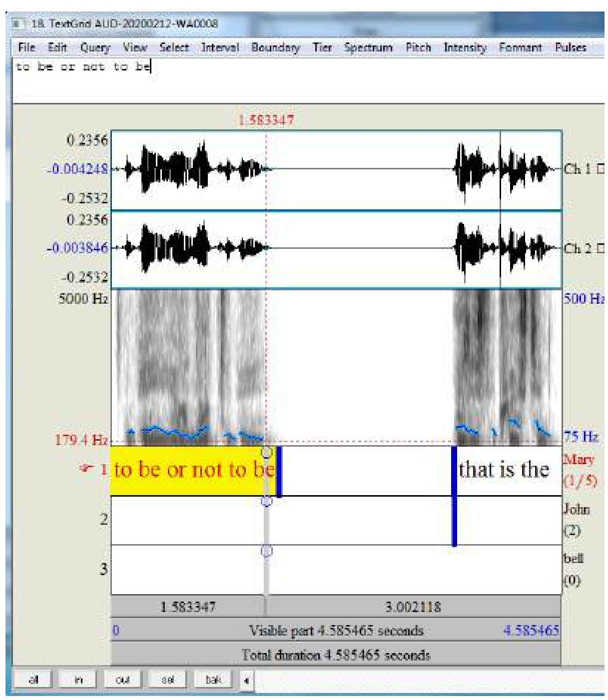
إجمالى الفترة الزمنية ٧,٢ ثانية، نجد أن أدريان ليستر قد اختلف تماما عن لورانس أوليفيه، حيث سجل أعلى نبر عنده فى المقطع الأخير من الجملة وقد بلغ ٣٠٤ هرتز. وكان أقل نبر فى الجملة فى المقطع الثانى "or not to be" حيث سجل ٣٠,٨ هرتز. وهنا أدريان ليستر لم يعطى أهمية لفعل الكينونة، وكأن المعنى المراد توصيله هنا هو التركيز على التساؤل، وليس الفعل. وهذا بدوره يختلف عن المعنى الأصلى المراد توصيله. كما يدل على عدم اكتمال المعنى المراد توصيله، فالنهايات الصاعدة –كما ذكرت الباحثة آنفا- تعنى أن معنى الجملة لم يكتمل. ولابد من استكماله بكلمات أخرى.

٥-دافيد تانينت:



شكل رقم (١٥)

قياس أعلى نبر عند دايفيد تانينت



شكل رقم (١٦)

قياس أدنى نبر عند دايفيد تانينت

يظهر الشكل أنه على مدار الجملة كلها فإن أعلى قمة للنبر peak of intonation كانت في مقطع "that is the question" حيث يمكن قياس head of intonation عند 862.5HZ

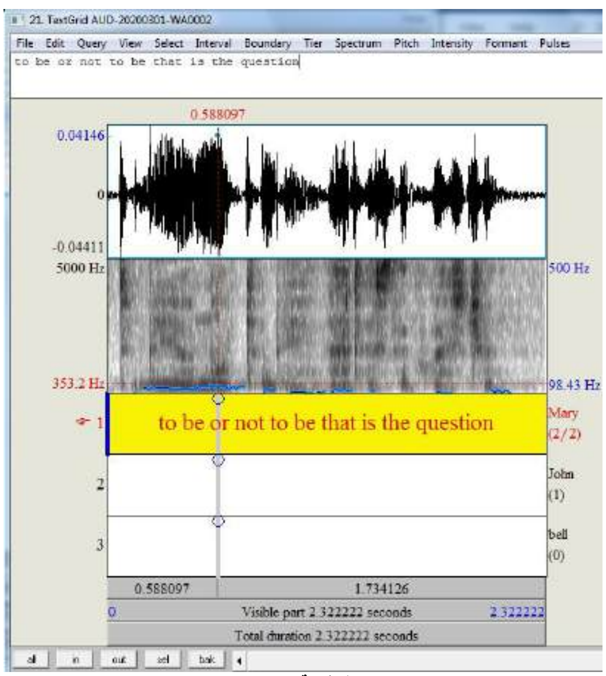
كما أن فإن أدنى قيمة للنبرفي الشكل الثاني لنفس الجملة كانت في مقطع "to be or not to be" حيث يمكن قياس intonation عند 179.4HZ. ومنحنى النبر هنا (هابط - صاعد)

	To beOr not to be	That is the question
أعلى نبر	654.5	862.5
أقل نبر	179.4	357.6

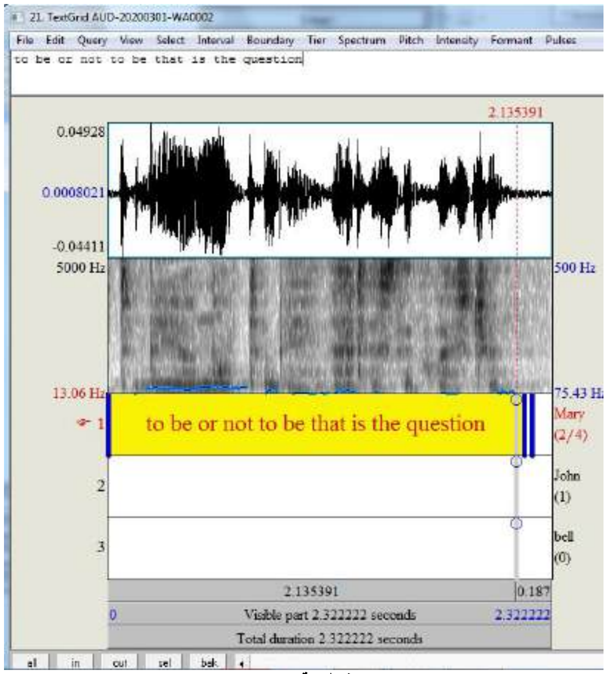
جدول رقم(٧) قياس النبر عند دايفيد تانينت

ومن خلال القياس السابق، الذى سجل تقريبا نصف الفترة الزمنية التى وجدت فى نموذج القياس، فإجمالى الفترة الزمنية هنا ٤,٥ ثانية، وقد قسمت الجملة إلى مقطعين هما to be or not to beK و that is the question. وقد سجل دايفيد تانينت أعلى نبر فى المقطع الثانى من الجملة، وهى التساؤل حيث بلغ ٨٦٢,٥ هرتز، وجاء أقل نبر فى المقطه الأول حيث سجل ١٧٩,٤ هرتز. أى أن منحنى التنغيم عنده سار عكس ما قام به النموذج القياسى فقد بدأ منخفضا، ثم ارتفع بشكل مفاجيء فى المقطع الثانى، مما يدل على عدم اكتمال المعنى، فهل هاملت هنا كان غارقا فى التفكير؟ أم أن دلالة ذلك هو استيقاظه فجأة وكأنه وجد إجابة الحيرة لديه؟ بتحليل المونولوج الأساسي كما كتبه شكسبير، فإن هذه الجملة دون غيرها لابد أن تظهر مدى عمق التفكير، والاستغراق فيه، ومن ثم تظهر مدى الضعف الذى يعانى منه هاملت. لذا يمكننا القول أن القياس هنا أظهر معنى آخر للجملة ، وهو مالا يتفق مع الهدف الأساسى من الجملة الأساسية.

٧- *ميل جيبسون:*



شكل رقم (١٧)
قياس أعلى نبر عند ميل جيبسون



شكل رقم (١٨)
قياس أقل نبر عند ميل جيبسون

تظهر الجملة في تتابع واحد دون فصل وأعلى نبر ٣٥٣,٢ HZ وأدنى قيمة له عند النهاية بلغت ١٣,٠٦ HZ .

سجلت الفترة الزمنية هنا أقل فترة زمنية للجملة، حيث استغرقت ٢,٣ ثانية، سجلت أعلى نبر فى بدايتها حيث بلغ ٣٥٣,٢ هرتز، وأقل نبر فى نهايتها ١٣,٠٦ هرتز. وقد اختلفت من حيث الشكل أو المضمون عن الجملة الأساسية، فإيقاعها جاء متسارعا، ولم يرتكز النبر على جزء أو مقطع دون غيره، ومن ثم فإن دلالة ذلك أن المعنى الذى تم توصيله، لا ينطوى على تأمل، أو استغراق فى التفكير. فلم يهدف ميل جيبسون هنا إلى طرح المشكلة على السامع، وإشراكه فيها. ولكن يبدو أنه قد وصل إلى النتيجة مسبقا.

أهم النتائج والتوصيات

وبناء على ما سبق يمكننا القول إن أهم نتائج القياسات جاءت كما يلى:

أولا:إمكانية إخضاع الانفعالات الصوتية للممثل للقياس، بل خلصت الباحثة أن ذلك يعتبر ضروريا سواء فى عملية اختيار الممثل لشخصية بعينها، أو التدريب عليها، وصولا لمرحلة العرض.

ثانيا: تراوحت تصنيف الأصوات ما بين الأصوات المتوسطة والغليظة. ولإذا كان هذا المعيار ليس حاكما فى قياس الانفعالات المنوط بالممثل تأديتها، إلا أنه معيار مهم لفى ختيار الممثلين لشخصيات بعينها، فشخصية هاملت وفق الرؤية الكلاسيكية لا ينبغى أن يؤديها ممثل ذو طبقة صوتية حادة؛ حيث لا يتناسب ذلك مع كونه يمثل الأمير الرومانسى الحالم (النموذج المثالى للأمير وفق مقاييس العصر الرومانسى وتحقيقا لمواصفات البطل التراجيدى) .

ثالثا: فيما يتعلق بقياس شدة الصوت إلى أن كل من كينيث براناه K، و أندرو سكوت S، وأدريان ليستر S يتشابهون مع لورانس أوليفيه R، ليس بنفس درجة قوة الصوت، ولكن فى طريقة التناول. فكل منهم قد بدأ عاليا فى المقطع الأول، ثم انخفض تدريجيا حتى وصل إلى أقل شدة فى المقطع الأخير. وقد اختلف عنهم كل من دايفيد تانينت D الذى لم يستخدم التنويع إلا فى فترة الصمت، وجاءت ا لجملة بنفس درجة القوة أما ميل جيبسون M فجاءت شدة الصوت ثابتة فى الجملة كاملة.

رابعا: اختلفت قياسات النبر لكل نموذج من النماذج المختارة عن نموذج القياس، مما أدى ذلك بدوره إلى تغيير المعنى المراد توصيله ، فقياس منحنى النبر عند كينيث براناه أثبت إخفاقه فى إيصال معنى مكتمل للجملة، حيث أنه بالقياسات العلمية اختفى النبر تماما فى المقطع الثالث للجملة، مما يعنى ذلك عدم اكتمال معنى للجملة. أما أندرو سكوت فابالرغم من تنوع شدة صوته، إلا أنها لم تأت متوافقة مع النموذج الأساسى، فقد سجل أعلى نبر فى مقطعى To be (الأول والثانى) بالرغم من تناقض المعنى

بينهما فالأولى كانت للإثبات، والثانية للنفى ومن ثم من خلال القياس لم يظهر هذا الاختلاف. وتأتى قياسات أدريان ليستر معاكسة تماما لنموذج القياس، حيث بدأ هابطا، ثم تصاعد تدريجيا وانتهت الجملة بمنحنى صاعد، مما يعنى ذلك أن تركيز الممثل لم يكن فى إبراز فعل الكينونة، ولكن فى التساؤل فى نهاية الجملة، الأمر الذى لابد أن يستتبعه توضيح حيث إن المعنى لم يكتمل. ويتكرر الأمر نفسه عند دايفيد تينانت حيث انتهى صاعدا . أما النموذج الأخير ميل جيبسون فقد ابتعد تماما عن التنوع فى الشدة أو التركيز على كلمة بعينها، وجاء نطق الجملة على وتيرة واحدة.

خامسا: تنوع استخدام فترات الصمت للتمهيد للانفعال التالى، صعودا أم هبوطا.

سادسا: العمل ببرامج التحليل الصوتى للممثلين هو ضرورة لا غنى عنها، حيث تخضع عملية التقييم إلى أسس علمية معيارية ينبغى القياس عليها، فالانطباع الذى يخلقه الصوت فى نفس المدرب هو معيار لا يمكن توحيده، أو القياس عليه.

التوصيات:

١- ضروة استخدام الوسائل العلمية لقياس الأداء الصوتى للممثل، للحكم على طريقة التدريب أو الآداء.
٢- ضرورة تزويد قاعات التدريب التمثيلى بفنى صوتيات لمتابعة عملية التقدم فى التدريبات الصوتية .
٣- العمل على توجيه مدربى التمثيل فى الأكاديميات المختلفة لعمل بحوث ودراسات معملية للصوت وتدريبه بالاشتراك مع أقسام الصوتيات فى الجامعات المختلفة.
٤- نشر المعرفة لدى مدربى التمثيل ببرامج تحليل الصوت الحديثة، كخطوة للانطلاق فى استخدامها فى تحليل الصوت.

المصادر والمراجع

١- أرسطو طاليس "فن الشعر" ترجمة: شكرى محمد عياد (القاهرة، دار الكتاب العربى للطباعة والنشر، ١٩٦٧) ص ص ١٩ :٢٠
٢- موريس فيشمان "تدريب الممثل" ترجمة: نور الدين مصطفى (القاهرة، الدار المصرية للتأليف والنشر، ب ت) ص ١٢
٣- عطا درغام "فن التمثيل والممثل-٢" (الحوار المتمدن- ع ٤٥٠٤- ٢٠١٤)
٤- مارى إلياس، وحنان قصاب حسن "المعجم المسرحى- مفاهيم ومصطلحات وفنون العرض " (لبنان، مكتبة لبنان ناشرون، ١٩٩٧) ص ٥٧
٥- ينظر فى : عبد الوارث عسر "فن الإلقاء والخطابة" (القاهرة، الهيئة المصرية العامة للكتاب،١٩٩٢)
٦- ينظر فى نجاة على "فن الإلقاء- بين النظرية والتطبيق" (الدار المصرية اللبنانية، ط٤، ٢٠٠٨)
٧- ينظر فى: سامى عبد الحميد و بيدى حسون " فن الإلقاء" (بغداد، جامعة بغداد، ط٤، ١٩٨٢)
٨- فرحان بلبل" أصول الإلقاء والإلقاء المسرحى" (القاهرة، مكتبة مدبولى، ١٩٩٦) ص ٨٨
٩- ويليم شكسبير" هاملت" ترجمة: عبد القادر القط (لبنان، دار الأندلس، ١٩٨٢)
١٠- ألكسندر دين" أسس الإخراج المسرحى" ترجمة سعدية غنيم (القاهرة، دار المعارف المصرية، القاهرة ، ١٩٧٢) ص ٣٥٧
١١- قنسطنطين ستانيسلافسكى " إعداد الممثل" ترجمة: الدكتور محمد زكى العشماوى، ومحمود مرسى (لبنان، بيروت، دار النهضة العربية للطباعة والنشر، ب ت) ص ٢٣٣
١٢- -ينظر فى : ابراهيم أنيس "الأصوات اللغوية" (القاهرة، مكتبة الأنجلو المصرية، ط٥، ١٩٧٥)
١٣- ينظر فى: بشير كمال "علم اللغة العام- الأصوات" (القاهرة، دار المعارف، ١٩٨٠)
١٤- عمر مهديوى" الهندسة اللغوية والترجمة الآلية- المفهوم والوظيفة" (بحث مقدم للمنظمة العربية للترجمة، حول الترجمة والحاسوب، ٢٠١٤) ص ١٣
١٥- ينظر فى: د.أحمد على على لقم "تطبيقات هندسة اللغويات العربية-واقع وآفاق" (حولية كلية اللغة العربية- إيتاى البارود، ع٣١)
١٦- محمد السكران" الهندسة اللغوية وتنمية العربية" (جريدة الأهرام، ع ٤٧٥٢٦، ١٩ يناير ٢٠١٧)
١٧- لينوكس موريس" شركات تستعين بتكنولوجيا تحليل الصوت البشرى لاختيار موظفيها" فبراير، ٢٠١٧ https://www.bbc.com/arabic/
١٨- وكالة أنباء الشرق الأوسط " الذكاء الاصطناعى يشخص اضطراب ما بعد الصدمة عبر تحليل صوت المريض" ابريل ٢٠١٩
١٩- أحمد الجعفرى" الخبراء الفنيين وحل ألغاز الجرائم.." يونيو ٢٠١٨ https://www.youm7.com/story/2018/6/16
٢٠- - سعاد الهرمزى" التحليل العلمى لصوت العندليب" ابريل ٢٠١٧ . https://www.almadasupplements.com/news.php
٢١- David Crystal "A Dictionary of linguistics and phonatics" (UK, Blackwell publishing, 2008) p.540
٢٢- ينظر فى : عصام نور الدين "علم الأصوات- الفونوتيكا" (بيروت، دار الفكر اللبنانى، ط١، ١٩٩٢) ص ١٨
٢٣- د.كبير بن عيسى "دليل مستعمل تطبيق تحليل الإشارات الصوتية ومعالجتها- برات PRAAT" (الجزائر، مركز البحث العلمى والتتقنى لتطوير اللغة العربية، ع٩، ٢٠١٩) ص ص ٢٠: ٢١
٢٤- ينظر فى: خرما نايف "أضواء على الدراسات اللغوية المعاصرة" (الكويت، عالم المعرفة، ١٩٧٩)
٢٥- ينظر فى: بسام بركة "علم الأصوات العام- أصوات اللغة العربية" (لبنان، مركز الإنماء القومى، ب ت)
٢٦- ينظر فى: عبد الجليل عبد القادر"الأصوات اللغوية" (عمان، دار الصفاء، ط١، ١٩٩٨)
٢٧- د.كبير بن عيسى "دليل مستعمل تطبيق الإشارات الصوتية ومعالجتها" سبق ذكره ص ٥
٢٨- برنامج برات بآخر تحديث https://phonetics-acoustics.blogspot.com/2016/04/praat-2016.html
٢٩- هاملت أحد أشهر مسرحيات وليام شكسبير (١٥٦٤- ١٦١٦) إن لم تكن أشهرها على الإطلاق وتصنف كمأساة، أو ما يسمى بـ(مأساة الانتقام) وهي نوع من المسرحيات كان شائعاً في ذلك الوقت . وتدور حول هاملت أمير الدانمارك، الذي يموت أبوه على يد عمه، ويتزوج عمه من أمه ويظهر الشبح لهاملت ويخبره بحقيقة موته، وهنا يبدأ هاملت بالانتقام من عمه ولكن تردده الدائم لا يمكنه من ذلك، وتنتهى المسرحية بموت الجميع.
٣٠- ممثل مسرح وسينما إنجليزى، أول ممثل يمنح لقب سير، لعب عدد من الأدوار المختلفة من التراجيديا اليونانية ،وعصر النهضة، إلى الأنواع المختلفة من الدراما الانجليزية والأمريكية. مقاطع صوتية لمونولج To be or not to be Laurence Olivier's wonderful 1948 film 1948 www.youtube.com
٣١- ممثل بريطانى مواليد ١٩٦٨ ، مقاطع صوتية لمونولج To be or not to be Adrian Lester speaks Hamlet's soliloquy from act III, scene 1 2016. www.youtube.com

٣٢- ممثل ومخرج أمريكى مواليد ١٩٥٦ حصل على عدد من جوائز الأوسكار، مقاطع صوتية لمونولج To be or not to be – Mel gebson movie of Hamlet (1990) www.youtube.com

٣٣- ممثل اسكتلدى مواليد ١٩٧١، بدأ مسيرته الفنية عام ١٩٨٧، وحصل على جائزة إيمى. مقاطع صوتية لمونولج To be or not to be David Tennant as Hamlet in a film of the Royal Shakespeare Company's award-winning production of Shakespeare's greatest play. Directed by Gregory Doran2010 www.youtube.com

٣٤- ممثل أيرلندى مواليد ١٩٧٦، حصل على عدد من الجوائز ومنها جائزة سير لورانس أوليفييه، مقاطع صوتية لمونولج To be or not to be Andrew scott 2018www.youtube.com

٣٥- ممثل ومخرج بريطانى من أيرلندا الشمالية، مواليد ١٩٦٠، مثل وأخرج العديد من أعمال شكسبير رشح لعدد من جوائز الأوسكار وجولدن جلوب وغيرها، مقاطع صوتية لمونولج "To be or not to be" From Hamlet, by Kenneth Branagh 1996 www.youtube.com

Summary

Through the The research aimed to evaluate the basis of the creation under controlled by modern technology. scientific analysis of voice using modern programs, so we can evaluate that the actor controls his or her voice in various emotions as a basic step in reaching the creation.

The problem of the study is measuring the emotions that the actor expresses, is achieved by analyzing the sound without affecting the creation. Therefore, the study divided into three sections, the first section, which is entitled "The Art of Acting and its Relationship to voice" it deals with the didacticism and its uses in the training of the actor .

In the second section, "the voice and the language engineering ",it deals with the most important terms which used in the field of language engineering.

And the third section is the applied part . I choice an analog audio performance model from the play Hamlet composed by Willem Shakespeare, and the performance of a number of different actors chronologically and temporally. The sounds were analyzed using the PRAAT program. The research ends with the most important .results, discussing them, recommendations, and finally the list of resources and references.

صديقة لاشين

- **أستاذ التمثيل المساعد بقسم الدراسات المسرحية -جامعة الإسكندرية.**
- **مدرب دولى معتمد من مؤسسة هورايزن الدولية لريادة الأعمال ٢٠١٩.**
- **عضو لجنة تحكيم مسابقة مهرجان الفنون المسرحية فى الجامعة وخارجها.**
- **مدير مهرجان قسم المسرح المحلى والدولى في دورته الأولي والثانية والثالثة منذ ٢٠١١**
- **عدد من الأبحاث المحلية والدولية.**

<u>النشاط الفنى</u>

- **مدرب تمثيل بمركز الفنون- التابع لوزارة الشباب والرياضة بجامعة الإسكندرية منذ ٢٠١٨**
- **المشاركة بلجنة تحكيم نصوص مسابقة المدارس " مهرجان شكسبير" ٢٠١٦**
- **مدرب للتمثيل في مشروع القراءة الكبري بمكتبة الإسكندرية ٢٠١٥.**
- **عضو مشارك للتدريب بالورشة الدائمة لتدريب الممثلين في الورشة الدائمة بالهيئة العامة لقصور الثقافة.**
- **عضو لجنة الندوات بمهرجان نوادي المسرح بالإسكندرية .**

- **عضو لجنة مشاهدة العروض المسرحية في مهرجان التذوق المسرحي الثالث .**
- **عضو لجنة تحكيم اختيار عروض (٦ مين) بمركز جوته الألماني .**
- **ممثلة معتمدة باتحاد الإذاعة والتلفزيون منذ ٢٠٠٩.**
- **عمل ورش تدريبية للإعداد الجسدي للممثل، والارتجال، والأداء التمثيلى، منذ ٢٠٠٩ حتى تاريخه**
- **المشاركة بالأداء التمثيلى في عروض مسرح الجامعة وقصور الثقافة منذ عام ١٩٩٢، وحصلت على العديد من الجوائز.**

The Acoustic Characteristics of Read and Spontaneous Colloquial Arabic Speech Corpora: A Pilot Study

Rudyna Ahmed

Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt

rudynaahmed23@gmail.com

Abstract— **Speaking styles are built by the social environment, dialect, gender and educational level. The speaker characteristics have to be known as a full description of the speaker. The ultimate goal for all speech technology applications is to produce high quality applications. Currently available speech technology applications are not known by a great amount of flexibility, especially not when it comes to different speaking styles. The main aim of this paper is to establish a precise and systematic description of the acoustic characteristics of different speaking styles which are divided into reading speech and conversational speech corpora through spectrographic analysis to clarify the significant differences between them. This study examined the production of Arabic three emphasized vowels [iː], [uː], [aː] in the speech of six native Egyptian speakers, either having a conversation with someone or reading a script. The words containing the target vowels were elicited using two ways of sampling spontaneous interactions and prompted monologues. Fundamental frequency, speaking rate, vowel space area and vowel duration were measured and compared. The result further reveals that there are significant differences between the two types of speech corpora on the acoustic level. Reading speech yielded higher F2, greater vowel space expansion for vowel /i:/. Moreover, vowels in conversational speech were shorter in duration than those in reading speech. Speaking rate was significantly fast in conversational speech. There was no significant change in the fundamental frequency.**

Keywords: **Acoustic analysis, Speech corpora, Colloquial Arabic, Speaking styles, Reading and Conversational Speech**

1 INTRODUCTION

The most important factor in corpus design is the intended use for this corpus. The purpose is to determine whether a corpus will be written or spoken language, or both, and what the exact varieties that will be used in it. Biber et al. [1] clarify a corpus as "a large and principled collection of natural texts" and he divided corpus into two types: written and spoken (speech). The highlighted area in this paper is speech corpora which is compiled for specific purposes, especially developing consumer applications.

Corpus is the vital source for research towards specific applications such as text-to-speech, automatic speech recognition and automatic dialogue systems which clearly influence the specific kind of data that will be needed. Automatic speech recognition systems are most effective when the language content is highly constrained, the environment free from extraneous noise, and when the system has been trained on an individual speaker [2]. There are many situations, however, in which these conditions are hard to be applied. Most recognition and text-to-speech systems have to operate in the real world outside the sound-proofed laboratory, and the real world is often noisy. They must deal with real people, whose speaking styles display wide variation.

As the range of interests in corpus collection vary to a great extent, then the methods used for speaking style will also vary. Differences include whether the speech occurs naturally or is elicited for the purpose of data collection. Consequently Speech corpus is divided into two types, firstly: dialogue corpus, which is defined linguistically as spontaneous, unscripted conversation. In this conversation the speakers must sustain the conversation. Another type of dialogue may be characterized as prompted monologues because of the long duration such as: broadcast interviews, the participant roles are asymmetrical. Like dialogue the term of monologue covers a wide variety of speech types that include reading aloud, unscripted but prepared speeches and story-telling [2].

Currently the great majority of text-to-speech and automatic speech recognition applications are not characterized by its variability, especially when it comes to the specific voice or speaking style. On the opposite, the focus has been on reading unrelated sentences. There is, however, a very practical need for different speaking styles which is used in a variety of applications. The range of applications ask for a variation close to that found in human speakers. Apart from these practical needs, there is the scientific interest in formulating our understanding of human speech variability in explicit models. Therefore, it is necessary to know whether there is a significant difference between the various speaking styles or not in order to be scaled and taken into account. In this paper we address this point in terms of acoustic phonetics and whether there are differences to be noticed in the acoustic variables that may differentiate and make a notable distinction between the various speaking styles especially reading and conversational speech.
Acoustic analysis can be informative because it affords quantitative and qualitative analyses that carry potential for speech description and for determining the correlates of perceptual judgments of intelligibility, quality, rate and prosody [3]. Differences in languages and dialects point to the need for further research in this area. The underlying motivation for this study is to examine distinctive acoustic variables that distinguish reading speech from conversational speech (spontaneous) in the colloquial Arabic which is as mentioned above a less explored area expanding this area of research beyond work on other languages.

2 LITERATURE REVIEW

Previous researches on speaking styles has identified a wide range of acoustic features that characterize the conversational and clear speech. These include features that serve to improve the overall acoustic salience of the signal such that it is more resistant to the adverse effects of background noise or a listener-related perceptual deficit [4]. The most studied features that may affect speaking styles are: speaking rate, fundamental frequency, duration and vowel properties.

A. Speaking rate

The study [5] claimed that the speaking rates for conversational speech is 160 to 205 word/ minute and 90 to 100 word/minute for clear speech. He also reported that the slower speaking rate for the clear speech was related to increases in the occurrence and average duration of pauses, and to increases in the duration of many sound segments. A study by [6] mentioned that the overall sentence duration increases of 51% for males and 116% for females, respectively in clear speaking style. Another study by [7] carried on five subjects focusing on clear and conversational speaking styles revealed that the speaking rates of clear speech is 144 to 200 wpm (average, 174 wpm) but in the conversational speech is 140 to 204 wpm (average, 179 wpm).

B. Fundamental frequency

Most of previous studies reported that clear speech has higher F0 and a larger range in F0, and this is due to the larger amounts of laryngeal tension that the clear speech requires. A study by [6] carried on two subjects one male and one female revealed that the mean F0 was increased by 1.1 and 5.4 semitones, and F0 range was increased by 6.2 and 5.8 semitones. However, study by [8] reported that changes in F0 are not consistent across the speakers.

C. Duration

Clear speech produced without a constraint on speaking rate generally has increased durations of speech segments, though not by the same amount or by the same percentage for all speech sounds [9][5][6]. In a study carried by [6] the researcher claimed that the vowel lengthening in clear speech for both English and Spanish relative to conversational speech, although that the amount of lengthening is less in Spanish.

D. Vowel properties

The great majority of studies stated that in clear speech the formant frequencies of vowels generally span a larger space F1 and F2 than conversational speech vowels [5] [6] [10] [11] [12] [13]. Bradlow in his research [6] found a similar amounts of vowel expansion in clear speech in both English and Spanish vowels. Krause and Braida [14] figured out a less consistent result in their comparison of formants of clear speech and conversational speech, in that only the tense vowels of one talker showed a larger vowel space. Clear speech has increased rates of F2 transitions [12], narrower formant bandwidths [14] and longer durations of formant transitions [8] but these results are not always consistent across all vowel classes.

3 METHODOLOGY

The lack of available Arabic speech corpora, especially Egyptian Arabic, was a serious problem in this research, there was no available data to carry out the analysis. Accordingly a hand-tailored speech corpus was recorded to fit our analysis.

A. Subjects

To elicit the reading and the conversation speech, six Egyptian Arabic native speakers, encompassing no regional accents participated in the study. All speakers live in Alexandria, their age ranged from 21 to 24 years old. The subjects were matched pairs in gender as in (table 1).

TABLE I
SUBJECTS

Number of subjects	**Gender**	**Age**
1	Male	21.
2	Male	22
3	Male	24
4	Female	21
5	Female	22
6	Female	24

B. Corpus

Approximately 20 sentences were read by every subject to obtain the speech sample in which every sentence must contain a word from the three content words that contain three Egyptian Arabic long emphasized vowels ([ɑ:], [i:], [u:]) (/ʔɪstwaːna/'cylinder', /mʊsṭɑtiːl/ 'triangle', /mæxruːṭ/'cone'). Also, the conversational speech sample was obtained by focusing on repeating the previous three content words by managing a natural conversation with the subjects. All vowels in the three content words appeared in stressed syllables.

C. Corpus Elicitation

In order to obtain comparable samples of the target vowels and sentences across speech styles, firstly spontaneous speech sample was elicited by giving the subjects picture in which there was geometrical shapes (e.g. the blue triangle is above the red cone and to the left of the green cylinder) which one of the two participants describes so that the other can re-create the same pattern. We have initiated a conversation about these geometrical shapes with the subjects trying not to deviate from the context. [15], in this way of elicitation the speech is spontaneous (unscripted), but very limited both in vocabulary. Secondly: reading speech sample task was elicited by asking the subjects to read the written orthographic transcription targeting approximately 20 sentences. The speech sample was recorded by Praat, with a sampling rate 44,100 Hz.

D. Acoustic Analysis

A total of 120 words and 120 sentences were acoustically analyzed in both conditions reading and conversational speech. Acoustic measurements were carried out in Praat. The analyzed acoustic variables are as follows:

1) Fundamental Frequency: Fundamental frequency is global talker characteristic that varies across genders. [5] claimed that, clear speech is characterized by a somewhat wider range of F0 with a slight bias towards higher F0 than conversational speech. In [6] study of clear speech, average (mean) F0 was increased by 1.1 and 5.4 semitones, and F0 range was increased by 6.2 and 5.8 semitones, respectively, for the one male and one female talker.

All F0 analysis were carried out by Praat. For each vowel produced by each subject the mean, minimum and maximum F0 were calculated ten times in the reading and conversational speech for each speaker. Vowel F0 were obtained from vowel midpoint.

2) Speaking Rate: The overall speaking rate is also a global talker characteristic, although it is not a voice quality characteristic it is one of the most important features that distinguishes clear speech and conversational speech within individuals [16].

Speaking rate was measured for each of the subjects for ten sentences in both reading and conversational speech. Speaking rate is expressed as the mean number of phonemes per second.

3) *Vowel Space Area*: In order to calculate each speaker vowel space area, we selected ten occurrences of each of the three vowels ([ɑ:], [i:], [u:]), from the sentences of our corpus. All words were content words and none was the final word in the sentence. F1 and F2 were measured using Praat. F1 and F2 were obtained from vowel midpoint. Although it is known that steady-state values for formant transitions are not found simultaneously for F1 and F2 [17], if at all, various proposals have been made for measuring steady-state values for formants. They were converted to the psychoacoustic Bark scale [18]:

$Z= \{26.81/ (1+1960/f\}-0.53$, where Z is the critical band value of a formant in Bark and f is a formant's frequency in Hertz.

Vowel space area was measured according to the following formula [19]:

$\text{Distance } (F1,F2)= \{\sum_i (F1_i- F2_i)^2\}^{1/2}$

4) *Vowel Duration*: In order to calculate each speaker vowel duration, we also selected ten occurrences of each of the three vowels (/ɑ:/, /i:/, /u:/). Vowel duration was calculated from the waveform and spectrogram by Praat.

4 RESULTS AND DISCUSSION

A. Global results

Global results are shown in Table 2.

TABLE 2
GLOBAL RESULTS

	Reading Speech	**Conversational Speech**
F0 mean (HZ)	185.95	190.12
F0 range (Hz)	70.53	66.33
Speaking rate (phoneme/sec)	12.19	18.24667
Vowel duration	0.119539	0.09236
Vowel space area mean value of /a:/ (Barks)	3.15	3.83
Vowel space area mean value of /i:/ (Barks)	10.63	9.36

Vowel space area mean value of /u:/ (Barks)	4.38	4.25

1) *Fundamental Frequency:* For vowel F0, there was a no significant effect of speech styles. Vowels in conversational speech were higher in F0 by 4.17 Hz.

2) *Speaking Rate:* The statistical analysis revealed a significant effect of speech styles on the speaking rate. Speaking rate in conversational speech were remarkably faster than reading speech by 26.66%.

3) *Vowel Space Area:* The overall mean vowel space area of conversational speech (6.02 Barks) was larger than that of reading speech (5.96 Barks), but there was no significant effect. However, the expansion of the vowel space area in reading speech was driven by a change in the F2 rather than the F1 dimension, especially in the F2 of conversational speech which was higher by 1.27 Barks for all speakers.

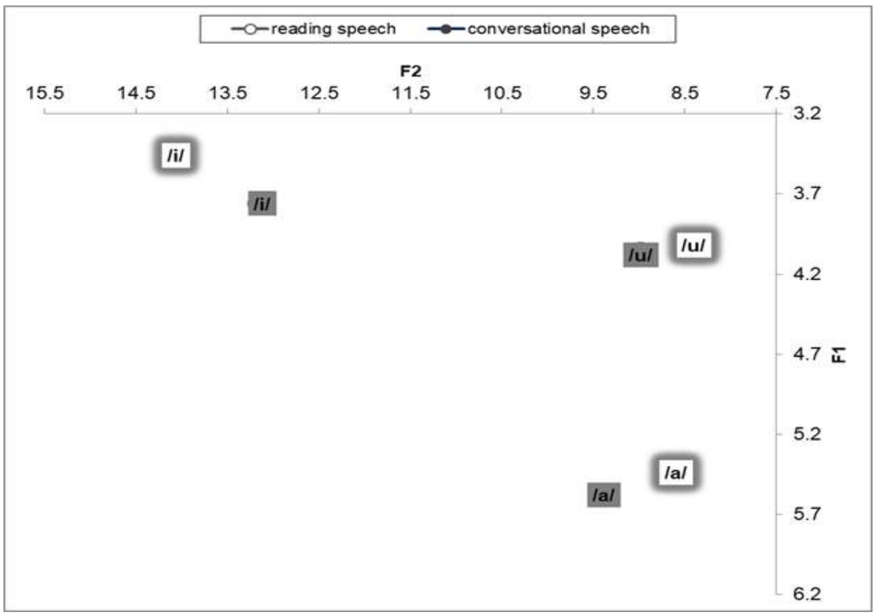


Figure 1: Vowel Space Area of Egyptian Arabic for Reading and Conversational Speech

4) *Vowel Duration:* The statistical analysis revealed a significant difference between reading speech and conversational speech with regard to vowel duration. Vowels in conversational speech were found to be shorter by an average 0.026 ms.

B. Individual results

Male1, male 3, female1 and female2 (Figure 1) correspond to a great extent to the global results but their average fundamental frequency in conversational speech is significantly high that may distinguish conversational speech from reading speech. On the contrary, the fundamental frequency of male 2 and female 3 is higher in reading speech. The other acoustic variables almost correspond to the global results.

TABLE3
INDIVIDUAL RESULTS OF CONVERSATIONAL SPEECH

Gender	**Vowel**	**F0 Mean (HZ)**	**F0 Range (HZ)**	**F1 (HZ)**	**F2 (HZ)**	**Vowel Duration**	**Speech Rate (phoneme/sec.)**

Male 1	/a:/	161.688	51	540.577	1113.391	0.0478	17.08
	/i:/			369.208	2046.312	0.0897	
	/u:/			375.647	1065.857	0.0817	
Male 2	/a:/	118.639	38	679.216	1328.074	0.0580	18.14
	/i:/			349.360	2028.851	0.1065	
	/u:/			407.043	1081.565	0.0792	
Male 3	/a:/	102.842	35	519.202	954.9057	0.0899	17.12
	/i:/			290.047	2039.266	0.0882	
	/u:/			403.348	972.9897	0.1207	
Female 1	/a:/	257.323	91	679.216	1328.074	0.0968	18.78
	/i:/			349.360	2028.851	0.1337	
	/u:/			407.043	1081.565	0.0809	
Female 2	/a:/	247.684	99	496.882	1149.673	0.1018	19.02
	/i:/			374.647	1978.328	0.1078	
	/u:/			379.977	1132.457	0.1027	
Female 3	/a:/	252.565	84	560.457	1030.569	0.0356	19.34
	/i:/			513.276	2088.497	0.0544	
	/u:/			475.339	1136.528	0.4784	

TABLE4

INDIVIDUAL RESULTS OF READING SPEECH

Gender	Vowel	F0 Mean (HZ)	F0 Range (HZ)	F1 (HZ)	F2 (HZ)	Vowel Duration	Speech Rate (phoneme/sec)
Male 1	/a:/	149.192	53	564.620	1005.286	0.1112	11.82
	/i:/			305.862	2220.534	0.1414	
	/u:/			359.143	1010.481	0.1468	
Male 2	/a:/	135.3293	40	638.207	1058.991	0.0742	15.12
	/i:/			327.755	2492.63	0.1112	
	/u:/			381.963	957.6422	0.0829	
Male 3	/a:/	113.947	44	513.543	917.4883	0.1060	10.83
	/i:/			273.001	2157.643	0.1273	
	/u:/			358.630	868.7661	0.1274	
Female 1	/a:/	220.623	117	638.207	1058.991	0.1103	10.6
	/i:/			327.755	2492.635	0.1288	
	/u:/			381.963	957.6422	0.1442	
Female 2	/a:/	224.384	96	482.033	1038.341	0.1096	10.72
	/i:/			353.546	2276.129	0.1537	
	/u:/			399.751	1034.272	0.1354	
Female 3	/a:/	272.226	73	536.695	987.8427	0.0964	14.05
	/i:/			472.736	2421.009	0.1284	
	/u:/			523.158	1050.382	0.1165	

5 CONCLUSIONS

The acoustic variables that were measured clarified that there is a difference between speaking styles and showed that there is need to work in this area in the future. We look forward to test our results by manipulating speech synthesis and evaluating the perception of the results. Many more studies, involving other style changes, relations and dialects other than Colloquial Arabic also need to be carried out in order to aid our understanding of the limits of individual variability.

REFERENCES

[1] Biber, Douglas/Conrad, Susan/Reppen, Randi (1998), Corpus Linguistics. Investigating Language Structure and Use. Cambridge: Cambridge University Press.

[2] Lüdeling, A. (Ed.) & Kytö, M. (Ed.) (2008). *Volume 1*. Berlin, Boston: De Gruyter Mouton.

[3] Kent, R. D. and Kim, Y. (2008). Acoustic analysis of speech. In M. J. Ball, M. R. Perkins, N. Müller, and S. Howard (eds), The Handbook of Clinical Linguistics. Oxford

[4] Bradlow, Ann & Bent, Tessa. (2002). The Clear Speech Effect for Non-Native Listeners. The Journal of the Acoustical Society of America. 112. 272-84. 10.1121/1.1487837.

[5] Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. Journal of Speech & Hearing Research, 29, 434–46.

[6] Bradlow, A. R. (2003). Confluent talker and listener-related forces in clear speech production. In C. Gussenhoven & N. Warner (eds.), Laboratory Phonology, 7 (pp. 241–73). Berlin & New York: Mouton de Gruyter.

[7] Krause, J. C. & Braida, L. D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. Journal of the Acoustical Society of America, 112, 2165–72.

[8] Krause, J. C. (2001). Properties of naturally produced clear speech at normal rates and implications for intelligibility enhancement. Unpublished doctoral thesis, MIT, Cambridge, MA.

[9] Ferguson, S. H. & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal hearing and hearing-impaired listeners. Journal of the Acoustical Society of America, 112, 259–71.

[10] Chen, F. R. (1980). Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level. Unpublished master’s thesis, MIT, Cambridge, MA.

[11] Ferguson, S. H. & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal hearing and hearing-impaired listeners. Journal of the Acoustical Society of America, 112, 259–71.

[12] Moon, S.-J. & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. Journal of the Acoustical Society of America, 96, 40–55.

[13] Kuhl, P. K. et al. (1997). Cross-language analysis of phonetic units in language addressed to infants. Science, 277, 684–6.

[14] Krause, J. C. & Braida, L. D. (2003). Acoustic properties of naturally produced clear speech at normal speaking rates. Journal of the Acoustical Society of America, 115, 362–78.

[15] Swerts, M./Collier, R. (1992), On the Controlled Elicitation of Spontaneous Speech. In: Speech Communication 11, 463 468.

[16] Ann R. Bradlow, Gina M. Torretta, David B. Pisoni. (1996) Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. Speech Communication 20:3-4, 255-272.

[17] Di Benedetto, M.-G. (1989). Vowel representation: Some observations on temporal and spectral properties of the first formant frequency. The Journal of the Acoustical Society of America, 86, 55–66.

[18] Traunmüller, H. 1990. Analytical expressions for the tonotopic sensory scale. J. Acoustic Society of America, 88, 97–100

[19] Bradlow, Ann & Torretta, Gina & Pisoni, David. (1996). Intelligibility of Normal Speech I: Global and Fine-Grained Acoustic-Phonetic Talker Characteristics. Speech Comm. 20. 255-272. 10.1016/S0167-6393(96)00063-5.

ACKNOWLEDGMENT

This work was supported by Dr. Sameh Alansary. I thank him for his insightful comments.

BIOGRAPHY

Rudyna: *Master Degree student at Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt*

She graduated from Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, (2019), her main areas of interest include Acoustic Analysis, Prosodic Analysis and Speech Corpora

الخصائص الاكوستيكية للمدونات اللغوية للكلام المقروء و المحادثة في العامية المصرية

Rudyna Ahmed

Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt

rudynaahmed23@gmail.com

ملخص **— اساليب الكلام تبنى من خلال البيئة الاجتماعية ، واللهجة ، والجنس ، والمستوى التعليمي وأيضًا حسب موقف ما معين يمر به المتحدث. الهدف النهائي لجميع تطبيقات تكنولوجيا الكلام هو إنتاج تطبيقات عالية الجودة. لا تُعرف تطبيقات تكنولوجيا الكلام المتاحة حاليًا بقدر كبير من المرونة ، لا سيما عندما يتعلق الأمر بأساليب التحدث المختلفة. الهدف الرئيسي من هذا البحث هو وضع وصف دقيق ومنهجي للخصائص الصوتية المميزة لأنماط التحدث المختلفة و التي تنقسم الى كلام مقروء و محادثة و ذلك من خلال التحليل الاكوستي الطيفي. تناولت هذه الدراسة نطق ثلاثة صوائت مفخمة و هم كالاتي: [i:]، [u:]،[a:] و يتم تسجيلهم لستة متحدثين للعربية المصرية ، و ذلك عن طريق اقامة محادثة و قراءة نص. تم قياس ومقارنة التردد الأساسي ، ومعدل الكلام ، و مساحة الفراغ الاكوستي و مدة لكل صائت. كشفت النتيجة عن وجود فروق بين الكلام المقروء و المحادثة على المستوى الاكوستي. أسفرت النتائج عن ارتفاع الشكل الثاني للصائت /i:/ ، و طول المدة الزمنية للكلام المقروء عن طول المدة الزمنية للمحادثة، و ان معدل الكلام ايضاً كان مرتفع عن المحادثة اكثر من الكلام المقروء. لم يتم ملاحظة اي تاثير ملحوظ من التردد الاساسي ليميز بين المحادثة والكلام المقروء.**

الكلمات الدالة: **التحليل الاكوستيكي، المدونات اللغوية، العامية المصرية، اساليب الكلام ، الكلام المقروء و المحادثة.**

Syllables Classification of ASR using Hybrid Visual Features in Fixed HMM

Doaa A. Lehabik [*1], Amr M. Gody [*2], Mohamed H. Merzban[*3], Sameh F. Saad [**4]

[*]*Electrical Engineering Department, Fayoum University, Fayoum, EGYPT*

[1]da1174@ fayoum.edu.eg

[2]amg00@fayoum.edu.eg

[3]mhm00@fayoum.edu.eg

[**]*Modern Sciences and Arts University, 6 October City, Giza, Egypt*

[4]dr.sam.far@gmail.com

Abstract: This paper presents a different approach to classifying speech phonemes. Two-hybrid techniques are used to emphasize the principle idea of this paper. The first hybrid model is constructed of fixed state, structured Hidden Markov Model, Gaussian Mixture, Mel scaled Best Tree, Convolution Neural network, Vector Quantization (FS-HMM-GM-MBT-CNN-VQ). The second hybrid model is constructed of fixed state, structured Hidden Markov Model, Gaussian Mixture, Mel scaled Best Tree, Convolution Neural. TIMIT database is used in this paper. All phones are classified into five classes and segregated into Vowels, Stops, Fricatives, Nasals, and Silences. The results show that using (FS-HMM-GM-MBT-CNN-VQ) is an available method for classification of phonemes, with the potential for use in applications such as automatic speech recognition and automatic language identification.

Key words: ***Automatic Speech Recognition, Classification technique, HTK, wavelet packet, Convolution Neural Network, Vector Quantization, Hidden Markov Model***

1 INTRODUCTION

Automatic speech recognition (ASR) is recognized as the independent computer-driven transcription which changes talked language into legible text. ASR lets a computer to get the words from a person that speaks into a microphone and alter them into written text. The phonetic features and articulation of various sounds are necessary for the classification into separate categories. This classification of sounds can be applied for applications like speaking rate evaluation, speech recognition, phone recognition, and language recognition. The broad phone classes are usually known as vowels, nasals, fricatives, stops, and silence. This categorization can improve speech recognition and hence categorization techniques were attempted. The current research presents two classification methods. The hybrid features model of MBT, CNN and or without VQ with fixed states of the Hidden Markov Model is used. Various Gaussian mixture numbers are used to get a greater rate of recognition. Now, Section 2 explains related work but, section 3 discusses each block in the proposed model. The experiment environment includes a database and experiment procedure in section 4. The results discussion would introduce in section 5 and conclusions would be shown in section 6.

.

2 RELATED WORK

Using classification techniques in ASR is the best preprocessing tasks to increase the recognition speech rate. Classification means that you had some categories and observed units and you want to assign these observed units to these categories. Classification is performed by matching feature vectors from various categories with the observed units.

In [1], expert classifiers are used for each broad phonetic class which classifies speech signals into vowels, stops, fricatives, and nasals and used the TIMIT database. The highest obtained phone accuracy was 74.2%. In [2], two feature extraction techniques to classify speech signals into five classes in the TIMIT database. These features are MFCC and time-frequency reassigned cepstral coefficients (TFRCC). For stops, the highest success rate (SR) is 53.74% by TFRCC but other classes achieved high SR when using MFCC. In [3], the authors classify the speech phonemes based on histograms of the reconstructed phase spaces into three classes Which are fricative, vowel, and nasal in the TIMIT database. The results achieved overall recognition rates of 61.59%, 34.49% and 30.21% for fricative, vowel, and nasal phonemes respectively. Using the Malayalam speech signal in [4], that used a two-stage system to spot the boundaries of vowels, nasals, and approximants. In the first stage, a speech signal is classified into six broad phoneme classes using an ANN but for second stages, frequency domain parameter named spectral peak frequency is suggested for accurate verification of nasals. Sonorant and non-syllabic features are used for verifying approximants and the syllabic feature is used for locating vowels. In [5], the authors classified speech signals into 19 classes which were 8 vowels and 11 consonants. This classification was made using two techniques that were Procrustes analysis and support vector machine (SVM). Procrustes analysis performed 91.67%, 91.37% accuracy for vowel and consonants respectively but SVM performed 89.05%, 88.94% accuracy for vowel and consonants respectively. In [6], support vector machine (SVM) is used to match phonemes into 6 classes in Gujarati language and the accuracy is 95.70 %. Producing in [7] new classification for broad phoneme by features that obtained immediately from a speech at the level of this signal. Broad phoneme classes comprise vowels, nasals, fricatives, stops, approximants, and silence. This classification is applied to three systems, each system is applied to three tests and results are 54%, 61%, and 46% for the combination on TEST 1, TEST 2 and TEST 3 respectively. Based on the pattern matching in [8], there is a method for accurate spotting of plosives which tested using TIMIT corpus. Results showed that the accuracy of spotting the plosives using the presented approach was very high. The distribution of insertions for the different classes is 34% for silences, 28% for fricatives, 23% for vowels, 7% for glides, 6.5% for nasals and 1.5% for affricates.

3 PROPOSED MODEL

In this model as indicated in Fig. 1, the input speech signal was resampled into 10 kHz to best distribute the wavelet tree structure through the significant band. Then framing it into small frames (20 ms). Mel Best Tree (MBT) images that are obtained from speech signal enter on Convolution Neural Network (CNN) then to Vector Quantization (VQ) to extract the features. This feature enters on HMM with various GMM to analyze speech signals into five classes; Vowels (V), Stops (S), Fricatives (F), Nasals (N) and Silences (Si). The state's number of HMM is fixed for each class. Each block of this model would be discussed in more detail

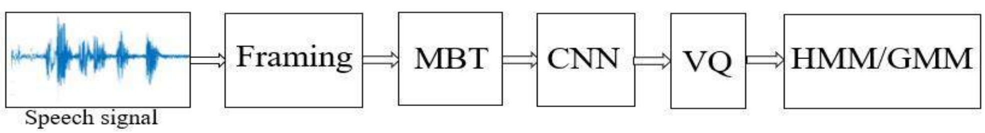


Figure 1: Block diagram of the proposed model

in the latter subsection.

C. Feature Extraction

1) Mel Best Tree

The idea of Best Tree Encoding should be introduced. The algorithm of creating the BTE feature file [9] starts by converting the speech signal into a collection of short frames. Then, the entropy of the Wavelet Packet Decomposition (WPD) coefficients is applied as a projection of these frames of the speech signal power into defined filter banks. The best tree that contains the significant signal power, using the acquired entropy, is obtained. The dynamic range of the BTE is normalized using the encoding algorithm in [10]. The new direction is considered in the version of BTE (BTE with Mel-filter). The algorithm of estimating the best tree is targeted in this version of BTE. It extracts the Mel frequency wavelet packet Best Tree Encoding

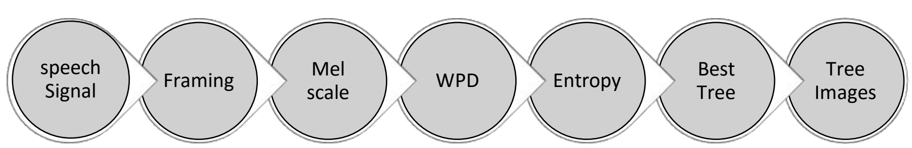


Figure 2: Block diagram of creating MBT

(MFBC) features from the WAV file. It considers applying Mel-scale before the conversion [11]. It also resamples the input to 10 kHz. In this research, we used Mel best tree images not encoding as in Fig. 2.

2) *Convolution Neural Network*

It is a branch of deep learning that used for image recognition and rarely in speech recognition [12] and we choose deep residual network (Resnet50) that act as a type of CNN with 50 layers. This network can organize images into 1000 object classifications. Subsequently, the network can learn rich features for a broad range of images. It has a size for an input image of [224 224 3]. The output features vector is included in 1000 features.

3) Vector Quantization

Vector quantization (VQ) is an efficient source-coding model. Vector quantization is a method that encodes the input vector into an integer number with an entry of a group of reproduction vectors. The reproduction vector is defined to be closest to the input vector. The coding efficiency is originated in the process of converting the vector into a compact integer representation. The performance of the vector quantizer, however, depends on whether the set of reproduction vectors, which are usually called code words, is accurately chosen such that the distortion is minimized.

D. Hidden Markov Model with Gaussian Mixture Model

Hidden Markov Model (HMM) is the ablest method used in automatic speech recognition. This system is produced for the Markov process with private parameters and we want to distinguish the hidden parameters from the observation [13]. The states are hidden, and the probability distribution for each is known as the variable which affects the states. Temporal data and states are usually identified as separate GMMs [14] in the HMM model. The transition matrix learns from training data and it is a known transition of state to another [15].

4 EXPERIMENT ENVIRONMENT

The later subsections explain the type of database that is used in this paper. the process of converting raw speech into features used in this classification and verification. MATLAB 2018b and visual studio 2015 are used as lab environment. The specifications of the laptop that used in the experiment are 8.00 GB RAM, 64-bit operating system, Intel(R) Core (TM) i5-8250U CPU @1.60 GHz 1.80 GHz and NVIDIA GeForce MX150 with 8061 MB Memory @4 GHz.

E. Database

The continuous corpus of TIMIT [16] is an acoustic-sounding speech made of English, recorded by a microphone at 16 kHz and 16-bit resolution. This database holds 6300 sentences (5.4 hours) in 630 speakers from 8 regional dialects of the United States (US). Each speaker articulated 10 sentences and all the sentences were identified with its phone level. The main version of TIMIT includes 61 phonics. The database is prepared to modify transcription files for the character recognition objective of this research. Vowels (V), Stops (S), Friction (F), Nasal (N) and Silence (Si) [17]. The following table presents the phone assigned to each classification.

TABLE 2

PHONES CLASSIFIERS

Classifiers	**TIMIT Labels**
Vowels (V)	aa, ae, ah, ao, ax, ax-h, axr, ay, aw, eh, el, er, ey, ih, ix, iy, l, ow, oy, r, uh, uw, ux, w, y
Stops (S)	p, t, k, b, d, g, jh, ch
Fricatives (F)	s, sh, z, zh, f, th, v, dh, hh, hv
Nasals (N)	m, em, n, nx, ng, eng, en
Silences (Si)	h#, epi, pau, bcl, dcl, gcl, pcl, tcl, kcl, q, dx

F. The Procedure of the proposed model

In this model, First, speech signal enters on MBT block. In this block; read speech signal and resample signal into 10 kHz. These samples are framed into 20 ms. The Mel-scale curve is implemented to convert all frequencies to Mel frequencies. Then apply wavelet packet decomposition and extract the best tree images. Second, Mel Best Tree images are entering as input on CNN block. In that extracting features from MBT images by using Resnet50. In this case; the features output vector is containing 1000 components. Third, these features enter as input to vector quantization block which features output vector is one component. Fourth, these features enter on HMM/GMM block. In this block; all classifiers are trained to utilize a fixed number of HMM states. In this model, we used four emitting states and two non-emitting states as in Fig.3 (The non-emitting states are required to identify the entry and the exit state in HMM model). GMM is used to act as spatial distribution probability density functions of attribute vectors of N-Dimensions. The mixture count is variable in this research. The system is tested with various Gaussian mixture counts (1, 2, 3, 4, 5, 6 and 7). HTK tool is utilized to build HMM-based speech processing tools. There is another model that is not used by VQ. In this, we used feature vector as 1000 components that were extracted from CNN block, then

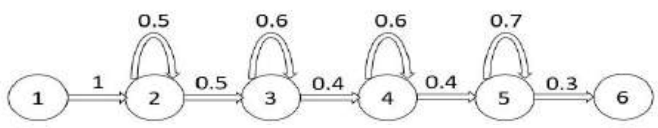


Figure 3: FS-HMM-VQ for all classifiers model

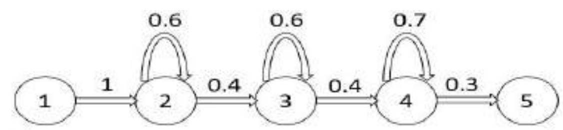


Figure 4: FS-HMM for all classifiers model

these features enter on HMM/GMM block which used five a fixed state (three emitting states and tow non-emitting states) as in Fig. 4.

5 RESULTS AND DISCUSSIONS

There are two models of features vectors that are applied: First, (MBT-CNN-VQ) is used and this vector consists of one component. Second, (MBT-CNN) is used and this vector is 1000 components as indicated in session 4. The results of these two proposed models (MBT-CNN-VQ) and (MBT-CNN) are illustrated in this section. Success Rate (SR) of each syllable is utilized for evaluating the proposed model. The comparison study is implemented to show the details and the key power in each specific feature set. We calculate the success rate (SR) in each case as in Table 2. The success rate can be defined by equation 1 and the result is shown as in Fig. 5 that shows the value of each SR against the Gaussian mixture (GM). In this equation: (D denotes as deletions), (s denotes as substitution) and (n denotes as the number of phones in the expected transcription). The following tables show the result in two cases with various numbers of GMM (1,2,3,4,5,6 and 7). Case1: Using (MBT, CNN, and VQ) features with fixed states number of HMM as in Fig. 6. Case 2: Using (MBT and CNN) features with fixed states number of HMM as in Fig. 7.

$$SR = \frac{N - D - S}{N} \qquad (1)$$

TABLE 3

EXPERIMENT RESULTS OF SUCCESS RATE

Mixture Count	Case 1 SR%	Case 2 SR%
1	43.42%	42.44%
2	47.36%	43.13%
3	50.21%	46.78%
4	51.79%	46.95%
5	54.97%	45.59%
6	**57.49%**	43.08%
7	57.07%	**47.76%**

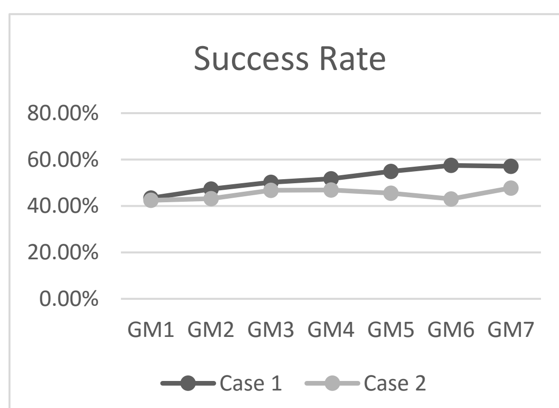


Figure 5: Success rate results

Table 2 shows that the best successive rate in two modules that achieved 57.49% when applying (MBT-CNN-VQ) features extraction with fixed state Hidden Markov model and Gaussian mixture number 6. Table 3 shows that the greatest success rate for each classifier. In case 1: vowels achieved 81.1% using a gaussian number 6, stops achieved 93% using gaussian number 1, fricatives achieved 70.5% using gaussian number 7, nasals achieved 94.7% using a gaussian number 4 and silences achieved 47.6% using gaussian number 7 as in Fig. 6. In case 2: vowels achieved 67.5% using a gaussian number 3, stops achieved 93.3% using gaussian number 1, fricatives achieved 87% using gaussian number 4, nasals achieved 72% using a gaussian number 2 and silences achieved 92.4% using gaussian number 6 as in Fig. 7.

TABLE 4

SUCCESS RATE FOR EACH CLASSIFIER IN EACH CASE

	SR for each classifier									
	Fixed states HMM and using vector quantization					Fixed states HMM and without using vector quantization				
GMM	**V**	**S**	**F**	**N**	**Si**	**V**	**S**	**F**	**N**	**Si**
1	53	**93**	0	91.8	29	65.6	**93.3**	28.4	45.1	63.2
2	63.9	87.1	3.5	92.1	30.3	64.6	89	52.8	**72**	53.1
3	71.9	76	15.9	93.9	31.6	**67.5**	92.1	60.1	61.7	63.9
4	73	58.5	48.4	**94.7**	35.2	59.6	76.2	**87**	66.4	72.9
5	78.3	57.5	58.7	90.8	41.4	49	75.1	72.4	61.1	89.9
6	**81.1**	67.7	66.9	81.7	45.5	39.8	75	64.7	65	**92.4**
7	78.2	77.9	**70.5**	70.9	**47.6**	65.1	75.1	71.4	51.3	90.4

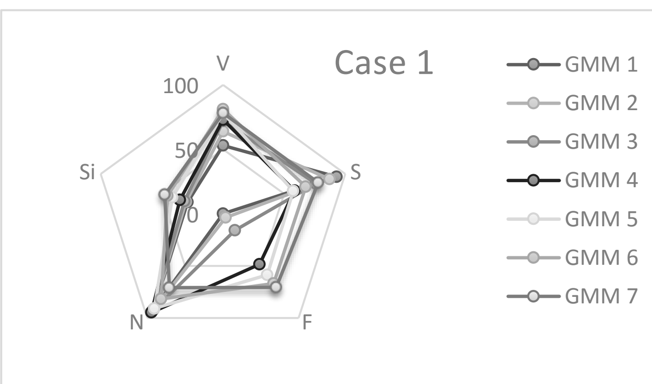


Figure 6: SR of classes in Fixed states HMM and using VQ

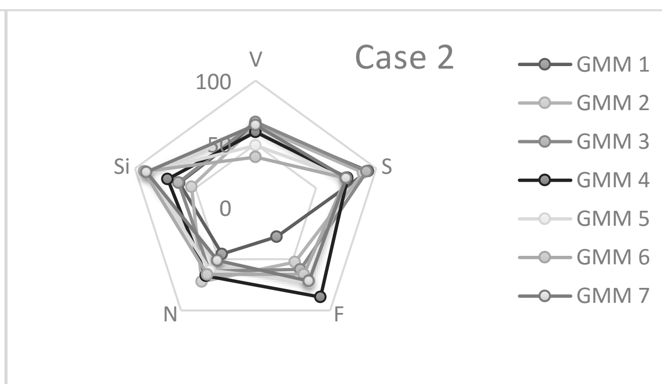


Figure 7: SR of classes in Fixed states HMM and without using VQ

The confusion matrix (error matrix) results of all experiments were showed in table 4 exhibit which is a table layout applied for a classification model, a classifier, or a recognizer as in our case. The performance summary is depended on test data, where its actual values are well identified. This table relates to the probability (success rate of the class) of the best decision. Vowels are the usual occurring sound group in the speech data. In GMM 1, 2 and 3 vowels are not recognized as silence. Vowels are recognized more in the first model than second. Stops and fricatives have small appearances in the TIMIT database. Stops are very far from silence in small gaussian, but they very near in nasals in the first model. Fricatives are very bad in the first model with small gaussian because they were recognized as stops and nasals more. Fricatives have increased in successive rate with the higher Gaussian mixture. Nasals are ideally recognized as nasals in the first model and good in the second model. For silence when using vector quantization, a result of successive rate is small because this class contain [17] (h#, epi, pau, bcl, dcl, gcl, pcl, tcl, kcl, q, dx) that (h#) had high duration in time but (bcl) had short time. So, vector quantization is not suitable for this class. But in the second model, we used all features vector (1000 vector size) not compressed vector (1 vector size). So, the result of this class is perfect with the high gaussian mixture.

TABLE 5

CONFUSION MATRIX

GMM		**Confusion Matrix**									
		Fixed states HMM and using vector quantization					**Fixed states HMM and without using vector quantization**				
		V	S	F	N	Si	V	S	F	N	Si
1	V	**13340**	6340		5492	0	**13466**	5642	684	430	302
	S	125	**6759**		358	0	63	**5832**	306	39	13
	F	353	2943		2539	0	109	2297	**1087**	201	130
	N	84	307		**4394**	0	44	1121	188	**1174**	78
	Si	590	3717		3929	**3365**	243	2477	308	265	**5662**
2	V	**16386**	3940	27	5310	0	**12933**	4079	1494	1403	123
	S	221	**6091**	4	680	0	70	**5438**	433	163	5
	F	540	2071	**193**	2686	0	134	1464	**2485**	563	56
	N	139	236	0	**4348**	0	41	526	330	**2368**	23
	Si	922	2701	23	4085	**3360**	298	2043	670	918	**4440**
3	V	**18540**	1896	158	5194	1	**13784**	4179	1103	945	417
	S	348	**5037**	33	1213	0	113	**5547**	189	140	36
	F	732	1165	**878**	2751	0	141	1321	**2958**	354	147
	N	158	123	6	**4391**	0	88	674	266	**1827**	105
	Si	1266	1544	145	4324	**3363**	340	1846	530	619	**5894**
4	V	**18665**	1020	683	5173	19	**11030**	1879	3322	1452	837
	S	498	**3604**	270	1786	0	140	**3821**	719	274	63
	F	569	485	**2981**	2117	5	91	332	**5648**	284	139
	N	130	62	54	**4393**	0	71	289	585	**2133**	134
	Si	1100	779	585	4122	**3576**	229	559	1384	608	**7486**
5	V	**19965**	947	929	3577	92	**8326**	1867	2326	1256	3214
	S	650	**3395**	380	1475	1	101	**3611**	471	232	394
	F	648	436	**3704**	1491	29	85	434	**4168**	288	782
	N	222	93	92	**4082**	5	50	270	428	**1936**	484
	Si	1278	652	844	3058	**4112**	100	265	675	223	**11203**
6	V	**20574**	1298	1223	2137	131	**6297**	1935	2059	1578	3939
	S	642	**4073**	506	786	5	64	**3627**	417	196	532
	F	682	548	**4251**	815	57	44	490	**3546**	370	1031
	N	343	217	190	**3411**	15	26	221	381	**2113**	511
	Si	1522	1019	1121	1817	**4565**	48	206	519	218	**12051**
7	V	**19441**	2169	1609	1474	167	**11969**	1649	1952	680	2125
	S	495	**4953**	459	441	7	157	**3267**	444	161	321
	F	590	773	**4602**	506	57	149	465	**3633**	211	627
	N	373	456	267	**2752**	31	86	310	413	**1294**	418
	Si	1304	1547	1289	1214	**4860**	179	287	500	149	**10471**

6 CONCLUSIONS

It has been indicated that the automatic speech recognition success rate is improved using a hybrid method of acoustic-phonetic approach and pattern recognition approach. The acoustic-phonetic approach is a statistical procedure, which depends on the HMM. The methodology of mixing (MBT-CNN-VQ) gives a greater success rate of correctness than the second model (MBT-CNN). The vector quantization technique also acts as a good role in achieves real results. The result was improved by using a various number of gaussian mixtures. To be specified in terms of particular class classification performance, the highest success rates are produced, using (FS-HMM-GM-MBT-CNN-VQ) as of approximately 81.1% for vowels at (GM 6) and as of 94.7% for nasals at (GM 4). When using (FS-HMM-GM-MBT-CNN), the highest success rates are achieved as of 93.3% for stops at (GM 1), as of 87% for fricatives at (GM 4) and as of 92.4% for silence at (GM 6). By comparing with the result in [1] that classifies speech signals into vowels, stops, fricatives, and nasals with 60.5%, 83.3%, 81.4%, and 75.9% success rate respectively. The successive rate (SR) of vowels and nasals are more than in [1] when used case 1 and more than 1 for stops and fricatives when used case 2. The results are developed by renewing the key parameters of the hybrid model. The key parameters are the features, HMM model and the numbers of the gaussian mixture. HMM, models can be renewed for most often occurring models by tying the states of seldom appearing phones. The silent class can be also altered for better HMM designs. A future work, concerning the modification of the HMM states from fixed state to variable states for each classifier and using RNN instead of CNN to obtain a higher recognition rate.

REFERENCES

[1] P. Scanlon, D. P. Ellis, and R. B. Reilly, "Using broad phonetic group experts for improved speech recognition," *IEEE transactions on audio, speech, and language processing,* vol. 15, no. 3, pp. 803-812, 2007.

[2] G. Tryfou, M. Pellin, and M. Omologo, "Time-frequency reassigned cepstral coefficients for phone-level speech segmentation," in *2014 22nd European Signal Processing Conference (EUSIPCO)*, 2014: IEEE, pp. 2060-2064.

[3] J. Ye, R. J. Povinelli, and M. T. Johnson, "Phoneme classification using naive bayes classifier in reconstructed phase space," in *Proceedings of 2002 IEEE 10th Digital Signal Processing Workshop, 2002 and the 2nd Signal Processing Education Workshop.*, 2002: IEEE, pp. 37-40.

[4] S. Salim, G. Deekshitha, A. George, and L. Mary, "Automatic Spotting of Vowels, Nasals and Approximants from Speech Signals," in *2018 International CET Conference on Control, Communication, and Computing (IC4)*, 2018: IEEE, pp. 272-277.

[5] J. Wang, J. R. Green, A. Samal, and Y. Yunusova, "Articulatory distinctiveness of vowels and consonants: A data-driven approach," *Journal of Speech, Language, and Hearing Research,* 2013.

[6] A. Chittora and H. A. Patil, "Classification of phonemes using modulation spectrogram based features for Gujarati language," in *2014 International Conference on Asian Language Processing (IALP)*, 2014: IEEE, pp. 46-49.

[7] G. Deekshitha and L. Mary, "Broad phoneme classification using signal based features," *International Journal on Soft Computing,* vol. 5, no. 3, p. 1, 2014.

[8] J. Keshet, D. Chazan, and B.-Z. Bobrovsky, "Plosive spotting with margin classifiers," in *Seventh European Conference on Speech Communication and Technology*, 2001.

[9] A. M. Gody, "Wavelet Packets Best Tree 4 Points Encoded (BTE) Features," in *The Eighth Conference on Language Engineering, Ain-Shams University, Cairo, Egypt*, 2008, pp. 189-198.

[10] A. M. Gody, R. A. AbulSeoud, and M. M. Ibraheem, "Hybrid Model Design for Baseline-Context-Independent-Mono-Phone Automatic Speech Recognition," *International Journal of Engineering Trends and Technology (IJETT)–Volume,* vol. 27.

[11] A. Gody, R. Abul Seoud, and M. Ezz El-Din, "Using Mel-Mapped Best Tree Encoding for Baseline-Context-Independent-Mono-Phone Automatic Speech Recognition," *The Egyptian Journal of Language Engineering,* vol. 2, no. 1, pp. 10-24, 2015.

[12] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458,* 2015.

[13] P. Bansal, A. Kant, S. Kumar, A. Sharda, and S. Gupta, "Improved hybrid model of HMM/GMM for speech recognition," 2008.

[14] G. Xuan, W. Zhang, and P. Chai, "EM algorithms of Gaussian mixture model and hidden Markov model," in *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*, 2001, vol. 1: IEEE, pp. 145-148.

[15] J. C. Brown and P. Smaragdis, "Hidden Markov and Gaussian mixture models for automatic call classification," *The Journal of the Acoustical Society of America,* vol. 125, no. 6, pp. EL221-EL224, 2009.

[16] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, "DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM. NIST speech disc 1-1.1," *NASA STI/Recon technical report n,* vol. 93, 1993.

[17] C. Lopes and F. Perdigao, "Phone recognition on the TIMIT database," *Speech Technologies/Book,* vol. 1, pp. 285-302, 2011.

BIOGRAPHY

Amr M. Gody obtained the B.Sc. M.Sc., and PhD. from the Faculty of Engineering, Cairo University, Egypt, in 1991, 1995 and 1999 sequentially. He met the teaching staff of the Electrical Engineering Department, Faculty of Engineering, Fayoum University, Egypt in 1994. He is the Acting director of Electrical Engineering department, Fayoum University in 2010, 2012, 2013, 2014 and 2016. His contemporary research areas of interest involve speech processing, speech recognition and speech compression. He is author and co-author of several papers in national and global conference proceedings and journals such as Springer(International Journal of Speech Technology), the Egyptian Society of Language Engineering (ESOLE) journal and conferences, International Journal of Engineering Trends and Technology (IJETT), Institute of Electrical and Electronics Engineers (IEEE), International Conference of Signal Processing And Technology (ICSPAT), National Radio Science Conference(NRSC), International Conference on Computer Engineering &System (ICCES) & Conference of Language Engineering(CLE).

Doaa A. Lehabik obtained the B.Sc. degree in Electrical Engineering – Communications and Electronics Department with very good with honor degree, from the Faculty of Engineering - Fayoum University in 2014. She met the M.Sc. program in Fayoum University - Communications and Electronics Department in 2015. She got the Pre-Master degree from Fayoum University with very good with honor degree, in 2016. Her areas of interest involve Speech recognition and Image recognition.

Mohamed H. Merzban obtained the B.Sc. from Fayoum university, M.Sc. from Cairo university, and PhD. from Egypt-Japan university for science and technology (E-JUST). He joined the teaching staff of the Electrical Engineering Department, Faculty of Engineering, Fayoum University, Egypt in 2006. He is now a lecturer at Fayoum university. His current research areas include computer vision and its applications to Robotics.

Sameh F. Saad obtained the B.Sc. degree in mechatronics engineering from High Institute of Engineering, Giza, Egypt, in 2000, the M.S. degree in mechatronics engineering from Mechatronics Laboratory, Paderborn University, NRW, Germany, in 2004 and the Ph.D. degree in electrical engineering at Cairo University, Giza, Egypt in 2012. From 2016 to 2018, he was a Researcher and a Lecturer with the October University for Modern Sciences and Arts, Giza, Egypt. His research interest includes the development of mobile robots and autonomous vehicle, automation of aquaponics ecological system, and automation of mechatronics systems.

التعرف التلقائي على الكلام بأستخدام تصنيف المقاطع الصوتية عن طريق نموذج ماركوف ذو الهيكلة الثابتة

دعاء أحمد لحيبك *[١], عمرو محمد جودي*[٢] , محمد حمدي مرزبان*[٣], سامح فريد**[٤]

* قسم الهندسة الكهربية, جامعة الفيوم , مصر

[1]da1174@ fayoum.edu.eg

[2]amg00@fayoum.edu.eg

[3]mhm00@fayoum.edu.eg

** جامعة العلوم والاداب , 6 اكتوبر , الجيزة , مصر
4dr.sam.far@gmail.com

ملخص
يقدم هذا البحث طريقتين مختلفتين لعملية التصنيف المستخدمة فى عملية التعرف التلقائى على الكلام. اول طريقة تتكون من هذه الخصائص المميزة للصوت (تطبيق نظام الميل سكيل على افضل شجرة من ال WPD مع شبكة التداخل العصبيه مع VQ) وثاني طريقة تتكون هذه الخصائص (تطبيق نظام الميل على افضل شجرة من ال WPD مع شبكة التداخل العصبيه) . تم استخدام بنيات مختلفة لنموذج ماركوف الخفى ذو الهيكلة الثابتة. تم تصنيف المقاطع الصوتية الى ٥ مقاطع وهي حروف متحركة (V) و حروف لا تحتوي على كلام (S) و حروف احتكاكية (F) و حروف انفية (N) . وصامت التي لا تحتوي على اي كلام (Si) . تم استخدام قاعدة البيانات TIMIT في هذا البحث . وتم استخدام عدد مختلف من GM الذي يتكون من(١ او ٢ او ٣ او ٤ او ٥ او ٦ او ٧). وحيث أن نجاح عملية التعرف التلقائى على الكلام يعتمد بشكل كبير على إجراء عملية التصنيف بطريقة صحيحة. لذا فإن هذا البحث يركز على الوصول لعملية التصنيف الصحيحة لرفع الكفأة لنظام التعرف التلقائي على الكلام عن طريق تحسين مصفوفة الانتقال TM وتغيير عدد GM. اول طريقة هي التي تعطينا افضل نتائج 57.49% . ونستخدم ال HTK وذلك لشهرتها الواسعة في مجال ال ASR.

الكلمات الدالة
التعرف التلقائي على الكلام, تقنية التصنيف, تحليل حزمة الموجيات (WPD), شبكة التداخل العصبية, المتجهات الكمي (VQ) , نموذج ماركوف الخفي

The Perception of Arabic Vowel Length by Native and Non-native Listeners An Experimental Investigation

Eman Kassem [*1], Lamyaa Tawfik [**2]

[*] *Phonetics and linguistics Department, Faculty of Arts, Alexandria University*
Alshatby, Alexandria, Egypt
[1] Eman.kassem@alexu.edu.eg

[**] *Phonetics and linguistics Department, Faculty of Arts, Alexandria University*
Alshatby, Alexandria, Egypt
[2] lamyaatawfik@hotmail.com

Abstract— **The length of a vowel - in many languages- is considered a distinguishing phonemic aspect; a listener could perceive an utterance in two different manners depending on the length and the quality of a specific vowel. [1] highlights the difference between duration and length stating that vowel duration is the actual timing an utterance takes, whereas length refers to the time of perceiving a vowel.**

Since a vowel length difference in Arabic leads to changes in meaning, the relative length of a vowel turns to be a distinguishing aspect in the Arabic language. However, to what extent there could be a perceptual and acoustic difference between long and short vowels? And what are the acoustic features upon which the native listeners and nonnative Arabic learners [Indonesian listeners) may depend to differentiate between short and long vowels?

This research aims at studying the perception of both native Arabic listeners and foreign listeners of the vowel duration in the Arabic language based on an experimental approach; it also clarifies their dependence on their cognitive relevance of the vowel length, through their responses on the processed vowel length.

Keywords: **vowel length perception, shortening, lengthening, merging, native and non-native listeners, psychoacoustic study.**

1 INTRODUCTION

With respect to all psycho-acoustic relations, it is worth noting that acoustic variation does not necessarily lead to an equivalent perceptive change [1]. Moreover, since the quality of a vowel conveys a great significance in communicating information, there could be – in general - certain differences in the quality of a long vowel and its short counterpart. [2] Stated that There could be both quantitative and qualitative differences between certain long and short pairs of vowels, whereas for other pairs, only quantitative ones occur .

Arabic includes six basic vowels (Fat<u>h</u>a/Kasra/<u>d</u>amma) /æ/ /i/ /u/ and the long vowels (*alif, yaa* and *waw*) / æ:/ /i:/ /u:/. Classifying Arabic vowels into *fatha, Kasra* and *damma*, does not imply that Arabic has no other vowels. The two vowels /ɛ/ and /o/ have been added along with the other basic vowels as they occur in a great number of Arabic dialects [3]. Such two vowels and the aforementioned emphatic vowel / ɑ /have been added to the present study as they do exist in the Egyptian dialects; they also play a main functional role in identifying meaning.

In this research , it was noticed that meaning change could also occur as a result of length change as well; If the vowel / ɛ / in /bɛ:t/ is shortened, it gives another word with different meaning, such as /bet/ meaning "a girl" as pronounced in Egyptian colloquial Arabic [4]. Thus, the vowels /ɑ/ , /o/ and /ɛ/ are dealt with as independent ones in the present study.

Giving due concern to vowels from a phonological perspective, certain Arabic long vowels do have short counterparts, such as /æ/, / ɑ / and /ɛ/, whereas others do not such as /i , u/. Vowel quality changes might be occurred with some vowels when they turned into short one; for example the word /di:b/ which means " fox" in Arabic , when it is pronounced with short vowel , it turns to vowel / ɛ/ /d ɛb/, also vowel /u/ in the word /ku:b/ . For the latter, a change in the quality of the vowel may occur as it is clear in /i/ and /u/.

Ancient linguists were mainly concerned with the quantity of vowels rather than the quality. The contemporary linguists hold different views concerning the relation between the quantitative and qualitative aspects of a vowel in what concerns the limits differentiating long and short vowels, and whether such limits are restricted to duration or other acoustic features related to the quality of a vowel (Intensity, fundamental frequency and the formants).

The contemporary linguists are divided into two main groups: *The first group* deems that the differences between a long vowel and its short counterpart, from a quantitative perspective, dictate a change in quality.

According to a study made by Omar,A.M [5] indicates that both a long and a short vowel have independent phonemes functionally causing a change in meaning and having minute differences in acoustic features concerning the length or the shortness of the same vowel. This is also manifested by a number of studies, such as that of (6], [4], (7]- [8].

Such conclusions are manifest through a previous experimental study by Al-Ani, S. [9] in which he represented the difference between the long and short vowels through their measured formants as follows:

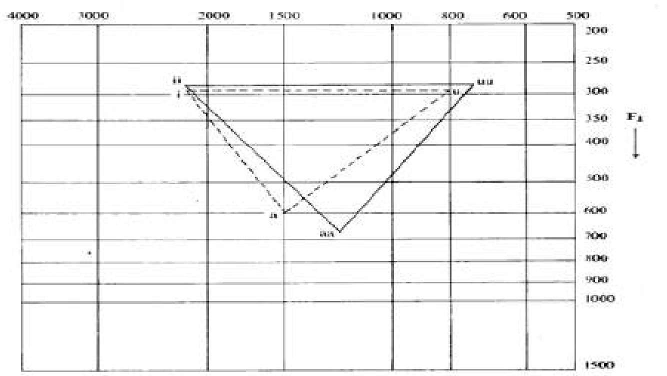


DIAGRAM 1

Figure [1] indicates the difference between the short vowels and their long counterparts through measuring F1 and F2 [9].

The second group deems that the differences between a long vowel and its short counterpart are a quantitative one; they state that no qualitative difference occurs, and if any, they are considered to be mere secondary. A study made by [10] also stresses the quantitative difference; he remarks that since fatha, kasra, damma are parts of their long counterparts *alif*, *jaa* and *waw*, as mentioned by the old grammarian [11.], they thus possess the same quality features. As for the quality, there is no difference whatsoever between the two groups.

This is also manifested by a number of studies, such as that of . [12],[13]- [14].

In Arabic, the relative length of short and long vowels have been measured by some linguists; [9] found that short vowels duration are in the range between 100-150 msec., while long ones are about 225 – 350 msec. [4] are about 95 msec. for short and about 112 msec. for long vowels in monosyllabic words. [15] found that the short vowels are about [117 msec.) while the long vowels about (282 msec.) in monosyllabic words.

In other languages, some of the experiments showed the importance of vowel duration as a prominent cue to differentiate the vowel qualities, whereas it has a little weight compared with the other spectral information.

In Holland language, [16] has conducted an experiment (by changing the spectral information of the vowel, while its durational information hold the same.) its results indicated that native language tend to depend on spectral information, while learners of English as second language depend on durational information.

Tsukada, K. [17 & 18] proved that listeners depend on spectral characteristics beside vowel duration in their perception of vowel length by examining the perception of vowel length pair of words in three languages; Arabic, Japanese and Thai. In Japanese language, [19] found that duration is the main cue in the perception of vowel length, while F0 is a secondary feature. In French language, [20] found that native French language does not depend on the difference in vowel durations as a perceptual cue, while the French learners of English native language do.

In English,. [21] found that the change in vowel durations has a little weight as a perceptual cue in identifying the quality of the vowel. This was an opposite result of. [22]' study who found that duration is an important perceptual cue.

Some studies like [23] found that vowel type (backness and height) may affect its duration. They found that low back vowels are affected by change in duration, while high front vowels were difficult to be affected.

.[24] came to the conclusion that Indonesian Language is syllable-timed language, they usually do not have lexical stress nor do they have vowel length contrast -as opposite to Arabic language- while the latter is stressed-timed language having a vowel length contrast.

2 EXPERIMENTAL STUDY

The present study thoroughly examines the approach of the experiment adapted. The approach is tackled at two main parts; the first one is manifested into three processes done on recorded words: (a) lengthening the short vowel, (b) shortening the long vowel, and finally (c) merging a part of the long vowel into the short one to make the later long. Acoustic measurements of the basic features (duration, F0, F1, F2, F3 and jitter/shimmer features) of the vowels are performed before and after the above mentioned three processes. The second part is a subjective–perceptual test for the results of the three processes.

A.. Procedures:

-1 Selecting the Linguistic Samples:

The material selected for the experiment is made up of six pairs of monosyllabic words /CVC/ /CVVC/ where the /V/ and the /VV/ represent both short and long basic Arabic vowels / æ / /i/ /u/ and the /ᵅ/ /ε/ /o/ as used in colloquial Egyptian; the difference in the length among such vowels do have a great role in differentiating meaning.

Vowel	Long V: CVVC		Short V: CVC	
	Word	Meaning	Word	Meaning
/æ/	\dææb\	Melted	\dæ bb\	To step feet
/ε/	\bεεt\	House	\bεtt\	A girl
/i/	\diib\	Bear	\dib\	No meaning
/o/	\toob\	A dress	\tob\	The Arabic infinitive for "to ask forgiveness"
/u/	\kuub\	A glass	\Kub\	No meaning
/ᵅ/	\tᵅᵅb\	To be well done	\tᵅbb\	To fall down (in colloquial Egyptian)

Table II- I

- Closed consonants /b t d k/ are tested at initial and final positions of syllables.
- Choosing word pairs with long high vowels /u/ and /i/ are reported to have no short counterparts in Arabic except when the quality is changed. Accordingly, subjects were trained to pronounce the short counterpart of the vowels in the same quality form even when the resulting words make no sense, e.g., /dib/ and /kub/.
- The words were pronounced by five male subjects between 22- 35 years old with no speech deficiency.
- The researcher explained the experiment to the subjects and trained them on how to pronounce the words before the actual recording; they were asked to pronounce each pair several times, and only correct pronunciations in context are selected.

-2 Recording the Linguistic Data:

The words were recorded in the department of phonetics and linguistics, University of Alexandria using a Computerized Speech Lab (CSL), 4500 kay elemetrics. The material of the experiment is hundred and eighty samples for the aforementioned words, including 5 subjects X 3 times of repetition for each word X 12 words (6 short vowels and 6 long ones).

-3 Acoustic Analysis:

- Programs used for Analysis:

Acoustic analysis of the selected recordings was done using the following speech acoustic analysis programs:

➢ Computerized speech lab (CSL), 4500 kay elemetric [being the main program of recording and analysis]
➢ Praat version 6005- win 32 for making the three processes, shortening, lengthening and merging by using copy , cut, and past options.
➢ E section, speech filling systems [SFS] for the LPC analysis.
➢ Winsnoori speech analyzer, version 1.34. is adapted in this respect to test the clarity of the merging process for the listener.

- Analysis Steps and Measurements:

Some measurements for the vowels were done before applying the aforementioned three processes: (duration, I, F0, F1, F2, F3 and jitter/shimmer).

<u>Basic Vowels (short/ long/ duration measurements):</u>

- Measuring the average duration of the basic short vowel in /dæb/ [60 – 100 msec.] of all the subjects with their repetitions (for all the vowels).
- Measuring the average duration of the basic long vowel in /dæ:b/ [160 – 250 msec.] of all the subjects with their repetitions (for all the vowels). (see Figures: 1-2 A/ 1-2 B)
- **<u>P.S.:</u>** The average (100 msec.) of short vowels and (250 msec.) of long vowels were chosen to help applications on shortening and lengthening processes.

Spectrogram analysis of long vowel in the word /dæ:b/ and the short one in /dæb/

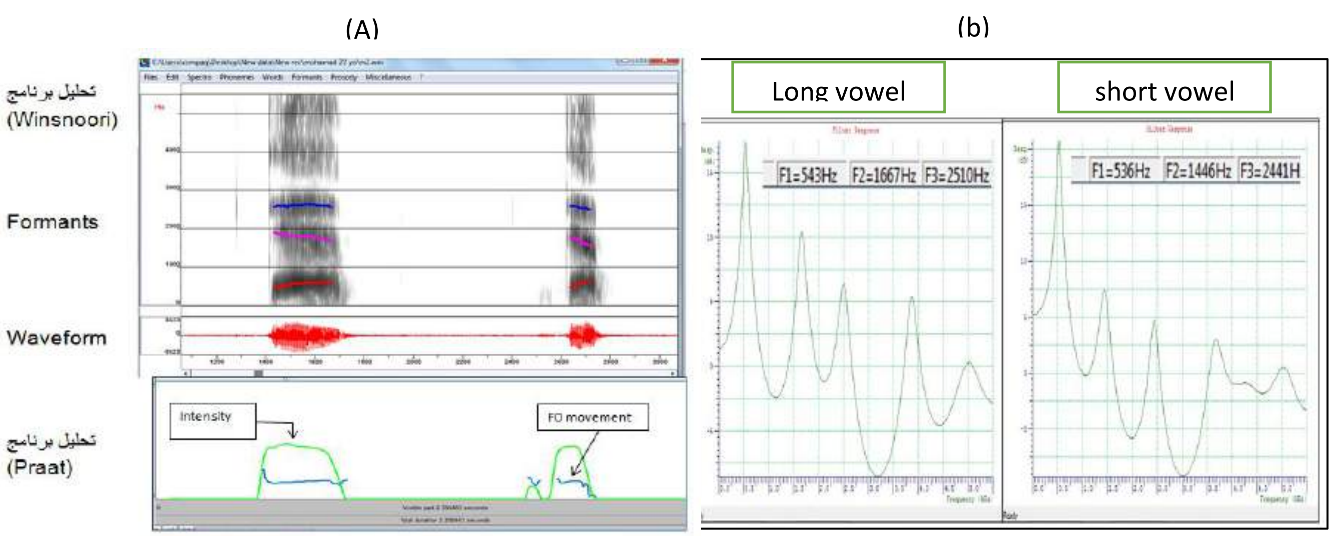


Figure 2-1 A - spectrogram and the waveform of the long and the short basic forms of the vowel /æ/. According to the steps followed, the duration of the short vowel and its fundamental frequency and [F1/F2/F3] in /dæb/ as well as that of /dæ:b/ were measured. Differences in the intensity of the two vowels and the FO were measured by Praat. Figure 2-1-B: LPC spectrum of the long vowel in /dæ:b/(right) and the short one in /dæb/ (left as measured by EsectionSFS)

<u>The first process: *Shortening a Long Vowel* Reaching the Length of a Short One:</u>

This is done by deleting a part of the steady state of the long vowel to equal the length of the short one (see Figure 2-2):

Figure: 2-2 indicates the steps of identifying, cutting a part of the steady state of the long vowel (right), and the result of the shortening process (left).

Comparing both the original short vowel and the result of the shortening process. See the following Figures 2-3 A (spectrogram, waveform, and F0/Intensity). See also Figure 2-3 B which shows the LPC spectrum.

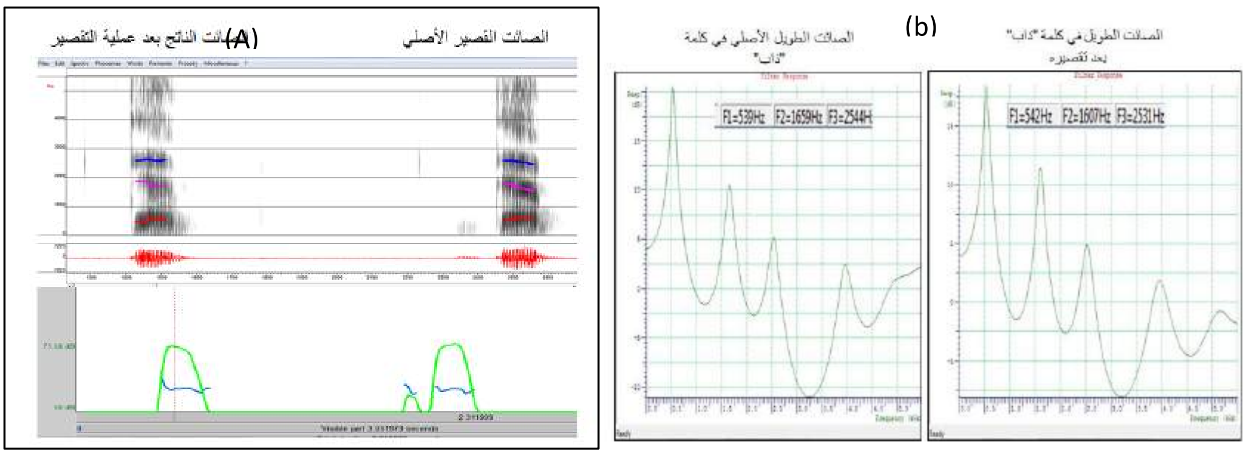


Figure: 2-3 A represents a comparison between the original short vowel /a/ (right), and the shortened vowel (left) in spectrogram, waveform and the (F0/intensity). Figure 2-3 B represents the LPC analysis of the original short vowel /a/ (left), and the shortened one (right).

In cases of pasting, adding or copying a part of the waveform, identifying the steady state of a vowel was done first; this part starts and ends with a full wave; cutting, pasting, adding and copying were not, hence, randomly proceeded in order to avoid sound clicks (see Figure 2-4).

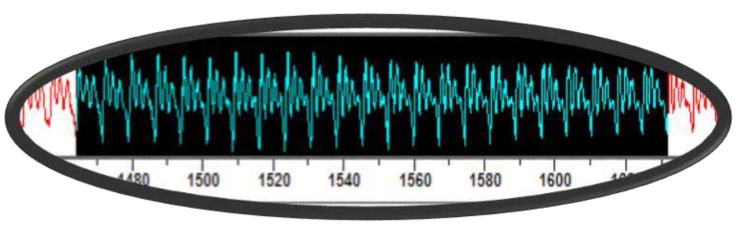


figure 2.4 select wave that will be deleted. this part starts and ends with a full wave; cutting, pasting, adding and copying were not, hence, randomly proceeded in order to avoid sound clicks.

<u>The second process: Lengthening the Short Vowel in its Same Wave:</u>

The short vowel /æ/ in /dæb/ was lengthened by pasting a section of steady state of its band ***several times*** to equal the length of the original long vowel. This procedure follows the coming steps:

- Identifying the section of the short vowel needed to be copied.
- Copying the selected section and pasting it in the same place several times (see Figures: 2-5/2-6 A, B)

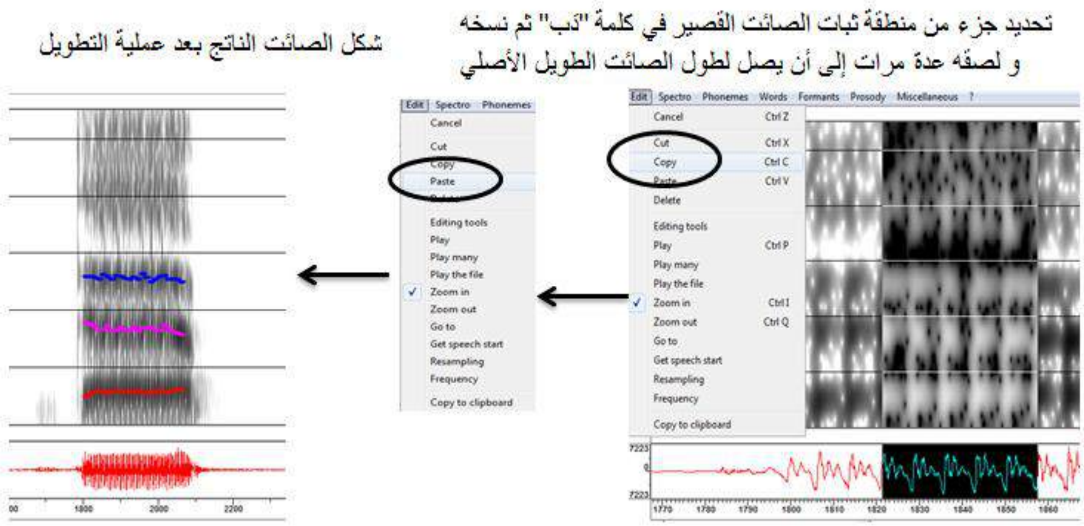
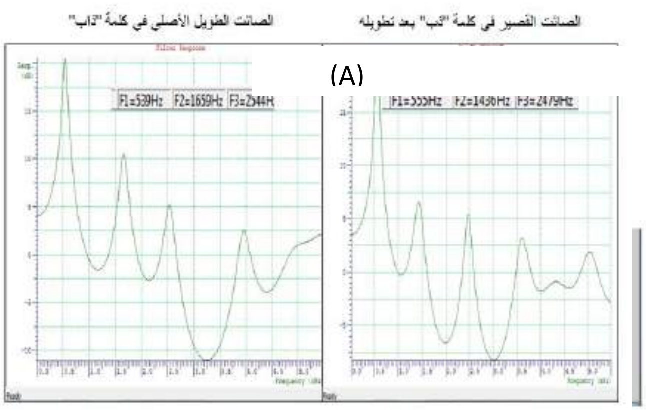


Figure: 2-5 indicates the process of identifying a part of the steady state of the short vowel /a/ for copying and pasting it into the same place of the short vowel several times. The left spectrogram represents the result.

(b)

(A)

(B)



1.
2.
3.
4.
5.
6.
7.
8.
9.

Figure: 2-6 A represents the spectrogram of the vowel /a/ after the process of lengthening (right), and the original short vowel (left) attached with their intensity and F0 movement. Figure: 2-6 B LPC spectrum of both vowels; after lengthening (right) and the original short one (left)

<u>The third process: Merging Part of the Wave of the Long Vowel to the Wave of the Short One to Make the Later long:</u>

The procedure was done as: Copying a part of the steady state of the long vowel and Paste it to the wave of the short one to make it long.

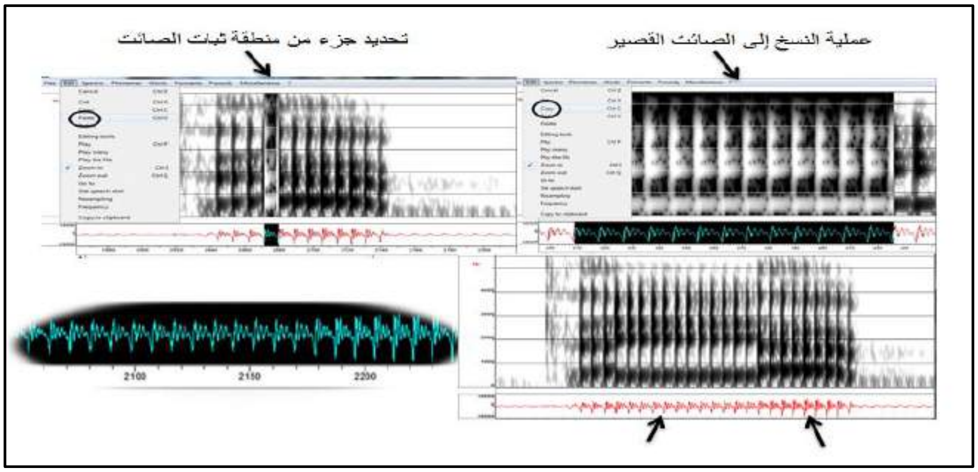


Figure: 2-7 represents the following steps of the merging processes: copying a part of the steady state of the long vowel (left) and paste it into the short vowel (right). The result of lengthening the short vowel is seen in the fourth spectrogram (down/ right), the difference between the two waveforms is noticed by the two arrows.

The following is the comparison between the original long vowel and the result of merging process.

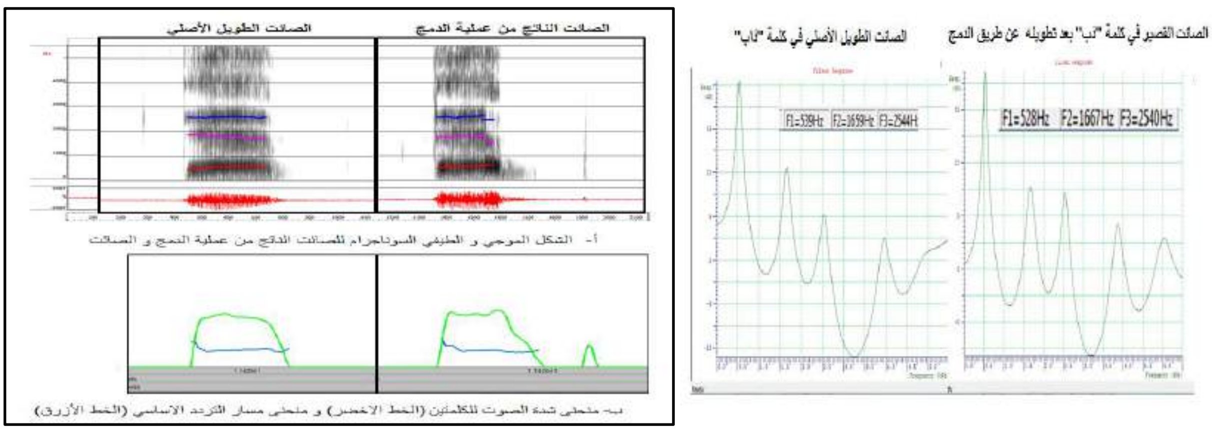


Figure: 2-8 A represents the comparison between: - The sonogram of the original long vowel (left) and the result of the merging process (right). – The intensity and F0 curves (the lower diagrams) for the two words /dæ:b/ /dæb/ and Figure 2-8 B represents LPC spectrum of both vowels; after lengthening by merging (right) and the original long one (left).

B. The second part of the experiment is the Perceptual Test:

-1 Subjects

- **Native Arabic listeners:**

Fifty participants between (20-35 years old), 25 females and 25 males, from undergraduate and postgraduate students at the department of phonetics were selected. The listening experiment was then done on the spoken material resulted from the previous acoustic procedure.

-2. Presenting Recorded Material to Native Listeners:

- The listening experiment took place in the studio of the department of phonetics, faculty of Arts, University of Alexandria. Each person was tested individually.
- Each pair of words was presented to the listeners, without pictures, using PowerPoint. The meaning of the words is given under each pair .
- The researcher explained the recorded material for the listeners, indicating that the difference lies in the length of the vowel, such as /dæ:b/ and /dæb/.
- The testing sheet of the listener includes six pairs of words. The participants were asked to choose and tick the word corresponding to what they had heard. Table (2-2)
- In applying all the previous procedures, the material was presented to the listeners in random order.

- **Nonnative Arabic listeners**

Data was presented to 40 nonnative Arabic listeners who were Indonesians (20 male and 20 female). They were 17 – 18 years old, their native language was Bohasa. They were Arabic learners at “The secondary extension institute of Arabic and Islamic studies in Egypt". They were at different levels (from level 1 to level 6). It is noteworthy that their native language does not have vowel length contrast. As the recorded material presented to native listeners (See: 2.2.2), aforementioned recorded material steps & instructions were presented to Indonesian listeners

Vowels		**بيت**	**بِت**
/ɛ/		**/bɛ:t/**	**/bɛt/**
		√	
O	**1**		√
S	**2**		√
L	**3**	√	
M	**4**	√	√

Table II-II: The testing sheet includes 4 cases of recording: Original record (O), Shortening (S), Lengthening (L) Merging (M).The listener was guided to consider numbers (1, 2, 3, 4) rather than symbols (O, S, L, M). The participants were asked to choose and tick the word corresponding to what they had heard.

3 RESULTS

This research is considered as a psychoacoustics study since it includes two main directions: perceptual and acoustical aspects. The first is the human ability to distinguish and classify what he/she hears in an abstract manner. the second aspect is the acoustical measurements and changes of the acoustic features in the production which is the concrete evidence for listener`s perception.

The research is based on an experimental approach; it clarifies the cognitive relevance of the vowel length of the listeners; natives and Arabic learners (whose mother tongue have no difference in vowel length), through their responses on processed vowel durations.

A. Acoustical results

-1 Durational features:

The results of this study showed that, the relative length of the short vowel was 70 - 90 msec. and the long vowel was 170-220 msec. with the average length of short to long was approximately 1: 2.5, while the results of several studies in Arabic showed that the length of the long vowel was twice as long as the short vowel [25], [26], [9], [13], [27]-[17]

The measurements of the relative vowel length differ from a study to another.

1. [9] study showed that the length of long vowel was is 300-350 m / s and the short vowel was 130-150 msec.
2. [4] The results of showed that the relative length of the short and long 95 m long was 112 msec.
3. [15] while the relative length of vowels was 117 msec. for short and 282 msec. for long.

The relative length of the different studies varies according to the spoken people and the vocal context of the spoken pronunciation as well as the different dialect and gender differences.

-2 Spectral information (vowel quality):

The study [13] showed that some long vowels have short counterpart that differ in quantity and in quality, the latter is a secondary feature

[28] study showed that vowels duration may affect vowel quality in some vowels, for example high closed vowels /i, u/ do not change their quality according to vowel length. While some other vowels such as /ʌ, ɛ, æ, ɑ / are affected by the length in their quality. This result was confirmed by the [29]' study applied to the Hungarian language. The study of [30] also examined the extent to which duration affects the listeners' perception of the length and quality of the vowel in Swedish language, through the process of shortening some long vowels. The results showed that vowel duration is a distinct element in the Swedish language, and also there are some vowels are affected in their quality.

[9], studied duration and components of short and long vowels. His results showed that the change in the quality of the long and short (æ / æ æ) was more than the change in the U / UU and (I / II).

The present study enhancing these results, native listeners observed a quality change in some vowels in the shortening process i.e. as they described that the long vowel / æ / in the word [dæ:b] was changed to the vowel [ˤ] when it was shortened so the resulted word was [dˤb], while high vowels / i and u/ qualities kept on.

-3 fundamental frequency f0 for vowels:

Some studies have demonstrated the idea of the intrinsic pitch of vowels, [31] confirmed that the values of f0 vary somewhat by changing the quality of the pattern. High vowels [u, i] have higher f0 values than the rest of vowels.

[18] study, in the Japanese language, showed that there is a relationship between the fundamental frequency and the temporal absorption in the perception of vowel length, since the low fundamental frequency is associated with the long vowels.

Results of this study showed that Short vowels are higher in f0 values than long, especially high vowels [u, i]. even in shortening and lengthening processed vowels but the merged processed vowels were the lowest in f0 values.

Acoustic measurements of Jitter and Shimmer (which measure the amount of irregularity of cycle to cycle F0 and Intensity) indicated significant values of irregularity
of merging waveforms. (Figure: 3.1)

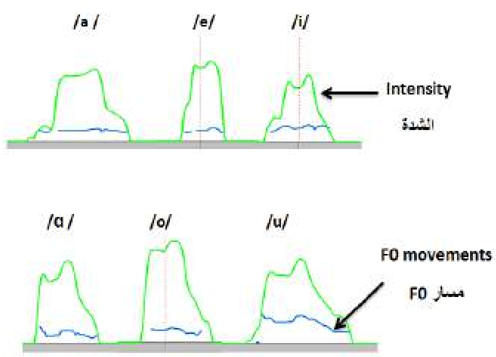


Figure: 3-1 Short vowels are higher in f0 values than long, especially high vowels [u, i]. even in shortening and lengthening processed vowels but the merged processed vowels were the lowest in f0 values, also F_0 contour shows irregularity

-4 F1 – F2 distance and the cognitive space:

[31] mentioned that [F1-F2 plot] is considered as a mental map that helps listeners to categorize the vowels precisely.

- In this study, formant frequencies of the perceived **original** short and long vowels were represented in the F1-F2 plot (see Figure: 3.2), where the long vowels are located on the edges of the F1-F2 plot map, while the short vowels were in the center.

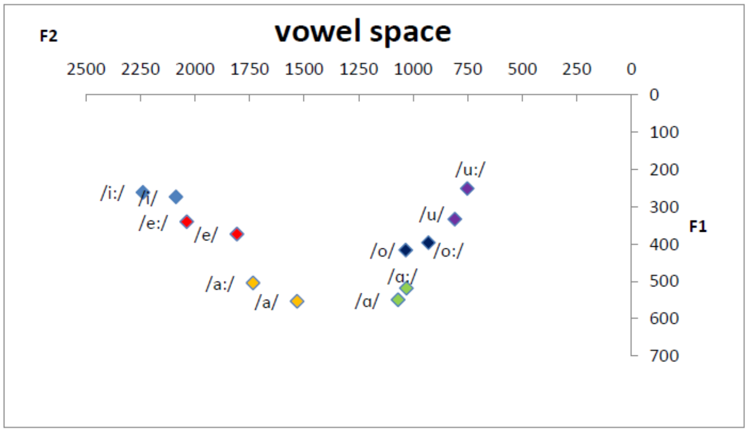


Figure:3.2 represents F1-F2 vowel space of the average values of the original short and long vowels.

- ➢ ***F1-F2 vowel space of the processed vowels; shortening, lengthening and merging* processes, it shows that:**
- ➢ *Long and lengthened vowels* specially the front vowels [æ, ɛ, i], are located on the edges of the F1-F2 plot map, while short, shortened and merged vowels tend to be in the center of the F1-F2 vowel space chart.
- ➢ *Back vowels* [u, o, ɑ], original and their processed ones tend to be close to each other with slight differences.

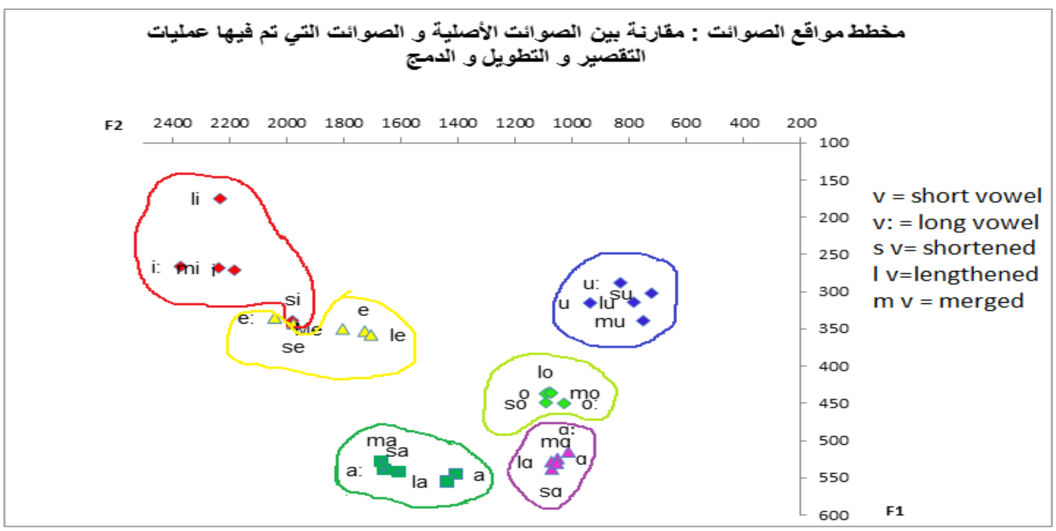


Figure: 3-3 indicates F1-F2 vowel space of original and processed vowels.

B. Perceptual Results

-1 Perceptual results Native Arabic listener

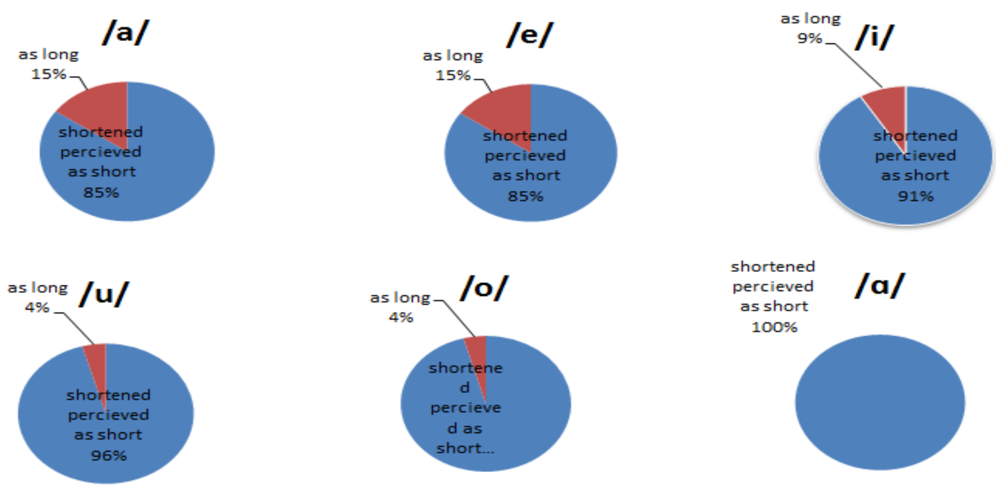
- *Shortening of the long vowels:*

In respect to the perceived ***length***; the shortened vowels were perceived as short ones with the percentage of about (96 – 100%) for the back vowels, and
(85 – 91%) for front vowels.

In respect to the perceived ***quality***; 42% of the listeners perceived a change in quality when the shortened vowels compared with the original ones, especially the quality of /æ/ vowel. (See Figure: 3-5)

Perceptual results for shortening process

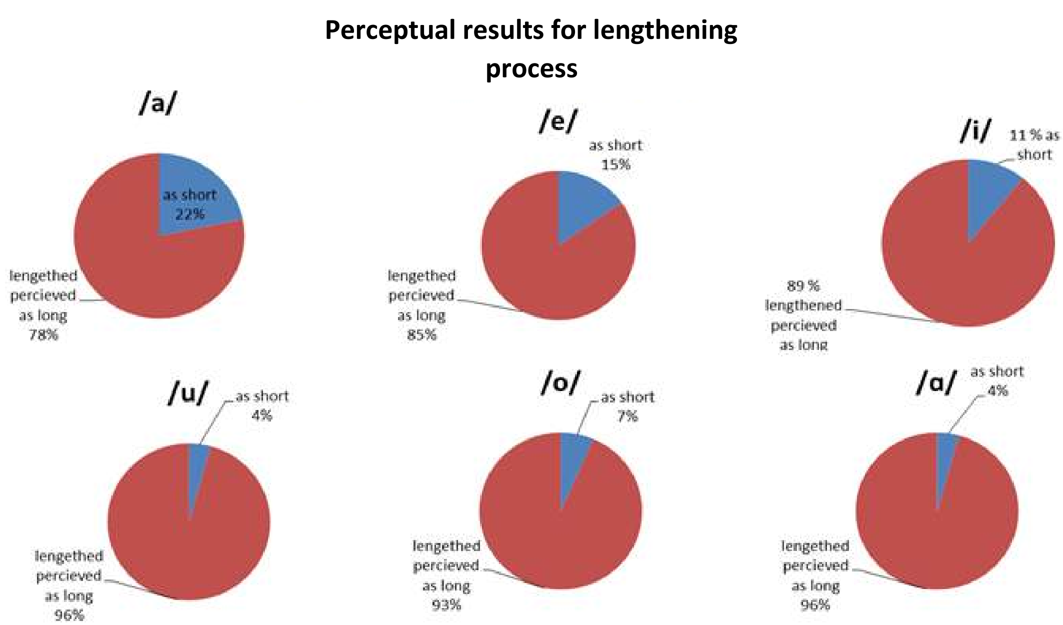
أ- النتائج الإدراكية لعملية تقصير الصائت



➢ *Lengthening of the short vowels*

In respect to the perceived ***length***; the lengthened vowels were perceived as long with high percentage values of about (78 – 96%). Back vowels were higher percentage than front ones.

In respect to the perceived ***quality;*** 40% of the listeners perceived a significant change between the lengthened vowels and the original long ones, especially for the front vowels. (see figure: 3-6)



➢ *Merging process*<u>;</u>

The listeners' judgments were at significant variance from the merged vowels quantity as well as their quality.

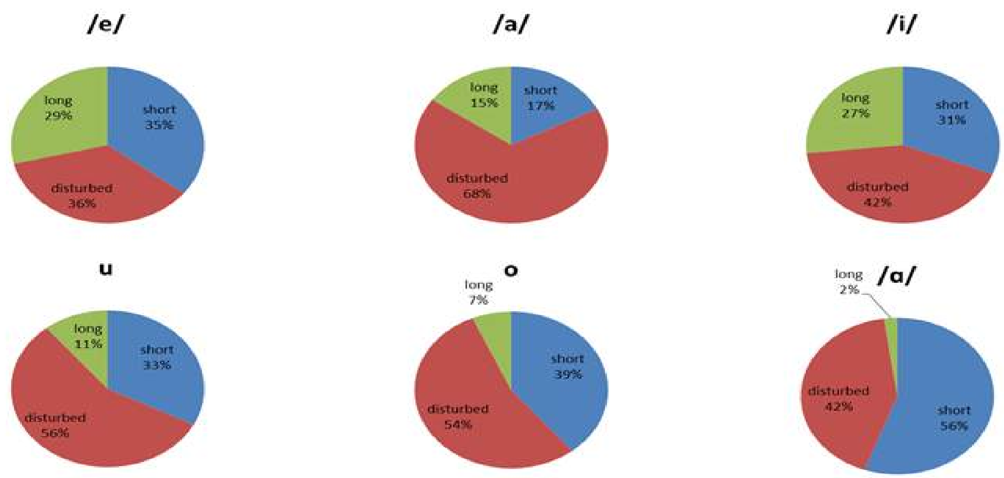
In respect to the perceived ***length;*** back merged vowels were perceived as long with percentage about (2%- 4%) and were perceived as short about (33% – 56%).

The front merged vowels were perceived as long about (15% – 29%) and as short about (17% – 35%). (See figure 3-7).

In respect to the perceived ***quality;*** all the merged vowels were perceived as disturbed and unclear- as the listeners' observations – with about (36 – 68%):

- The vowel /æ/ was the highest percentage, 68% of the listeners judged it as disturbed and described it as if a sequence of two different vowels.
- 56% of the listeners judged the back vowels [/u/ , /o/ , /ɑ/] as disturbed and described them as if there is a sound like glottal stop inside the vowel, e.g., / t^{ɑɑ}b / ------→ / t^{ɑ} ? ab /, / kuub / ----→ / ku ? ub /.
- 42% of the listeners perceived the front vowels / i / as disturbed and described them as if there is a glottal stop inside the vowel.
- For the vowel / ε / about 29% judged it as long, 35% judged it as short, and 36% judged it as disturbed.

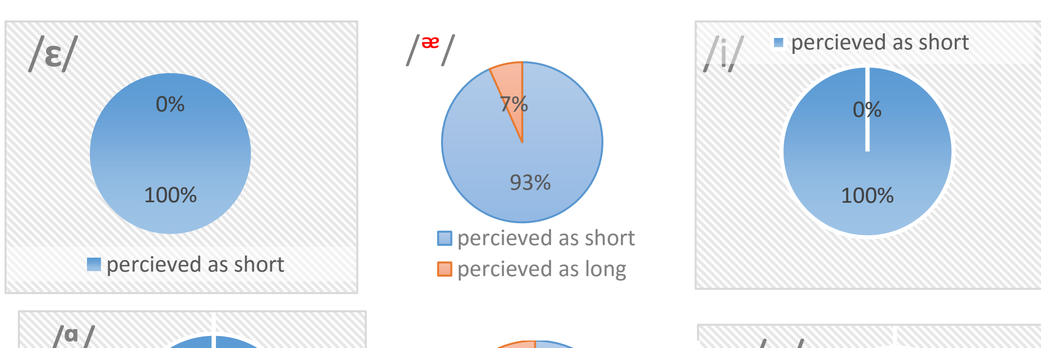
٣ . ج - النتائج الإدراكية لعملية تطويل الصائت القصير عن طريق الدمج



-2 Perceptual results for nonnative listeners

➢ *Shortening of the long vowels:*

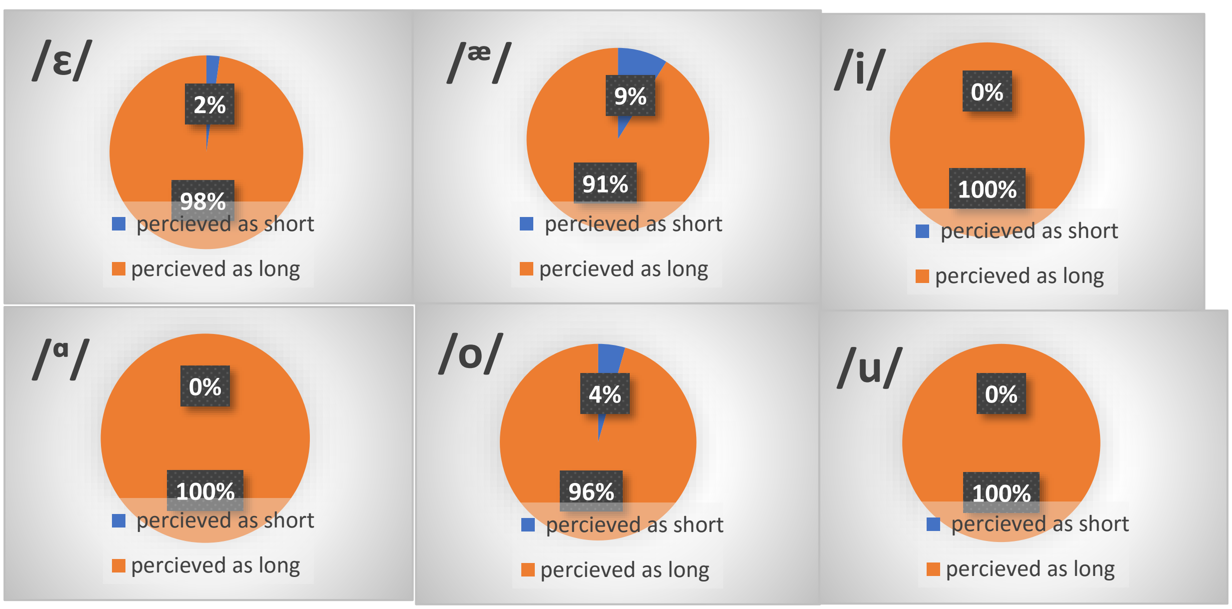
In respect to the perceived ***length***; nonnative listeners perceived the shortened vowels as short ones with high percentage values of about (93% - 100%). In respect to the perceived ***quality***; none of the nonnative listeners perceived change in quality when the shortened vowels compared with the original ones, (See Figure: 3.8).



➢ *Lengthening of the short vowels*

In respect to the perceived ***length***; nonnative listeners perceived the lengthened vowels as long ones with high percentage values of about (91 - 100%).

In respect to the perceived ***quality***; none of the nonnative listeners perceived any change in quality when the shortened vowels compared with the original ones, (see figure: 3-9).

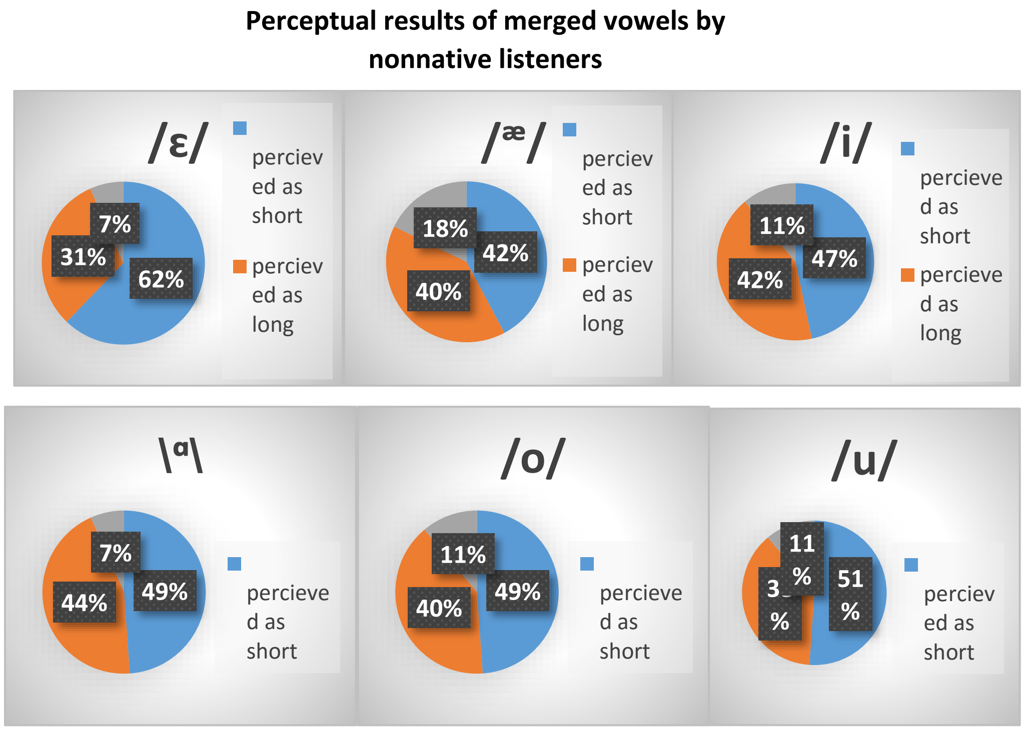


➢ *Merging process;*

In respect to the perceived ***length;***

- ✓ High percentage values range from (42% - 62%) of nonnative listeners who perceived all vowels lengthened by merging as a ***short*** vowel.
- ✓ Less percentage values range from (31%-44%) of nonnative listeners who perceived all vowels lengthened by merging as a long vowel (see figure 3-9)

In respect to the perceived ***quality;*** the lowest values percentage of nonnative listeners range from (7%-18%) who perceived all the merged vowels as disturbed and unclear- as the listeners' observations .



4 DISCUSSION

This part aims to correlate the results of acoustical measurements (i.e durational and spectral information) with listeners response to vowel length. As was mentioned in results, measurements proved that long and short vowels are different in their spectral features (f_0, F_1, F_2, intensity, pitch contour and waveform). Also, in the processes of shortening and lengthening, a slight change was occurred in these spectral features. For perceptual results, responses of native and nonnative listeners are naturally different.

A. Native Arabic listeners perception

When native Arabic listeners heard the original short and long vowels they could easily discriminate between them, this is because They have a prototype of long and short vowels stored in their minds according to vowel spectral features in their native phonetic system of Arabic.

➢ *In case of proceeded sample (shortening and lengthening)*

❖ When we made the shortening of the vowel by deleting a part of its steady state, and also lengthening short vowel by reduplicating length of its own wave, some acoustic information was changed. Though this change was a slight change, but it had a perceptual correlate.

❖ It was noticed that the Arabic native listener perceived back vowels differently from front vowels. In back vowels, most of them perceived proceeded shortened vowels as a short one in high percentages (96-100%) specially proceeded shortened vowel \ɑ\ which was perceived as a short vowel by all native listeners. While in front vowels, lower percentage of native listeners (85- 91%) perceived proceeded short vowel as a short one. In case of shortening long vowels, perception of proceeded front vowels which are shortened is more resistant in their perception than proceeded back shortened vowel. This is because that when we made shortening of the front vowels, their quality has been changed. Unlike back vowels which keep its original features. This is may be due to back vowels are reinforced by articulatory features like rounding, backness and emphasis for vowel \ɑ\, these features did not change during shortening or lengthening of the back vowel. While front vowels are characterized by easy articulation specially vowels \e, a\ which involve relaxation of articulators, and accordingly they are easily changeable in their quality

- ❖ Most of Arabic native listeners (85-100%) perceived proceeded shortened vowel as a short one and lengthened vowel as a long one i.e. they depended on duration for discrimination of vowel length. But many of them (40%) noticed a *quality change*. These results agreed with the studies [4], [5], [6], [7], [8] – [9] who cited that listener depend on spectral information (qualitative) as well as durational aspect (quantitative).
- ❖ Also, some of them perceived the proceeded shortened vowel as its origin long vowel but it is pronounced in a fast manner. For the proceeded lengthened vowel few of them described it as a short vowel but were pronounced in a prominent or slow manner.

➢ *In case of lengthening by merging two different waves:*

- ❖ (lengthened short vowel was done by merging its wave to a wave of original long vowel), it was noticed that differences were resulted acoustically and perceptually.
- ❖ Acoustically, the resulted merged vowel is characterized by disturbance in shimmer and jitter, pitch contour, intensity contour, waveform and spectrographic analysis.
- ❖ Accordingly this disturbance was perceived by native listeners as a disturbed vowel i.e. the highest percentage of listeners described the merged vowel as a disturbed vowel in all six vowels.

Hence;

- ✓ The difference between long and short vowel is not only a difference in duration, but also it involves difference in spectral characteristics
- ✓ Arabic native listeners do not depend only on duration in perceiving vowel length, but also, they depend on spectral information too.
- ✓ This was strongly proved in the process of merging long with short vowel, in which listeners made what is called perceptual cue weighting. They tended to depend on spectral characteristics more than durational characteristics in their perception of the proceeded merged vowel. And accordingly, because the disturbance that happened in the spectral features of the proceeded merged vowel, they perceived it as a disturbed vowel.

.

B. Nonnative Arabic listeners

Perception of nonnative Arabic listeners (Indonesian listeners) was different from perception of native Arabic listeners. they depended basically on durational differences in their perception of vowel length.

- ❖ They perceived proceeded short vowel as a short vowel
- ❖ And perceived proceeded long vowel as a long vowel
- ❖ In merging vowel, the highest percentage of them perceived proceeded long merged vowel as a short vowel, also a lower percentage of them were perceived it as a long vowel. while the lowest percentage of them perceived it as a disturbed vowel

This was due to that Indonesian language has no long vowel in its phonetic system. It has only short vowel [32]. So, vowel length is not a distinctive phonemic feature in the Indonesian language.

They could perceive long and short Arabic vowels (and also shortened and lengthened vowels) correctly by depending only on duration of the vowel. They just hear a long timing and a short timing of the same vowel, without any notice that long vowel is a pattern that differed from the short vowel. unlike native speaker who depended on spectral information and noticed a quality change in some vowels and also noticed that the vowel identity did not change by shortening (as they described it was pronounced in a fast manner when it was shortened) and also in lengthening vowel (as they described it was pronounced in a stress or slow manner when it was lengthened.)

When nonnative listeners heard merged vowels, firstly most of them perceived it as a short vowel because they enclose what they hear to their native phonetic system, and because they have no long aspect of vowels in their native language, they perceived it as a short vowel. Secondly other group of them who perceived merged vowel as a long vowel, this was because that they only depend on duration and heard just a long timing vowel. The lowest percentage of them perceived it as a disturbed vowel – unlike the previous two cases – they could notice some spectral features that is presented in the disturbance that occurred in the process of merging. These results agreed with [33] study who investigated the acquisition of long and short vowels of Japanese language by 20 Indonesian speakers, results showed that They made few errors in listening to short vowels, however, many errors in long one, and final positioning was very difficult. [32] assumes that Indonesian perhaps will generalize the English vowels with the Indonesian vowels that they have where only short vowels exist.

Indonesian has only six vowels, which are [I], [ə], [ɑ], [o], [u], [e], and three diphthongs and [32] showed that Pronunciation problems faced by foreign language learners are caused by differences found between the learners‟ language and the target language. The students have tendency to pronounce English vowels as the way they pronounce vowels in their mother tongue. The students also tend to pronounce the word longer if the word has double same vowels, even though the word truly is not pronounce long. The students also tend to pronounce every word shortly if they have only one vowel. Likewise; in this research when Indonesians listeners heard long duration, they respond it as long vowel, and when they heard short duration they respond it as a short one, but in the case of listening to the merged vowels they responded it as a short vowel because only short vowel exists in their native phonetic inventory.

5 CONCLUSIONS:

Conclusions of this study could be summarized as follows:

- ✓ ***In Arabic, the acoustic differences between long and short vowels are not only durational features, but also in spectral information. i.e., quantity and quality.***
- ✓ ***Perceptually, however, native Arabic listeners tend to depend on length of the vowels as well as spectral information.***
- ✓ ***Vowel length is not a distinctive phonemic feature in some languages like Indonesian language. Accordingly, nonnative listeners of Arabic could discriminate between long and short vowel by depending only on duration (unlike native listeners of Arabic).***
- ✓ ***From the afore-mentioned perceptual percentages, shortening and lengthening of the vowel from their own waveform showed a significant change in quality by natives, but no significant values by non-natives.***
- ✓ ***Merging two different wave qualities showed high significant values of perceptual percentages by natives, while non-natives showed insignificant ones.***
- ✓ ***The high vowels /i/ and /u/ are more resistant in changing than other vowels by natives and nonnatives. This is proved in the original vowels as well as all kinds of processed vowels.***
- ✓ ***Finally, in second language acquisition, it is very important to consider influence of native language on the listener who learn another foreign language***

ACKNOWLEDGMENT

Great thanks to Phonetics and linguistics department specially to our professor Dr. Mervat Fashal for her guidance and regular support. And Dr. Sameh Alansary - the head of the department- for his permanent encouragement . Another thanks to students of phonetics department who participated in this this experiment, and Indonesians students in “The secondary extension institute of Arabic and Islamic studies in Egypt".

BIOGRAPHY

Eman Mohammed Mohammed Kassem

- ***Demonstrator*** in phonetics and linguistics department 2003, **M.A in Experimental Phonetics "Influence of punctuation marks stemming of suprasegmental features in understanding read-text " Psycho-acoustic study 2010.**
- ***Assistant lecturer*** in phonetics and linguistics department 2011
- **PHD in Experimental Phonetics "Contributions of Nasals and Vowels to Speaker Identification in Arabic"2014**
- ***Lecturer*** in phonetics and linguistics department 2015
- ***Academic Advisor*** in the Forensic Phonetic and linguistic Program in 2018.
- ***Academic supervisor*** in the Forensic Phonetic and linguistic Program from September 2019 untill now.
- ***Teaching experience in*** different subjects; Experimental Phonetics, Speech Perception, Acoustic Phonetics, Physiological Phonetics and Auditory Phonetics.

Lamiaa Mohammad Tawfik Aldawy

- ***Postgraduate researcher*** in phonetics and linguistics department
- **B.A in phonetics and linguistics department in 2010**
- **M.A in Experimental Phonetics "perception of durational differences of vowel length in Arabic an experimental study "2016.**

الملخص العربي

إدراك طول الصوائت فى اللغة العربية عند المتحدث الأصلى وعند متعلمى اللغة العربية من غير الناطقين بها

دراسة تجريبية

[1*] إيمان محمد قاسم, [2**] لمياء محمد توفيق

[1*]قسم الصوتيات و اللسانيات , كلية الآداب, الشاطبي , جامعة الإسكندرية

[1]Eman.kassem@alexu.edu.eg

[2**]قسم الصوتيات و اللسانيات , كلية الآداب, الشاطبي , جامعة الإسكندرية

[2]lamyaatawfik@hotmail.com

الملخص

يعد طول الصائت المدْرك - في العديد من اللغات - من العلامات الفارقة للتمييز الصوتي؛ حيث يتضح ذلك في إدراك المستمع لنفس المنطوق بطريقتين مختلفتين إعتماداً على الطول الزمني للصائت ونوعية الصائت.

قام فراي (١٩٥٦) بإلقاء الضوء على الفرق بين الطول الزمني والطول الإدراكي للصائت، حيث أن الطول الزمني للصائت هو الوقت الفعلي الذي يستغرقه الصائت أثناء نطقه، بينما الطول الإدراكي للصائت هو الطول المدرك للصائت من قبل المستمع.

إن اختلاف الطول المدرك للصائت في اللغة العربية يؤدي إلى تغييرات في المعنى؛ ولذلك تعتبر علامة فارقة في لغتنا العربية. ولكن إلى أي مدى يمكن أن يكون هناك إختلافات إدراكية وفيزيائية (قياسية) بين الصوائت القصيرة والطويلة؟ وما هي السمات الصوتية التي قد يعتمد عليها المستمعون العرب والمستمعون الإندونيسيون للتمييز بين الصوائت القصيرة والطويلة؟

يهدف هذا البحث إلى دراسة إدراك كل من مستمعي اللغة العربية الأصليين والمستمعين الأجانب لطول الصائت في اللغة العربية من خلال دراسة تجريبية؛ وتوضح الدراسة أيضاً مدى إعتمادهم على الصلة المعرفية المخزنة/ المدركة لديهم لطول الصائت، من خلال تسجيل إستجاباتهم على طول الصائت المعالج.

REFERENCES

[1] *Fry, D.B. (1956). 'Duration and intensity as physical correlates of linguistic stress'. Journal of the Acoustic Society of America 27, 765-768.*

[16] *Escudero, P. (2001). The role of the input in the development of L1 and L2 sound contrasts: language-specific cue weighting for vowels. Proceedings of the 25th Annual Boston University Conference on Language Development, Cascadilla, pp. 250-261.*

[17] Tsukada, K., (2009). An acoustic comparison of vowel length contrasts in Arabic, Japanese, Thai: Durational and Spectral data, International Journal on Asian Language Processing; 19:127-138.

[18] Tsukada, K., (2013).Vowel length categorization in Arabic and Japanese: Comparison of native and non-native Japanese perception, Speech, Language and Hearing, 16(4):187-196.

[19] Kinoshita, K., Behne, D., M., and Takayuki, A., (2002). Duration and F_0 as perceptual cues to Japanese vowel quantity, 7th international conference on spoken language processing, Denver, Colorado, USA: 757-760.

[20] Gottfried, T., L and Beddor (1988). Patrice speeter: Perception of temporal and spectral information in French vowels, Haskins laboratories

[21] Sawusch. J., R., (1996). Effects of duration and formant movement on vowel perception, ICSLP, vol.4: 2482- 2485,

[22] Kassem, E., M. (2014), Contributions of nasals and vowels to speaker identification, PhD thesis, Alexandria university.

[23] Abramson, A.S. & Ren, N. (1990). Distinctive vowel length: Duration versus spectrum in Thai. Journal of Phonetics, 18, 79-92.

[24] Dauer,R.M.(1983), "stress-timing and syllable-timing reanalyzed", Journal of Phonetics,11:51-62.(in reference at the end)

[25] *Wahba.k, Sh., (1988) ,The acoustic analysis of colloquial egyptain Arabic vowels: An experimental study, M.A thesis, Alexandria university.*

[26] Saadah, E (2011). The production of Arabic vowels by English L2 learners and heritage speakers of Arabic. PHD.

[28] Hillenbrand, J., M., Clarkm, M.J., Houde, R.A. (2000). Some effects of duration on vowel recognition, J Acoust Soc Am; 108(6):3013-22.

[26] Mady, K & White, L (2008). The long and the short and the final: Phonological vowel length and prosodic timing in Hungarian, Brazil, Speech Prosody, ISCA Archive: 363-366.

[30] Hadding-Koch, K. & Abramson, A. S. (1964). Duration versus spectrum in Swedish vowels: Some perceptual experiments. Studia Linguistica, XVIII, 94-107.

[31] Hayward, K., (2000). Experimental Phonetics -. England, pearson Education Limeted.

[32] *Riadi, A, Rufinus, A, Novita, D, (2013). STUDENTS' PROBLEMS IN PRONOUNCING SHORT AND LONG ENGLISH VOWELS, English Education Study Program, Language and Art Education Department Teacher Training and Education Faculty of Tanjungpura University, Pontianak*.

[33] Franky R, YOKOYAMA, N, ISOMURA, k, USAMI,Y, KUBOTA, Y., (2012). The Acquisition of Japanese Vowel Length Contrast by Indonesian Native Speakers: Evidence from Perception and Production, Journal of the Phonetic Society of Japan, Volume 16 Issue 2 Pages 28-39 available at https://www.jstage.jst.go.jp/article/onseikenkyu/16/2/16_KJ00008228944/_article

المراجع العربية :

[٢] بريتل مالمبرج. الصوتيات \ ترجمة محمد حلمي هليل- الإسكندرية : عين للدراسات و البحوث الإنسانية و الإجتماعية, ١٩٩٤

[٣] محمد صالح الضالع. الصوتيات والفونولوجيا مقدمة معاصرة للقارئ العربي.- جامعة الإسكندرية, ٢٠٠٣

[٤] سعد عبد العزيز مصلوح. التناسب الزمني بين الحركات القصيرة و الطويلة دراسة صوتية معملية في القافية العربية. مجلة معهد اللغة العربية.ع ٢, ١٩٨٤

[٥] أحمد مختار عمر. دراسة الصوت اللغوي.- القاهرة: عالم الكتب ١٩٩١.

[٦] غالب فاضل المطلبي. في الأصوات اللغوية دراسة في أصوات المد العربية.- العراق: دائرة الشئون الثقافية و النشر, ١٩٨٤

[٧] عبد الحميد زاهيد. حركات العربية دراسة صوتية للتراث العربي/ تقديم التهامي الراجي الهاشمي.- مراكش, المطبعة و الوراقة الوطنية, ٢٠٠٥

[٨] محمد سالم الرجوبي. الحركة الإعرابية بين القيم الصوتية و القيم الدلالية. مجلة الجامعة الاسمرية, ع ٢٠, ليبيا, ٢٠١٤

[٩] سلمان حسن العاني. التشكيل الصوتي في اللغة العربية فونولوجيا العربية.- جدة. النادي الأدبي الثقافي, ١٩٨٣

[١٠] عبد الفتاح عبد العليم البركاوي. مقدمة في علم الأصوات العربية. القاهرة – كلية اللغة العربية, ٢٠٠٤

[١١] أبي الفتح عثمان ابن جني / تحقيق د. حسن هنداوي. سر صناعة الإعراب ,الجزء الاول.- دمشق: دار القلم ١٩٩٣.

[١٢] كمال بشر. علم الأصوات.- القاهرة : دار غريب للطباعة والنشر والتوزيع , ٢٠٠٠

[١٣] منصور محمد الغامدي. الصوتيات العربية.- الرياض: مكتبة التوبة, ٢٠٠١

[١٤] محمد احمد زكي. المَدْ في العربية دراسة صوتية موجزة. مجلة جامعة بابل للعلوم الصرفة والتطبيقية. م ١٩, العراق, ٢٠١١

[١٥] يحيى علي أحمد. طول الحركة في اللغة العربية وعلاقته بالبنية المقطعية. مجلة جامعة دمشق, م ٢٩, ع ٣+٤, ٢٠١٣.

[٢٧] إبراهيم أنيس. الأصوات اللغوية.- القاهرة : مكتبة انجلو, ١٩٧١

An Investigation of the Correlation between Perceived Pauses and Syntactic Structures

Israa Elhosiny*1, Mervat Fashal*2

*Phonetics and Linguistics Department, Faculty of Arts, Alexandria University, Alexandria, Egypt

[1]israa_elsayed@yahoo.com

[2]Mervat.fashal@alexu.edu.eg

Abstract— **Natural-sounding speech synthesis requires close control over the temporal structure of the speech flow. Therefore, the relationship between prosodic phrase boundaries in terms of pausing and the syntactic structure has been investigated in read dialogue and continuant speech in Arabic. Both the speakers' production and the listeners' perception of pausing are considered and mapped to the syntactic structure. In order to describe the correlation between pauses and phrase boundaries in the proposed speaking styles, recall and precision rates were counted.**

Keywords: **speech production, speech perception, phrasal categories**.

1 INTRODUCTION

> ***"In the speech sound wave, one word runs into the next seamlessly; there are no little silences between spoken words the way there are white spaces between written words. We simply hallucinate word boundaries when we reach the end of a stretch of sound that matches some entry in our mental dictionary."***
>
> **(Pinker, 1995*)***

Pauses are useful for the speaker [1] and for the listener [2]. In fact, there is a positive correlation between the need for pause time during the speech sequence and the level of processing required by the task. Lists of digits and letters presented at fast rates or at low signal-to-noise ratios are recalled more accurately when pauses are inserted into the stimulus sequence. Similarly, speakers tend to increase the ratio of speech to pause time as the utterance becomes more complex.

Grammatical pauses appear to be more associated with different types of processing than the non-grammatical pauses. They occur between clauses and used for structural and semantic long-term planning in speech production. By contrast, non-grammatical or within-clause pauses are concerned with last-minute word selection [3].

Prosodic features are not sufficient, to analyze and generate the structure of the speech, but also grammatical analysis on many linguistic levels. Several researchers have discussed how prosody, morpho-syntax and discourse structure are related to each other. Linguistic structure has been shown to play a vital role in pausing strategies which signal information flow of the utterance, thereby helping the listener to interpret the message uttered by the speaker [4]. Swerts and Geluykens [5] found that speakers in monologues use pauses of various lengths to signal information flow in terms of topic structure. Shriberg et al [6] reported that new topics are often realized by some combination of silent pauses, low boundary tones and/or pitch range resets in English. Also, Hirschberg [7] argued that phrases introducing a new topic can be characterized by an initially wider pitch range preceded by a longer pause, and on average they are louder and slower than other phrases. Van Donzel [8] studied prosodic features of discourse boundaries for Dutch on the basis of clause, sentence and paragraph division, as well as the prosodic features of information structure in terms of the New–Given taxonomy. She found that discourse boundaries in spontaneous speech are realized by silent pauses and high boundary tones. These studies show that there is a relationship between prosody and (at least) higher linguistic structure, such as discourse in terms of topic, theme, and New–Given taxonomies.

Moreover, several researchers have investigated the relationship between morpho- syntactic structure and prosody. Most of the studies deal with the automatic prediction of prosodic phrase boundaries, given some linguistic information, used in text–to– speech systems. Some studies show that full syntactic analysis is not needed for the prediction of prosodic boundaries, while others claim the opposite. For example, in text–to–speech systems, phrase breaks are often predicted by distinguishing between content and function words. For prosodic phrase boundary detection, detailed but incomplete syntactic analyses were used by Bachenko and Fitzpatrick [9] by implementing the Phi rule-based algorithm developed by Gee and Grosjean [4]. Wang and Hirschberg [10] as well as Ostendorf and Veilleux [11] used PoS and syntactic constituent structure together with some acoustic information (such as pitch accent, phrase duration, and position to the last break) to predict phrase breaks. However, Ostendorf and Veilleux reported that good phrase prediction can be achieved without using any detailed PoS, or syntactic information. Taylor and Black [12] assigned phrase breaks on the basis of part-of-speech sequences only, although they suggested that syntactic parsers giving reliable parse trees might facilitate phrase break assignment. These studies show that prosodic phrase boundaries do not necessarily correspond to syntactic phrase boundaries. Most of the researchers agree that there is a relation between prosody and syntactic structure on one hand, and between prosody and discourse structure on the other hand. However, most of the studies performed on this topic investigate one of these relations either for non–spontaneous or for spontaneous speech.

The aim of this study is to investigate some aspects of the relation between the prosodic, and syntactic structure in spontaneous as well as in non–spontaneous speech. Additionally, both the speakers' production and the listeners' perception of pausing are considered and mapped to the linguistic structure. For spontaneous speech, we use a continuant speech, and for non-spontaneous speech, we study the acted dialogue for a children story. We investigated the pausing strategies in the speaking styles in terms of syntactic phrasal boundaries. Next, we will describe the data and method used for investigating the relationship between pausing and linguistic structure.

2 DATA AND METHOD

In order to investigate, primarily, the relation between pauses and syntactic structure, a modern standard Arabic speech sample is the most suitable material. The following subsections presents how the speech sample collected, prepared and analyzed.

A. Speech Material

As the Modern Standard Arabic; MSA, is the representative method used in Arab world, the researcher preferred to start the syntax-acoustic interfaced research by the MSA not by colloquial Arabic. MSA considered as more regular then colloquial. We selected an acted children story as a dialogue to represent the non-spontaneous speech. Moreover, a continuant speech has been selected to represent the spontaneous speech. The speech sample of dialogue is segmented into sentence; a sentence is defined as a turn. The first 10 sentences were selected which represent 106 words. For spontaneous speech, the first 9 sentences from a continuant speech have been selected. The sentence boundaries have been detected by the length of sentence;

the maximum length for the sentence is 18 words. The number of words in spontaneous speech was 171 words.

B. *Syntactic boundaries annotation*

The research assumption is, pauses may split some phrasal units in Arabic. Speech sample has been written as a text to enable syntactic annotation. Table (1) expresses some annotated examples of the expected phrasal boundaries (bold font). The last word from each expected phrase has been marked beside its final word position. Around 30% of sentences' words in the overall data. Both speaking styles, have been marked manually as phrasal boundaries. The ratio between marked and unmarked phrasal boundaries has been represented as in figure (1).

TABLE 1: EXAMPLES FOR MARKED PHRASES.

phrase	*Function*	*example*
NP: Vocative particle + **NP**	Vocative style	أيتها الحمامة **الضعيفة**
VP: V $+_{obj\ pron}$ + **N**	Verbal phrase	أكلتكم **جميعا**
NP: N + Prep + **NP**	2nd Verb argument	ألقي إلى واحدا من **صغارك**
VP: V + **PP**	2nd Verb argument	صعدت **إليكم**
JP: J + **A**	absolute Object	ماهر **جدا**
PP: Prep + N + **NP**	Predicate argument	ماهر في تسلق **الأشجار**
VP: V + PP + **NP**	Temporal noun	آتي إليكم صباح **غد**

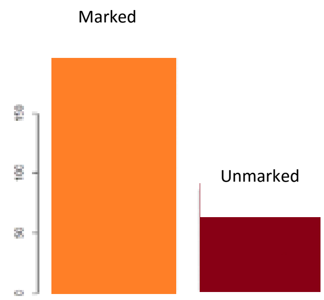


Figure 1: Marked vs unmarked phrase boundaries.

C. *Acoustic Detection of Pauses*

In the traditional linguistic definition, normal speech flow is considered to be interrupted by a physical pause whenever a brief silence can be observed in the acoustic signal (i.e., a segment with no significant amplitude). Which exact duration of the silence is considered sufficient for the constitution of a physical pause depends on its linguistic context [13]. Intra-segmental pauses are related to the occlusions of the vocal tract in normal speech production. Example: In the word “happy” as in Figure (2), the pause component of the Voice Onset Time (VOT) for the consonant /p/ corresponds to a silence of 96 ms [14].

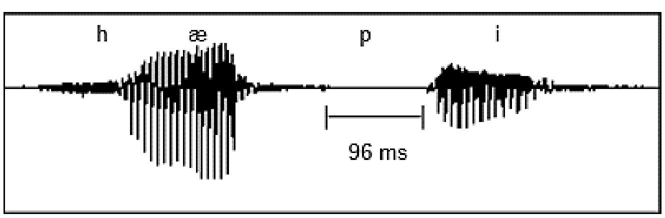
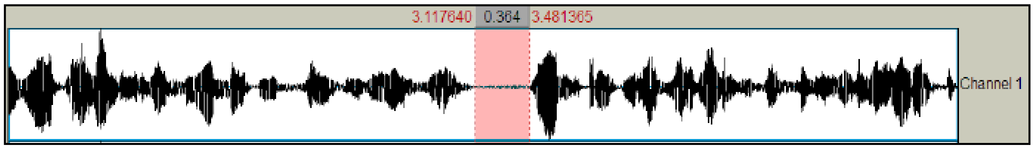


Figure 2: The acoustic signal of “happy” showing an intra-segmental pause.

Both speaking styles samples have been segmented into spoken sentences in isolated .MP3 formatted files. A database has been built to store sound files and their written sentences with their IDs. In addition, the acoustic signals for each sentence have been stored as a picture file. Sentences were segmented into written words to be stored in the data base. The database designed as it is the best way for data handling, retrieval, exploration and calculation. The words table inside the database is linked to sentences table to store information about phrasal boundary, silence after the stored words, etc.

Observed silence intervals; more than 100 ms, are measured by Praat for each sentence' words and stored beside word it proceeded as in figure (3).

ID	sound file	sentence	Pattern	wave form	high_pitch
1	Package	أيتها الحمامة الضعيفة ألقي إلي واحدا من صغارك وإلا صعدت إليكم وأكلتكم جميعا فأنا ماهر جدا في تسلق الأشجار.	NP VP VP[VP COO VP] NP[NP ADJP[ADJP PP]]	Package	368



ID	Word	duration_pause	constituent
2	الحمامة		
3	الضعيفة	0.364	NP
4	ألقي		
5	إلي		
6	واحدا		
7	من		
8	صغارك		NP – COO
9	وإلا		
10	صعدت		
11	إليكم		PP – short – EOF (VP)

Figure 3: database and speech information storage

3 Results and Discussion

In order to hold a comparison of pauses between the two different speaking styles, we investigate three different dimensions of communication; production, perception, and context. The following subsections discuss the relationship between the production of pauses and the linguistic context in which pauses appear, as well as the perception of the pauses and the linguistic environment in which people actually perceive them for the two speaking styles: acted dialogs and continuant speeches.

D. Production and linguistic features

In this section, the silent intervals characteristics detected in the two speaking styles will be described with special attention directed to the syntactic context in which pauses appear. The mean duration and word/pause ratio of silent intervals are calculated as in table (2). There are differences in the duration and frequency of pauses between the two speaking styles. Hence, the time it takes to pronounce a word on average differs between the styles suggesting greater variation in speech tempo. Our results, indicating the variation of pausing patterns across speaking styles.

TABLE 2: FEATURES OF ACOUSTIC PAUSES

Speaking style	***Pause Duration***	***Word/Pause Ratio***

Acted dialog	435.6 ms	4.25
Continuant speech	963 ms	3.24

For describing the correlation between silent intervals and supposed phrasal boundaries in the two speaking styles, recall and precision rates were counted. Precision and recall are calculated in terms of positive and negative classifiers. The predicted condition can be one of the following:

TN / True Negative: case was negative and predicted negative

TP / True Positive: case was positive and predicted positive

FN / False Negative: case was positive but predicted negative

FP / False Positive: case was negative but predicted positive

Precision and recall are then defined as in Definitions (1) and (2) [15]:

$$(1)\ \textit{Precesion} = \frac{TP}{TP + FP}$$

$$(2)\ \textit{Recall} = \frac{TP}{TP + FN}$$

Recall (R) describes the percentage of acoustic pauses that appear in boundary positions, see Definition (3). Precision (P), on the other hand, gives the percentage of phrasal boundaries that corresponds to pauses in the speaking styles, see Definition (4).

$$(3)\ P = \frac{\textit{Turly predicted pauses}}{\textit{Truly predicted pauses} + \textit{wrongly predicted pauses}}$$

$$(4)\ R = \frac{\textit{Turly predicted pauses}}{\textit{Truly predicted pauses} + \textit{unpredicted pauses}}$$

The correlation between silent intervals and phrasal boundaries for each speakingstyle is shown in Figure (4). It is obvious that phrasal boundaries and pauses do not always coincide, and we find clear differences between the speaking styles. In the acted dialog style, pauses often appear at a phrasal boundary positions (shown by the high recall = 100%). In the continuant speech, greater number of pauses existing in a non-phrasal boundary positions (lower recall= 93.6%). On the other hand, looking at the phrasal boundaries and their acoustic correlate in terms of silent intervals (i.e. the precision rates) we find that in dialog, 6 predicted phrasal boundaries were not followed by a pause (80% precision), although the continuant speech do this to a greater extent (88% precision)

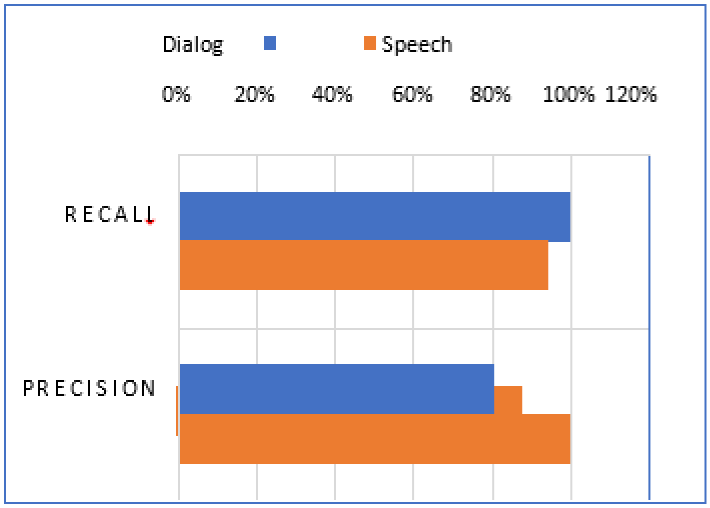


Figure 4: Recall and precision rates for acoustic pauses and phrasal boundaries in acted dialog and continuant speech.

For the Syntactic context of the acoustic pauses, the majority of silent intervals can be found at sentence boundaries. In acted dialog, pauses appear entirely at sentence boundaries and between phrases, e.g. in after noun phrases (21%), adverb phrases (6%), verb phrases (12%). While in continuant speech, after noun phrases (26%), adverb phrases (2.2%), verb phrases (24.4%) and prepositional phrases (17.7%).

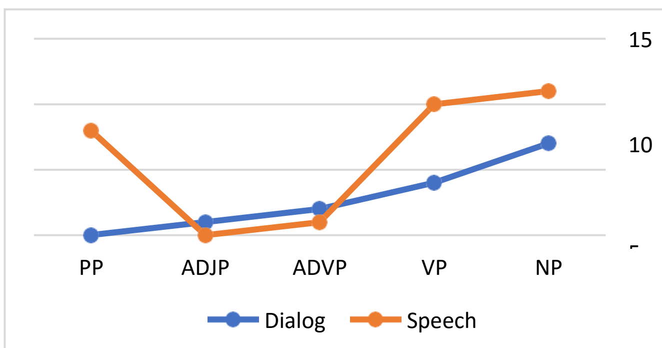


Figure 5: The ratio between occurrences of phrase boundaries in both speaking styles.

E. Perception and linguistic features

In order to investigate how often and in what linguistic context people actually perceive silent intervals, the frequency and position of the perceived pauses were examined. The distribution of the perceived pauses, labeled by the eleven subjects, are to a large extent evenly distributed across the speaking styles (as opposed to the distribution of silent intervals). The average "words per perceived pauses ratio" is highest in acted dialog (4.1) followed by the continuant speech (3.2) as in figure (6).

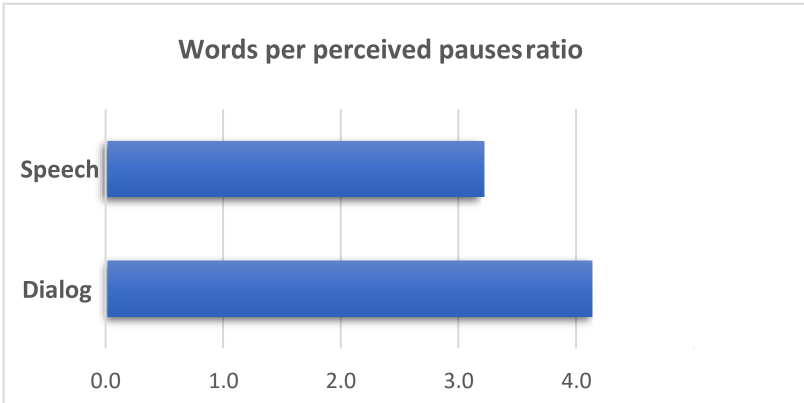


Figure 6: The words/perceived pauses ratio.

Concerning the relation between perceived pauses and phrasal boundaries, recall and precision rates are given in Figure (7) using definitions (1) and (2). The results reported here are based on instances where at least 7 of the 11 subjects perceived a pause.

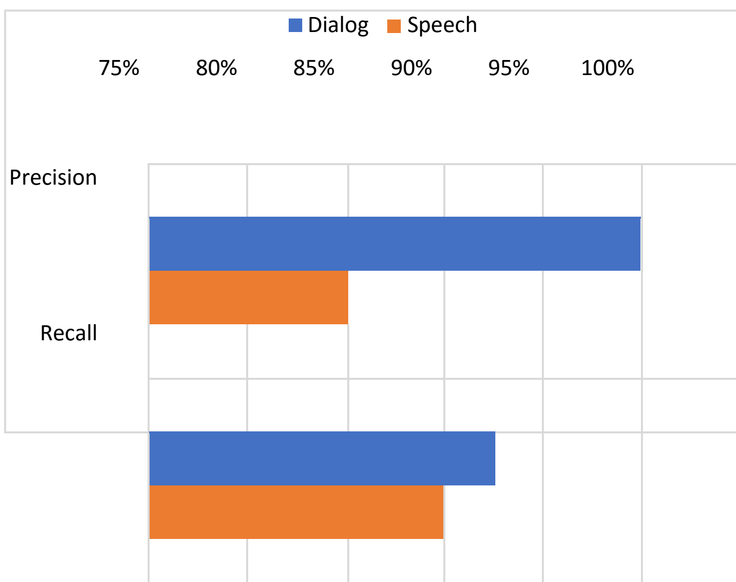


Figure 7: Recall and precision rates for perceived pauses and phrasal boundaries in acted dialogs and continuant speeches.

In the acted dialog style, the majority of the perceived pauses are located at phrasal boundaries (high recall = 93%), while in the continuant speech, we found the majority of perceived pauses located between words positions (low recall = 89%). It is observed that, the perceived pauses are rare at phrasal boundary positions in spontaneous speech, as shown by low precision (85.48%), while more frequently occurring in the acted dialog (100%). Additionally, in the acted dialog style, pauses are perceived at sentence and phrase boundaries. In continuant speech, pauses are located between phrases, e.g. in front of NPs, AdvPs, PPs, conjunctions or verbs, as well as within phrases. Even though the linguistic context of the perceived pauses is similar to the context described for the silent intervals for the various speaking styles, the acoustic and perceived pauses do not necessarily overlap [16].

F. Production and perception of Pauses

To give an overall picture of the correlation between the silent intervals and pauses perceived by each of the 11 subjects participating in the listening test, recall and precision are measured. Here, recall (R) describes the percentage of the acoustic pauses that were actually perceived, while precision (P) gives the percentage of perceived pauses that corresponds to acoustic silence.

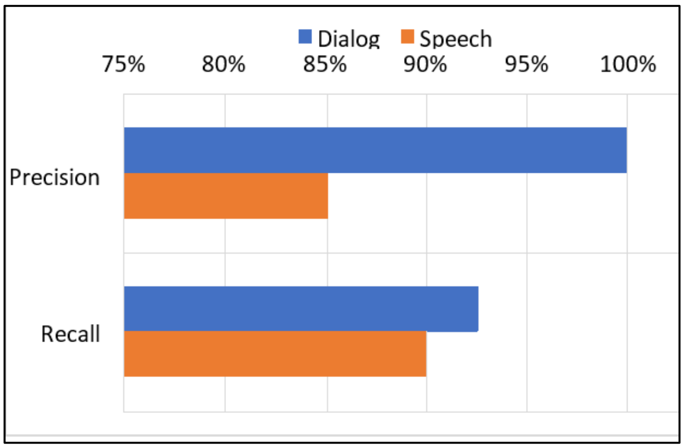


Figure 8: The correlation between acoustic and perceived pauses in the three speaking styles.

The results reported here are based on instances where at least 7 of the 11 subjects perceived a pause. It is clear that the correlation of the acoustic and perceived pauses varies across the speaking styles. In the acted dialog, a considerable number of pauses are perceived by the subjects (100% recall), but many of the perceived pauses do not have any correlates in acoustic silence (93% precision). In the continuant speech, the majority of the acoustic pauses are perceived by the listeners, and many of the perceived pauses actually have an acoustic correlate (represented by higher precision = 96.36%).

4 CONCLUSION AND FUTURE WORK

In this study, we investigated the phenomena of pausing in two different speaking styles in Arabic: acted dialogue, and continuant speech. Additionally, we examined the syntactic context that corresponded to intervals of acoustic silence and listener perceived pauses. Our results show large differences across the speaking styles. In the acted dialog, all acoustic silence intervals are perceived by the listeners, but there is a number of perceived pauses do not have an acoustic correlate in silence. In the continuant speech, the majority of the acoustic pauses are perceived by the listeners, and many of the perceived pauses actually have an acoustic correlate. Considering the syntactic environment in which the acoustic and perceptual pauses appear, we observed that small silence intervals is perceived: if it occurs in Connection to phrasal boundaries, while if this small silence is found in the middle of phrase, the listeners do not perceive those intervals as pauses. Not surprisingly, we also showed that pause length have an effect on the listeners' perception; the longer the silent intervals are, the better the chance that the perceived pause is actually an acoustic silent interval. Questions we find important to explore in future work concerning syntactic variation in connection to pausing structure. Since the speech sample is not large enough to enable predicting more syntactic boundaries necessary for pausing. Other fields for future work include the investigation of the relation between the duration of silence interval and the type of the phrasal boundaries. We propose in the end of this study, to complete working on enlarging and improving the speech-syntactic database and enhancing it with different kinds of linguistic information such as semantic and pragmatic. Integrated speech- linguistic database will open the way for better language understanding and speech synthesis in the future.

REFERENCESD.

[1] F.Goldman-Eisler, Psycholinguistics: Experiments in Spontaneous Speech. Academic Press, New York, 1968.

[2] D.Aaronson, *Temporal course of perception in an immediate recall task.* J. Exp. PsychoI. 76:129-140, 1968.

[3] F. Goldman-Eisler, *Speech production and the predictability of words in context*. Q. J. Exp. Psychol. 10:96, 1958.

[4] J. P. Gee, and F.Grosjean, *Performance Structures: A Psycholinguistic and Linguistic Appraisal*. Cognitive Psychology, pp. 411–458, 1983.

[5] M. Swerts, and R.Geluykens, *Prosody as a Marker of Information Flow in Spoken Discourse*, Language and Speech, vol. 37, pp. 21–45, 1994.

[6] E.Shriberg, A.Stolcke, D.Hakkani-T¨ur, *Prosody-Based Automatic Segmentation of Speech into Sentences and Topics*, Speech Communication, vol. 32, pp. 127–154. 2000.

[7] J. Hirschberg, *Communication and prosody: Functional aspects of prosody*. Speech Communication, *36*1, 31-43, 2002.

[8] M. E. V. Donzel, *Prosodic aspects of information structure in discourse*. 1999.

[9] J.Bachenko, E. Fitzpatrick, *computational grammar of discourse- neutral prosodic phrasing in English*. Computational linguistics, 155-170 . 1990.

[10] Wang, M. Q., and Hirschberg, J. Automatic classification of intonational phrase boundaries. *Computer Speech & Language*, *6*2), 175-196. 1992.

[11] M.Ostendorf and N.Veilleux, *A hierarchical stochastic model for automatic prediction of prosodic boundary location*. *Computational Linguistics*, 27-54. 1994.

[12] P.Taylor and A. W. Black, *Assigning phrase breaks from part-of-speech sequences*. 1998.

[13] A.Butcher, *Pause and syntactic structure*. In W. Dechert & M. Raupach (Eds.), Temporal variables in speech (pp. 86-90). Mouton. 1980.

[14] B.Zellner, *Pauses and the temporal structure of speech*. In *Zellner*, B. Pauses and the temporal structure of speech, in E. Keller (Ed.) Fundamentals of speech synthesis and speech recognition. *(pp. 41-62). Chichester: John Wiley.* (pp. 41-62). John Wiley. 1994.

[15] L. Olson, *Advanced Data Mining Techniques*, Springer, (1st edition February 1, 2008), page 138, ISBN 3-540-76916-1. 2008

[16] B.Megyesi and S.Gustafson-Capkova, *Production and perception of pauses and their linguistic context in read and spontaneous speech in Swedish*. In INTERSPEECH. 2002.

[17] E. MacNeal and S.Pinker, *The Language Instinct: How the Mind Creates Language*. 1995.

BIOGRAPHY

Israa Elhosiny
Principal Grammar Developer in the Arabic Computational Linguistics Center Bibliotheca Alexandrina.
PhD. Student in the Department of Phonetics and Linguistics. Faculty of Arts, Alexandria University (2016).
She obtained her MA, from Department of Phonetics and Linguistics 2015.
She obtained her BA, from Department of Phonetics and Linguistics 2004.
She obtained the UNL certificates; CLEA250, CLEA750, CUP250, and CUP500. She attended the X UNL School organized by the UNDL foundation and Bibliotheca Alexandrina (7-11 October 2012).
She has an experience in Python programming, and obtained the New Horizon certificates in Java programming (OOP, JDBC and JSF) (2017).
She is also Principal Grammar Developer of Arabic Computational Linguistics Center in Bibliotheca Alexandrina. She is working in the Universal Networking Language project in Library of Alexandria since 1-6-2005 till now.
She has an experience in Arabic language analysis and generation and text tokenization. She participated in building grammars using UNL for library information system (LIS) and Knowledge Extraction sYStem (Keys).
She has an experience in automatic Arabic diacritization. She participated in building grammars, and developing a diacritization evaluation tool.
She is a member in many scientific organizations: (1) Egyptian Society of Language Engineering, Cairo, (2) Universal Networking Language foundation, United Nations, Geneva, Switzerland.

Mervat Fashal
Professor in Department of Phonetics and Linguistics, Faculty of Arts, Alexandria University, Alexandria, Egypt.
She obtained her BA, from Department of Phonetics and Linguistics in 1979.
She obtained her MA, from Department of Phonetics and Linguistics in 1986.
She obtained her PhD in 1991.
She is a member in many scientific organizations: (1) Egyptian Society of Language Engineering, Cairo, (2) Egyptian Society of Phonetics, Ain Shams University, Cairo.
She has articles in the field of acoustics, prosody, experimental phonetics, speech perception and speech recognition and speaker identification in the field of forensic phonetics.

دراسة الوقفات الصوتية المدركة وعلاقتها بالمركبات النحوية

إسراء الحسيني[1]، ميرفت فشل[2]

قسم الصوتيات واللسانيات، كلية الآداب، جامعة الإسكندرية
[1]israa_elsayed@yahoo.com
[2]Mervat.fashal@alexu.edu.eg

ملخص—— إن التوليد الآلي للكلمات بصورة شبه طبيعية يتطلب التحكم في التركيب الزمني لإنتاج الكلام. لذلك فإن العلاقة بين حدود العبارات التطريزية من حيث الوقفات الصوتية والتركيب النحوي وجب دراستها في نوعين مختلفين من الكلام؛ المحادثة والكلام المتصل. تركز هذه الورقة البحثية على دراسة العلاقة بين كل من إنتاج المتكلم للكلام وإدراك المستمع له وعلاقتهما بحدود العبارات التركيبية في الجملة المنطوقة. من أجل وصف الترابط بين الوقفات الصوتية وحدود العبارات في كل من أساليب الكلام موضع الدراسة، اعتمد الباحث على حساب نسبة كل من الدقة والضبط **precision** والاسترجاع**recall.**

الكلمات المفتاحية: انتاج الكلام – فهم الكلام – تصنيف العبارات

Automatic Arabic Speaker Recognition Using Gaussian Mixture Model

, Amna Dheif*1, Aya Nabil*2, Rehab Arafat*3 Mervat Fashal*4

*Phonetics and linguistics Department, Faculty of Arts, Alexandria University
Alshatby, Alexandria, Egypt

1amnaadheif@gmail.com
2aya.nabil253@gmail.com
3rehab.arafat@gmail.com
4mervat.fashal@alexu.edu.eg

Abstract— **This article is presenting an experiment for Arabic speaker identification using the Gaussian Mixture Model (GMM). The speech signals of the speaker were described using the Mel-Frequency Cepstral Coefficients (MFCCs). Three experiments were held on a free open-source dataset, which consists of 20 words spoken by 50 male Arabic speakers. The fifteen speakers were used both in text-dependent and text-independent experiments. The first experiment yielded an accuracy of 84.48%. While the second experiment yielded an accuracy of 83.45%. The third experiment yielded an accuracy of 94.59%.**

Keywords: **speaker recognition (SR), Speaker Identification, Speaker verification, Mel-Frequency Cepstral Coefficients (MFCCs), GMM model, text-dependent recognition, text independent recognition**.

1 INTRODUCTION

Automatic Speech recognition is an interdisciplinary field that involves may methodologies that aim at the automatic recognition and identification of spoken language by means of applications. Automatic Speaker Recognition (ASR) is the process of recognizing the identity of persons based on their voices using an application. The process of speaker recognition can be categorized into **speaker identification** and **speaker verification** (ASV). [1] **Speaker verification** is the process of automatic verifying a person's identity from his voice. In other words, it decides if the speaker is whom he claims to be. This process can be used for verifying the identity of a person in applications like banking by telephone or voice mail. While the purpose of automatic speaker identification is to decide who the person is, without a priori identity speaker. It is the procedure of mapping of the speech signal of an unknown speaker to a database of known speakers.

There are some factors that must be taken in account that may lead to recognition and identifications errors. These factors could be misspoken or misread phrases. The verification process also may be affected be the emotional state of the speakers. Other environmental factors could be as the noise, the microphone placement or using different microphones for enrolment.

On the other hand, other factors could make the verification process a hard task. The variety of speaker's age, gender, mood and sickness may also affect the process and could make the recognition more complicated job and may lead to low rate of recognition. Also, accents will differ between speakers. The variety of the physiology of the organs of vocal tract of speakers will lead to variability in the speech signals.

The first speech recognition systems worked on isolated word or letter recognition and were speaker dependent. The next systems were to work on continuous speech, with a vocabulary of approximately a thousand words. Current state of the art systems is working on conversational or spontaneous speech in noisy and limited bandwidth domains. [2]

2 FEATURE EXTRACTION TECHNIQUES:

In the phase of extracting acoustic features form the speakers, the most commonly feature sets used in speech recognition are Mel frequency cepstral coefficients (MFCCs) and perceptual linear prediction (PLP) [2]. And the Mel frequency cepstral coefficients (MFCCs) are the most used method to represent the speech spectrum

in speaker recognition systems. The MFFCs will be used in this experiment to extract the features in the voice signal. MFCC focuses on series of calculation that uses Cepstrum with a nonlinear frequency axis following Mel scale. To obtain melcepstrum, the voice signal is windowed first using analysis window and then Discrete Fourier Transform is computed. The whole calculations will de discoed in the section 6.2.
Perceptual Linear Prediction (PLP) coefficients are improved spectral representation. It target is to model the psychoacoustics of hearing, by means of implementing three properties of the human auditory system: the nonlinear frequency response of the human ear; the critical bands in the cochlea; and the non-linear amplitude response. [2]

3 MACHINE LEARNING MODELS:

The classical machine learning models applied in the field of speech recognition somehow revolves around Hidden Markov Model (HMM) and Gaussian Mixture Model (GMM). [3] Hidden Markov models are generative models based on stochastic finite state networks. The acoustic model provides the likelihood of a set of acoustic vectors given a word sequence. Markov models are stochastic state machines with a finite set of N states. Given a pointer to the active state at time, the selection of the next state has a constant probability distribution. Thus, the sequence of states is a stationary stochastic process. It assumed that a given state depends on the existence of the previous states.
In the other hand, GMM can be thought of as a single state HMM [3]. GMM is a probabilistic model that supposes that all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. One can think of mixture models as generalizing k-means clustering to combine the information about the variant structure of the data as well as the centers of the latent Gaussians. GMM is preferably applied because it is more reliable than HMM, although HMM possess more accuracy. Also, GMM operates the output much faster than HMM.[3]
The GMM parameters can be thought as being analogous to a set of formant-like features. The component means correspond to the **formant locations**, the standard deviations to the **formant bandwidths** and the component energies to the **amplitudes**. The GMM parameters are extracted, then, they are ordered according to their frequency values. They are ordered as the component with the lowest mean is the first component and so on [2].
Usually, either of them is used independently or with a Deep Neural Network (DNN). However, sometimes they both can be used together in combination [3].

4 ARABIC EXISTING MODELS

(Tolba, 2011) Presented an approach for automatic speaker identification on Arabic speakers depended on Continuous Hidden Markov Models (CHMMs). The Mel-Frequency Cepstral Coefficients (MFCCs) were selected to describe the speech signal. The general Gaussian density distribution HMM is developed for the CHMM system. Ten Arabic speakers (Tolba, 2011)[4] Presented an approach for automatic speaker identification on Arabic speakers depended on Continuous Hidden Markov Models (CHMMs). The Mel-Frequency Cepstral Coefficients (MFCCs) were selected to describe the speech signal. The general Gaussian density distribution HMM is developed for the CHMM system. Ten Arabic speakers were participated in this experiment. The identification rate was 100% for text dependent experiments, but 80% for text-independent experiments.

Djemili, Bedda, & Bourouba (2007) [5] proposed a new hybrid approach that combines different models (statistical models and support vector machines). This technique gets benefit from the advantages of both models. Support vector machines (SVMs) are used in the training phase to divide the whole speakers’ space into small subsets of speakers within a hierarchical tree structure. During the testing phase a speech token is assigned to its corresponding group and evaluated using gaussian mixture models (GMMs). They showed that the proposed method can significantly improve the performance of text independent speaker identification task. The improvements are up to 50% reduction in identification error rate compared to the baseline statistical model.

Al-Ani, Mohammed, & Aljebory (2007)[6] presented a hybrid approach for Arabic speaker identification, where the wavelet transform and neural networks are used together to form the. Features are extracted by applying a discrete wavelet transform (DWT), while a neural network (NN) is used for formulating the system database and for handling the task of decision making. Evaluation of the system depends on false acceptance ratio (FAR) and false rejection ratio (FRR) performance. The participated speakers were 25 randomly aged male and female speakers.

The evaluation criteria parameters obtained are; FAR=14.5% and FRR=24.5%.

Naseem & Deriche (2006) [7] discussed a new system using the Dempster Shafer theory of evidence. They depended on a combined classifier based on the Dempster-Shafer theory outperforms the individual LPCC and MFCC classifiers. They developed a new approach for combining the results of the two different to improve the classification results of the LPCC and the MFCC. Results of their work are reported in this following table.

Aldhaheri & Al-Saadi (2004) [8] presented a new technique for text-independent speaker identification for noisy speech is presented. They based on finding the ratio of the singular values of the feature vectors of the unknown speaker and each of the N reference features stored in the database. The i**th** reference feature that gives the largest ratio is considered the feature of the unknown speaker.

They archived an accuracy rate of 99.5% for clean speech and 77.5% for noisy speech.

Table (1): Results of Naseem & Deriche's work (2006)

Method	Training	Testing	Recogntion
MFCC	5	3	85%
LPCC	5	3	83.3%
NNEF	5	3	90%

5 CHALLENGES OF AUTOMATED SPEAKER RECOGNITION

In order to perform the task of speaker recognition, a series of acoustic features are extracted from the speech signal, and then pattern recognition algorithms are trained on these features. Thus, the quality of speech signals is critical for the system performance. The are many factors that may be considered as challenges in that audio dataset. Some factors could be back to speaker characteristics. The speech signal differs among speakers due to the different anatomy and physiology of the vocal tracts of the speakers. Variations in the speech signal also could be back to the gender of the speakers, the ager of the speakers, speaking at different rates or using different accents of the language. A procedure to deal with this variability is through the construction of speaker-dependent speech recognition systems. But this requires a new system for each speaker to be constructed [9]. *However, Speaker-independent* systems have some flexibility, in the point that these systems are designed to recognize any speaker. Practically, speaker-dependent systems will make fewer errors than a speaker-independent system.

Other factors that could also lead to variations in the speech signals are environmental factors, in other words, the acoustic environment in which the speech is recorded, along with any transmission channel. This may have a significant impact on the accuracy of a speech signal, and accordingly on the speaker recognition process.so in case of recording outside the laboratories, it is very important to separate different acoustic signals or noises found in the environment [9]. Also, the position the microphone to the speaker and the movements of the speaker's head relative to the microphone may lead to some variability of the signal. So, all these factors must be taken in account during the process of speaker recognition.

6 PROPOSED MODEL:

This section explains the methodology followed in the current study. Two experiments were conducted, the first and second experiment's goal is to build a text-independent speaker recognition system for 50 speaker and the third experiment's goal is to build text-dependent speaker recognition system also for 50 speakers.

The three experiments relied on the same features extracted from the speech corpus and also used the same machine learning model which is the Gaussian Mixture model (GMM) to build the speakers models . In the following sections, the speech corpus used, the features used to represent each speaker identity, and each experiment will be discussed in great detail. Figure (١) shows how the model works in the three experiments.

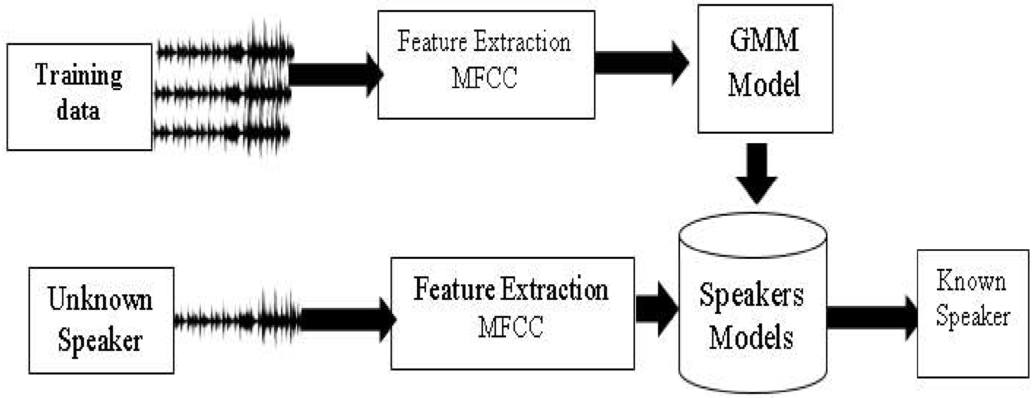


Figure (١) diagram for how the proposed model works in the three experiments

6.1 DATASET

This dataset used in this experiment is taken from "Arabic Corpus of Isolated Words". It is a free open-source dataset created by University of Stirling in Central Belt of Scotland. It can be downloaded from the official website of the university[1]. It consists of 9992 utterances, of 20 words spoken by 50 native male Arabic speakers. The data was recorded with a 44100 Hz sampling rate in .wav format files. Each file contains only one voice. Table (1) shows the twenty words of the Arabic Corpus of Isolated Words along with their transcription.

Table 2: The twenty words of the Arabic Corpus of Isolated Words Along with their transcription

Word	Transcription
صِفْرْ	s̲e.fer
وَاحِدْ	wa.ħid
إِثْنَان	ʔiθ.na:n
ثَلَاثَة,	θa.la:.θah
أَربَعَة	ʔar.ba.ʕah
خَمْسَة	xam.sah
سِتَّة	sit.ta

[1] http://www.cs.stir.ac.uk/~lss/arabic/[11]

سَبْعَةْ	sab.ʕah
ثَمَانِيَةْ	θa.maːniy.yah
تِسْعَةْ	tis.ʕah
التَّنْشِيطْ	ʔat.tan.ʃiːt
التَّحْوِيلْ	ʔat.taħ.wiːl
الرَّصِيد	ʔar.ra.siːd
التَّسْدِيدْ	ʔat.tas.diːd
نَعَمْ	na.ʕam
لَا	La:
التَّمْوِيل	ʔat.tam.wiːl
الْبَيَانَاتْ	ʔal.ba.yaː.na:t
الْحِسَابْ	ʔal.ħi.saːb
إِنْهَاءْ	ʔin.haʔ

6.2 FEATURE EXTRACTION

The general methodology of speaker recognition involves extracting discriminatory features from the audio data and feeding them to a training model that would use these features to segregate speakers from each other. The so-called *Mel-Frequency Cepstral Coefficients* (MFCC) are the most used in speech and speaker recognition.

The MFCC model imitates human perception processes, and therefore is considered the best model for speech/ and speaker identification. MFCC model have main stages a speech signal has to go through. The first stage is the pre-emphasis stage where high frequencies are amplified, and the noise is cut off from the signal. The Pre emphasis filter used in all the experiments was .97 filter. The second stage is framing where the speech signal is sliced into (overlapping) short frames. The size of frame in the current study was 25ms and the step between each two frames is 10ms. The third stage is windowing to keep the continuity of the first and the last points in the frame. The windowing function used in the current study is hamming window. The fourth stage is applying a fast Fourier transform is applied on each frame to identify the main frequencies present in each frame. Subsequently the filter banks are computed using the Mel scale which tells which frequencies humans discern as the difference between two closely spaced frequencies are not discerned as frequencies increase. Once the filter bank energies are computed, we take the logarithm of them. This imitates how humans perceive loudness of a signal. The last step is to apply the Discrete Cosine Transform (DCT) to the filter banks and eventually a multidimensional feature vector for every speech signal is yielded. In the current study delta coefficients are appended to the MFCC coefficients. The Following figure number (2) depicts the MFCC model steps.

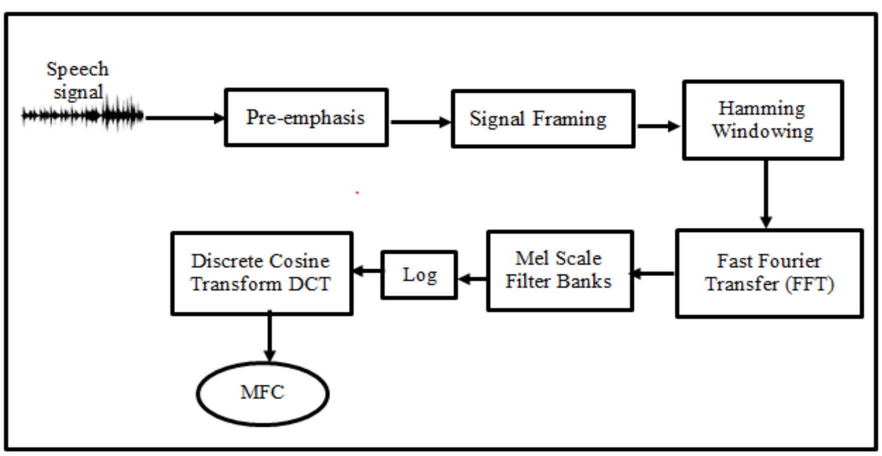


Figure (2): MFCCs Model Steps

6.3 EXPERIMENTS:

The experiments proposed in this paper were based on with Gaussian Mixture Models (GMMs) in order to identify Arabic speakers automatically from their voices using the **sklearn.mixture** python package.

6.3.1 EXPERIMENT ONE:

Experiment one addressed the goal of building a text-independent speaker recognition system using the early proposed Gaussian Mixture Model (GMM). The previously discussed features where fed to The GMM where it built a model for each speaker. The first thirteen (13) words of the Arabic Corpus of Isolated Words of each speaker represented the training material for the GMM model. The model was trained on the ten trials of each of the thirteen words for each speaker, thus, the training data consisted of 6,500 speech signals. The other 7 words remaining in the corpus represented the testing material for the GMM model. The model was tested on the ten trials of each of the seven words for each speaker, thus, the testing data consisted of 3,500 speech signals.

6.3.2 EXPERIMENT TWO:

Experiment two was also conducted to build a text-independent speaker recognition system using GMM. The last thirteen (13) words of the Arabic Corpus of Isolated Words of each speaker were alternatively used as the training material in this experiment, while the first 7 words in the speech corpus represented the testing material for the model. The concept behind the second experiment is to find out if the training data is different, how the accuracy of the model shall differ.

6.3.3 EXPERIMENT THREE:

Experiment three alternatively addressed the goal of building a text-dependent speaker recognition system. The first seven trials of all the 20 words of the speech corpus of each speaker represented the training material for the GMM model, thus, the training data consisted of 7,000 speech signals. The other last three trials of the 20 words remaining in the corpus represented the testing material, thus, the testing data consisted of 3,000 speech signals. The results of the text dependent speaker recognition are assumed to be higher than the text independent two experiments.

7 RESULTS

The first experiment yielded an accuracy of 84.48%. Figure (3) depicts the number of times each testing word caused the system to misidentify the speaker.

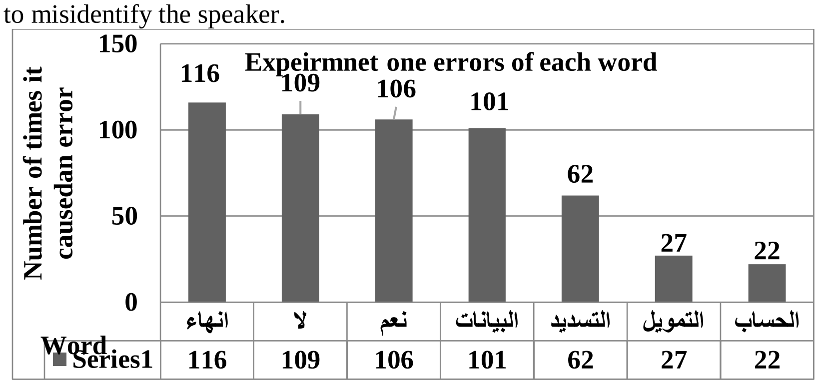


Figure (3) Experiment one errors of each word

The second experiment yielded an accuracy of 83.45%. Figure (4) depicts the number of times each testing word caused the system to misidentify the speaker.

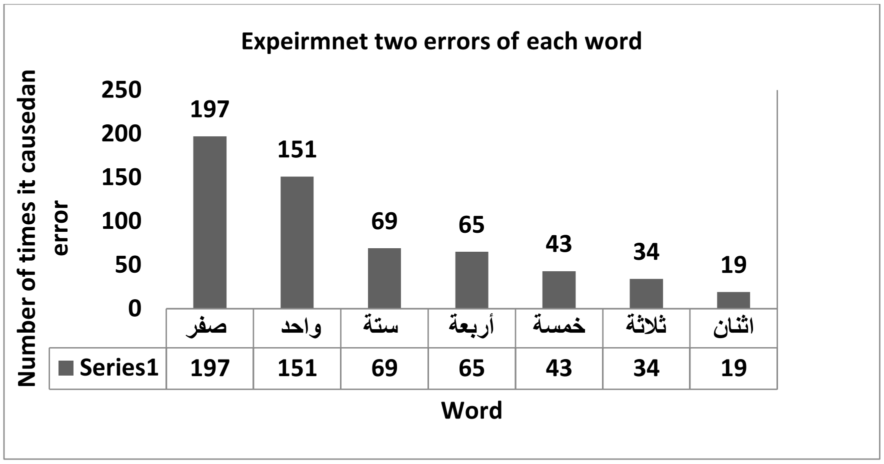


Figure (4) Experiment two errors of each word

The third experiment yielded an accuracy 94.59%. Figure (5) depicts the number of times each testing word caused the system to misidentify the speaker

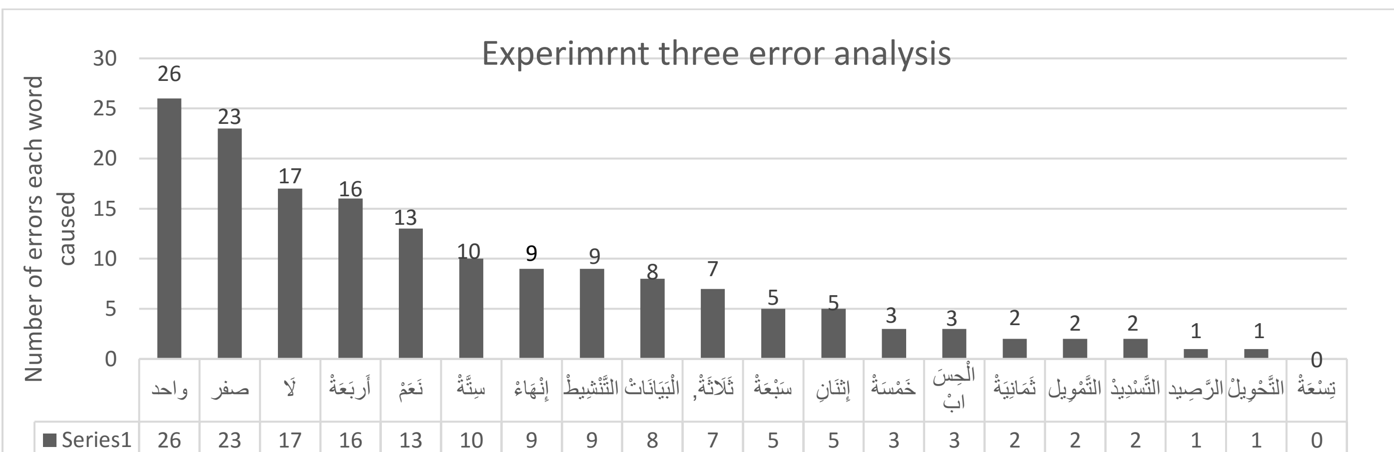


Figure (5) Experiment three errors of each word

8 DISCUSSION AND CONCLUSION

The thee experiments yielded good results regarding the task of speaker identification. Of course, the text independent speaker identification task is more challenging than the text dependent speaker one. This can be seen in the drastic ten percent drop of the accuracy from 94.59% in the text dependent task to 84% which represents the average accuracy of the text independent two tasks.

Changing the training data in the text independent two experiments had a small impact on the accuracy as the accuracy dropped by 1.03% in the second experiment.

Running an error analysis for all the experiments, it was found that, words that were short in length are the reason why the model behaved poorly. Table 2 illustrates the average length of each word in milliseconds and the average length of the sonorant sounds in each word in milliseconds. As shown the short words "لَا, "إِنْهَاءْ","نَعَمْ" and "الْحِسَابْ" are the ones that caused the highest errors in experiment one. The word "التَّسْدِيدْ" has a very short sonorant sounds length which made the model performance poor. The two words "التَّموِيل" and "الْبَيَانَاتْ" are both long in length and have the longest sonorant sounds yet they caused a high error rate.

In experiment two, the two words "صِفْرْ", "سِتَّةْ", "خَمْسَةْ"and "وَاحِدْ" caused the highest errors as they have short sonorant sounds. The least two words the caused errors are the longest two words in the testing set which are "ثَلَاثَةْ" and "إِثنَانِ". The word "أَربَعَةْ" has a long length and long sonorant sounds yet it caused a high error rate.

In experiment three, the short words are the ones that caused most of the errors except for "تِسْعَةْ" and "خَمْسَةْ". On the other hand, the long words caused the least errors except for "الْبَيَانَاتْ" and "التَّنْشِيطْ".

Table 3: the average length of each word and the average length of the sonorant sounds in each word in milliseconds

Word	Average Length in MSEC	Average Sonorant sounds length in MSEC
لَا	27	27
إِنْهَاءْ	28	28
تِسْعَةْ	35	20
خَمْسَةْ	40	20
ثَلَاثَةْ	42	32
صِفْرْ	45	20
أَربَعَةْ	46	46
سَبْعَةْ	46	30
سِتَّةْ	47	22
وَاحِدْ	54	38
إِثنَانِ	54	40
ثَمَانِيَةْ	55	45
نَعَمْ	65	65
الْحِسَابْ	65	40
التَّموِيل	66	65
التَّسْدِيدْ	71	30
التَّحْوِيلْ	72	35
التَّنْشِيطْ	73	35
الرَّصِيد	73	60
الْبَيَانَاتْ	78	78

To conclude, the three experiments show promising results for speaker identification. The researchers suggest that the error rate could be decreased if both the training data and the testing data are longer in length.

REFERENCES

[1] Joseph P. Campbell, JR. Speaker Recognition: A Tutorial. PROCEEDINGS OF THE IEEE, VOL. 85, NO. 9, SEPTEMBER 1997.

[2] A Gaussian Mixture Model Spectral Representation for Speech Recognition Matthew Nicholas Stuttle Hughes Hall and Cambridge University Engineering Department. 2003.

[3] Speaker- Identification over Call Records. Atul Anand, Bharti Parmar, Ankit Kumar. ITM university, India.

[4] Tolba, H. (2011). A high-performance text-independent speaker identification of Arabic speakers using a CHMM based approach. *Alexandria Engineering Journal*, 43–47.

[5] Djemili, R., Bedda, M. & Bourouba, H. (2007). A Hubrid GMM/SVM System for Text Independnent Speaker Identification. International Journal of Electronic, Computer, Enteritic, Electronic and Communication Engineering, 713-719.
[6] Al-Ani, 1. S.,Mohammed, T.S., & Aljebory, k. M. (2007). Speakers Identification: A Hybrid Approach Using Neural Networks and Wavelet Transform. *Journal of Computer Science*, 304-309.
[7] Naseem, I., & Deriche, M. (2006). A new Algorithm for Speaker Idenification Using the Dempster-Shafer Theory Of Evidence. King Fahd University of Petroleum and Minerals. Saudi Arabia.
[8] Aldhaheri, R. W., & Al-Saadi, a. F. (2004). Text-Independent Speaker Identification in Noisy Environment Using Singular Value Decomposition. King Abdulaziz University. Saudi Arabi.
[9] The Handbook of Computational Linguistics and Natural Language Processing. Alexander Clark, Chris Fox, and

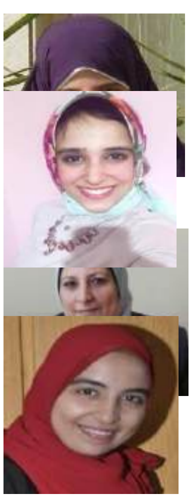
Shalom Lappin

[10] International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-2 Number-4 Issue-7 December-2012 119. Automatic Speaker Recognition System. Parul1, R. B. Dubey2
ECE, Hindu College of Engineering, Sonepat, India1,2
[11]Computing Science and Mathematics, Faculty of Natural Sciences, University of Stirling, Scotland

BIOGRAPHY

Amna Helmy: Assistant teacher at Phonetics and Linguistics department, Faculty of Arts, Alexandria University .She got a Bachelor of Arts in 2013 (Phonetics and Linguistics department – Faculty of Arts - Alexandria university – Alexandria - Egypt).She has graduated from Information technology institute (ITI) in 2014 intake 34.She worked as a demonstrator at Phonetics and Linguistics department , Faculty of Arts ,Alexandria university form 2013 till 2019 .She has discussed her Master thesis in 2019 about semantic role labeling. She had participated in three published papers.

Aya Nabil: Assistant teacher at Phonetics and Linguistics department, Faculty of Arts, Alexandria University She got a Bachelor of Arts in 2012 (Phonetics and Linguistics department – Faculty of Arts - Alexandria university – Alexandria - Egypt).She has graduated from Information technology institute (ITI) in 2013 intake 33.She works as a demonstrator at Phonetics and Linguistics department , Faculty of Arts ,Alexandria university form 2013 till now .She has discussed her Master thesis in 13-10-2018.



Rehab Arafat: Phonetics Laboratory' specialist at Phonetics and Linguistics department, Faculty of Arts, Alexandria University. She got a Bachelor of Arts in 2008 (Phonetics and Linguistics department – Faculty of Arts - Alexandria university – Alexandria - Egypt). She worked in UNL Center in Bibliotheca Alexandrina from 2008-2012. She has discussed her Master thesis in 18-3-2015.

Mervat Fashal: Professor of phonetics science, and in psycho-acoustic field, Department of phonetics and linguistics, Faculty of arts, Alexandria University. She had studied Ph.D. in Germany. She has articles in the filed of acoustics, prosody, experimental phonetics, speech perception and speech recognition and speaker identification in the field of forensic phonetics.

TRANSLATED ABSTRACT

التعرف الآلى على المتحدث باستخدام نموذج جاوس المختلط

آمنة ضيف[1]، آية نبيل[2] ، رحاب عرفات[3] ، مرفت فشل[4]

قسم اللغويات و الصوتيات، كلية الآداب، جامعة الاسكندرية

[1]amnaadheif@gmail.com

[2]aya.nabil253@gmail.com

[3]rehab.arafat@gmail.com

[4]mervat.fashal@alexu.edu.eg

ملخص:

تقدم هذه المقالة تجربة للتعرف الآلي على متحدثي اللغة العربية باستخدام نموذج جاوس الإحصائي . .(GMM) تم توصيف الموجات الصوتية للمتحدث باستخدام معاملات ميل التردد (MFCC). أجريت التجربة على مدونة صوتية مفتوحة المصدر، تتكون من ٢٠ كلمة يتحدث بها خمسون متحدثًا باللغة العربية. تم استخدام المتحدثين الخمسين في كل من التجارب المعتمدة على النصوص والتجارب غير المعتمدة عن النصوص. أسفرت التجربة الأولى عن نسبة دقة في التعرف الآلي بلغت ٨٤,٤٨ ٪. بينما أسفرت التجربة الثانية عن نسبة تعرف بلغت ٨٣,٤٥ ٪. والتجربة الثالثة بلغت نسبة التعرف فيها ٩٤,٥٩%.

الكلمات المفتاحية: التعرف الآلي على المتكلم (SR) ، التحديد الآلي للمتكلم، التحقق الآلي من المتكلم، معاملات ميل التردد، نموذج جاوس الإحصائي، التعرف الآلي المعتمد على النص، التعرف الآلي غير المعتمد على النص.

End-to-End Arabic Speech Recognition: A Review

Abdelaziz A. Abdelhamid[*, **1], Hamzah A. Alsayadi[*2], Islam Hegazy [*3], Zaki T. Fayed[*4]

[*] *Computer Science Department, Faculty of Computer and information Sciences, Ain Shams University, Cairo, Egypt*

[1]Abdelaziz@cis.asu.edu.eg

[2]hamzah.sayadi@cis.asu.edu.eg

[3]islheg@cis.asu.edu.eg

[4]zaki.taha@cis.asu.edu.eg

[**] *Computer Science Department, College of Computing and Information Technology, Shaqra University, Saudi Arabia*

Abstract: — **Automatic speech recognition (ASR) is a crucial field of science due to its massive applications that can be developed to help humans to improve their daily life tasks. Despite its long history, ASR remains an active and interesting research field in general and on Arabic language in particular. Arabic is one of the most widely spoken languages. However, current research is still limited on it due to its high variations and complex morphology. Therefore, this paper highlights the most recent techniques and key milestones of Arabic speech recognition to guide researchers who are interested in working on the Arabic language. There are many machine learning techniques applied in building ASR systems. For long time, hidden Markov models (HMMs)-Gaussian mixed models (GMMs) were standing as the best frameworks for ASR. However, in last decade, hybrid HMM-deep neural network (DNN) models and end-to-end deep learning models have been emerged as a breakthrough in improving the performance of ASR. End-to-end deep learning is distinguished as the most recent methodology in the field and represents the main focus of this review. Therefore, the proposed review discusses the most recent achievements of research on Arabic speech from the end-to-end methodology perspective. In addition, the currently available services and toolkits necessary for building end-to-end models are explained.**

Keywords: **Automatic Speech Recognition (ASR), End-to-End; Deep learning; LSTM; CTC; RNN; Attention-Based; HMM.**

1 Introduction

Arabic language is one of the five most important languages in the world [1]. It is the official language about 25 countries [1]. Arabic is one of the most widely spoken languages around the world with an estimated number of over 313 million speakers and with 270 million as a second language speaker Arabic is ranked as the forth after Mandarin, Spanish and English, it contain about 12 million words [2]. Moreover, it is the language of the Islamic holy book “Quran” with 1.8 billion Muslims around the world in 2015 and projected to increase to 3 billion in 2060 [3]. There have been relatively little speech recognition researches on Arabic compared to other languages [4]. Arabic is a morphologically rich language. Prefixes and suffixes, affixes for short, augment word stems to form words [5].

Arabic language is one of the oldest languages in the world and is a Semitic language and has high variation [6]. There are three types in Arabic language: 1) Classical Arabic is the language of Quran and Hadith (the religion language) and language of the old Arabic poetry. 2) Modern Standard Arabic (MSA) is based on classical Arabic without some features like syntax structure and diacritics. MSA used in formal communication, news, modern books, newspapers, and modern books. 3) Dialectal Arabic has multiple regional forms and is used for daily spoken communication in non-formal settings. With the advent of social media, dialectal Arabic is also written. Each type of Arabic language has different grammatical, lexical, and morphological standard. This makes hardness of developing Arabic NLP applications to process data from different varieties [7].

There are several type of dialectal Arabic, each type differs from one country to another. This problem is even more pronounced for dialectal Arabic due to the following reasons:

1) Additional prefixes, and sometimes suffixes, are informally introduced during the everyday use of language. 2) The amount of text data available for dialectal Arabic is usually much smaller than that for MSA, and hence it is not clear how to increase the vocabulary size to reduce out-of-vocabulary (OOV). 3)

Even if vocabulary is increased using some means, the sparse text resources will lead to poor estimates of the language model probabilities, and hence may hurt performance on a different front [8].

Automatic speech recognition (ASR) is the automatic way to transcribe the speech into text. It is used to make machines understand the human voice. Over the last decades, ASR technologies plays an important role in many areas such as education, personal computers, robotics, mobile phones, dictation, military, health, security systems, …etc. ASR is important because speech is the simplest ways of communication among people. In addition, ASR systems are under active development and adopted by a wide range of applications due to their functionality and simplicity. For instance, it is used in customer care applications where users can interact with a voice-enabled service instead of human interactions. This helps with serving higher number of customers and reduce lengthy service queues. Another example is using ASR in smart homes to control heating, lighting, and other appliances. In this context, Automatic Digit/Command Recognition is considered as one of the most challenging domains in ASR.

The growing importance of Digit/Command recognition through the increasing use of applications that help human-machine interaction by natural languages such as command systems via pronounced digits [9, 10].

There are many systems presented to show the importance of ASR. The most popular systems are: Microsoft SAPI, Dragon Naturally Speaking, and IBM via voice. Open source speech recognition systems are available too, such as domain speech-to-text system and The SPHINX-II [11, 12]. Most researchers developed ASR system based on Hidden Markov Models (HMMs) [7]. HMM is a statistical model where the system being modeled is assumed to be a Markov process with unknown parameters, and the challenge is to determine the hidden parameters, from the observable parameters, based on this assumption. The extracted model parameters can then be used to perform further analysis, for example for pattern recognition applications. Its extension into foreign languages (English is the standard) represent a real research challenge area [13].

Arabic Automatic Speech Recognition (ASR) is a challenging task because of the morphology, data sparseness, and lexical variety of the language [14]. Although there are too many people who speak Arabic, there is little research in Arabic compared to other languages [15, 16, 17].

The first system on Arabic ASR is used to recognize the modern standard Arabic (MSA). The most difficult problems in building highly accurate ASRs system for Arabic are the morphological complexity, predominance of non diacritized text material, and the enormous dialectal variety [16].

D. Vergyri et al. examine the usage of morphology-based language model at different phase in a speech recognition system for conversational Arabic [16]. K. Kirchhoff et al. [17] examine the recognition of dialectal Arabic and study the discrepancies between formal and dialectal Arabic in the speech recognition point of view. D. Vergyri et al [15] examine the automatic diacritizing Arabic text for use in acoustic model training for ASR. Reducing the entry barrier to build robust Automatic Speech Recognition (ASR) for Arabic has been a research concern over the past decade [18]–[21].

Different studies have been investigated in the literature to propose recognition systems using different approaches [8, 22, 23]. However, compared to other languages such as English, the number of research papers in Arabic language is limited [7]. In this review, some studies concerning ASR systems for the Arabic language will be discussed.

Large-vocabulary automatic speech recognition (ASR) for conversational Arabic poses several challenges for the speech research community. Most acoustic training features for Arabic ASR is transcribed in the Arabic nondiacritized form, which does not include short vowels and other diacritics that reflect differences in pronunciation, such as the fattaha, kassra, etc. In particular, the Arabic diacritized text (standard script) form (e.g. broadcast news corpora) is much easier to recognize than the Arabic nondiacritized form.

The nondiacritized texts have constraint for features of recognizer training model. This constraint is cause problems for both language and acoustic modeling. First, if the identity and location of the short vowels in the signal is not known, hence, this makes it difficult to train accurate acoustic models. Second, the lack of diacritics leads to a great set of linguistic contexts for a particular word paradigm; language models trained the nondiacritized features may be less predictive than diacritized trained. Both of these factors may lead to a loss in recognition accuracy. The work in [24] shows the significant of both word error rate and language model perplexity is increased when the Arabic text does not contain vowel information.

There are software applications developed for the automatic diacritization of Arabic by some companies (Sakhr, Apptek, RDI) [25]. However, these products use to predict diacritics using only possible morphological analyses, text-based information, and such as the syntactic context of words. In the context of diacritization for speech recognition, by contrast, acoustic data is available that can be used as an additional knowledge source [25].

Traditional automatic speech recognition (ASR) systems used a modular design. In these systems, different model are trained for pronunciation lexicon, acoustic modeling, and language modeling separately. In contrast, in end-to-end (E2E) approach, all these models are trained to convert the features of acoustic to text transcriptions directly, potentially optimizing all parts for the end task [26]. The goal of End-to-end ASR is to make is to simplify training the above module-based components into a single-network architecture within a deep learning technique, in order to fix these issues. End-to-end ASR approach typically depends only on both acoustic and language data without linguistic knowledge, and train the model with a single algorithm [27].

Unfortunately, they are also less interpretable: identifying what different parts do and what properties they capture is less straightforward. It is a common problem in many neural network models besides E2E ASR. Therefore, a line of work is concerned with deciphering the information captured by learned representations in neural models that are trained on some downstream task [28].

Previous work analyzed different neural representations and various properties, such as evaluating how phonetic information is captured in neural acoustic models [29, 30]. However, E2E ASR models are still relatively under-explored.

Therefore, the end-to-end makes it easy to develop ASR systems without expert knowledge. The end-to-end ASR architecture have several types such as recurrent neural network (RNN) transducer [31], attention-based encoder decoder [32], connectionist temporal classification (CTC) [33], and their hybrid models [34, 35].

Recently, the use of external language models has shown significant improvement of accuracy in neural machine translation [36] and end-to-end ASR [37, 38]. This approach is called shallow fusion, where the decoder network is combined with an external language model in log probability domain for decoding [27].

This paper is organized as follows: In Section 2, we briefly give background of the end-to-end techniques. In Sections 3, we try to describe the related work to Arabic ASR using some techniques such as HMM-based, neural network, recurrent neural network, deep learning and etc. In Section 4 we over view the work in Arabic ASR that used the end-to-end approach. Section 5 mentioned the End-to-End Arabic ASR services like API services and toolkits. Finally Section 6 presents the conclusion and the future directions.

2 BACKGROUND

In this section we present detailed information about end-to-end approach and the techniques that are used in end-to-end approaches.

G. End-to-end speech recognition approach

In recent years, with the advancement of deep learning, end-to-end solutions have emerged in many areas. A salient example is the wide application of deep convolutional neural networks (CNNs) to the classification task [39]. End-to-end represents a mapping between the sequence of input acoustic features and sequence of grapheme or words. In Conventional ASR, the language model, pronunciation, and trained acoustic components are trained separately as shown Fig. 6 [39].

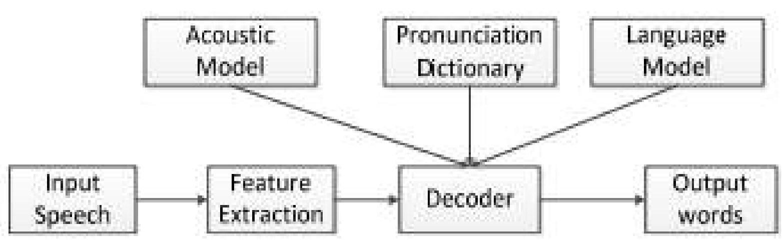


Figure 1: Conventional ASR Structure [39]

While end-to-end speech recognition greatly simplifies the complexity of traditional speech recognition, the models are not need to train separately, the pronunciation information or language can automatically learn as shown Fig. 2 [39].

Figure 2: End-to-end ASR Structure [39]

For more details, most end-to-end speech recognition models involve the following phases: 1) encoder, which realizes a mapping between the sequence of speech input and the sequence of feature; 2) aligner, which align the feature sequence to language; 3) decoder, which decodes the final identification result as in Fig. 3 [40].

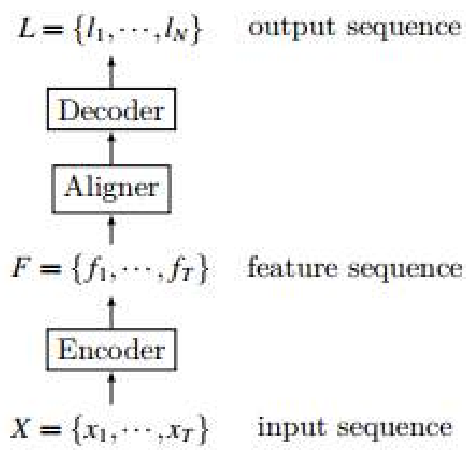


Figure 3: Function structure of end-to-end model [39]

End-to-end speech recognition is mainly based on deep recurrent neural network. A deep recurrent neural network alone is not enough for a speech recognition system. The first attempts used CTC, but it is incapable of learning the language and needs language model to clean up common mistakes. Instead of CTC, attention-based models were tested and they proved to be outperforming previous models, due to the ability of learning all components of a speech recognizer. Both methods still have some benefits over one another and recent works have started using a hybrid CTC/attention based architecture.

Using attention-based encoder–decoder models, the end-to-end idea can be naturally ported to speech recognition [38]. In principle, building an ASR system involves learning a mapping from speech feature vectors to a transcript (e.g., words, phones, characters, etc.), both of which that are in sequence, so no reordering is expected to take place. If we can learn such a mapping directly, all the components are optimized under a unified objective, which can enhance the final recognition performance obliterating the need for separate acoustic and language models [41].

H. Deep Learning

Building intelligent machines has fascinated humanity for centuries [42]. The field of Artificial Intelligence (AI), however, started to develop relatively recently, when programmable digital computers were conceived. The rise of deep learning [43] has recently contributed to renew the interest in AI and has allowed current technology to achieve higher levels of artificial intelligence.

Deep learning is actually a very general machine-learning paradigm that follows a compositionality principle to represent the world around us efficiently. Current deep learning implementation exploits deep neural networks, which are properly trained to progressively discover complex representations starting from simpler

ones. This principle can be applied in several practical problems, including the problem of recognizing human speech [43].
The deep learning paradigm is currently implemented with Deep Neural Networks (DNNs), that are Artificial Neural Networks (ANNs) based on several hidden layers between input and output. Each layer learns higher-level features that are later processed by the following layer [42].

I. Recurrent neural network (RNN)

The recurrent neural network was first developed in the 1980s [44]. RNN is a type of deep learning model that works best for handling sequential information. RNN assumes that all inputs, one or more hidden layers, and outputs are dependent on each other, unlike the traditional neural network. It keeps a memory of previous outputs and passes those as inputs from one-step of the network to the next as shown in Fig. 4. This way the network can have a deeper understanding of the application [45].

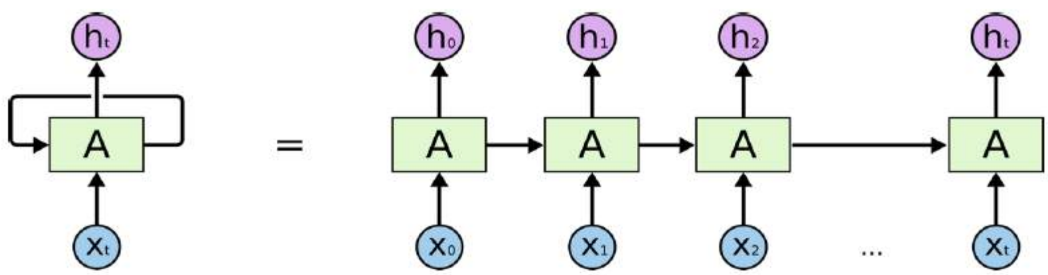


Figure 4: An unrolled RNN [44]

There are many applications of Recurrent Neural Networks such as Machine Translation, Robot control, Time series prediction, Speech recognition, Speech synthesis, Music composition, Handwriting recognition, Human action recognition, and etc. [44].
The above figure shows a chunk of neural network (A), that takes (Xt) and previous output as inputs and outputs a value (ht). The recurrence allows the network to pass information from one-step to the next [43]. This is the basic workflow of a RNN, but it is often used with bidirectional to get more accurate results.
In the training of RNN stage, the backpropagation algorithm used to calculate gradients and adjust weights in artificial neural network (ANN). It also applied in adjusting and modifying the weights after the updating of the feedback process [44]. The RNN uses the backpropagation through time (BPTT) method, this method utilizes working-backward manner to swap the weights of each unit according to the total output error [44].

J. Long Short-Term Memory (LSTM)

Standard RNN architectures have a problem with multiple hidden layers. When passing information from one hidden layer to another, the information might get lost, if there are many layers. LSTMs are a special case of RNN, can able to use the long-term dependencies and save information long time as a default [46].
The LSTM model is organized as chain structure form. However, the repeating module has a different structure. The standard RNN has a single neural network, while LSTM uses four interacting layers with a unique communication link as shown in Fig. 5 [46].

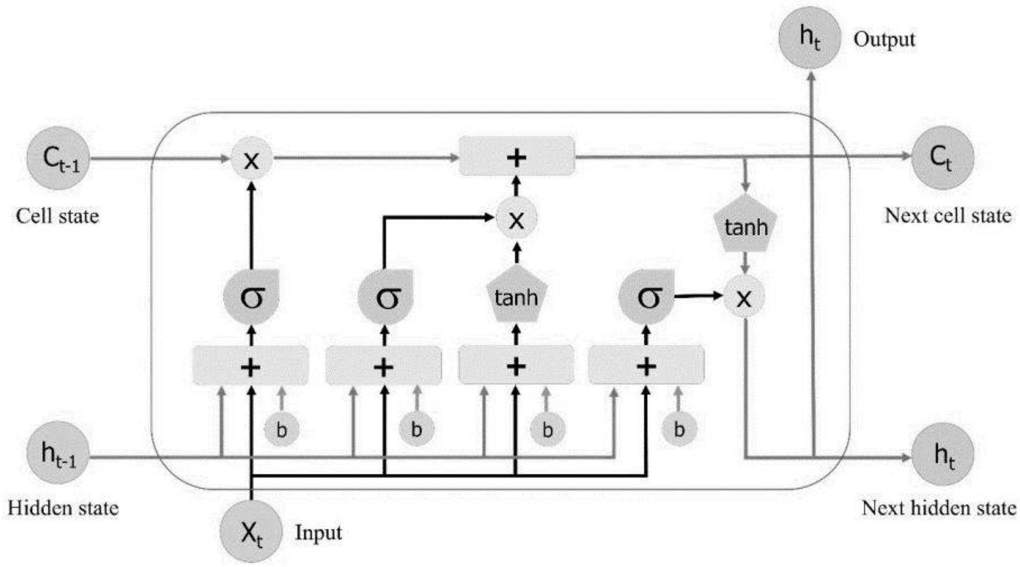


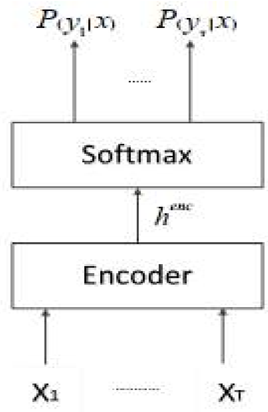
Figure 5: The structure of the Long Short-Term Memory (LSTM) neural network [44]

LSTM has many Applications like Robot control, Time series prediction, Speech recognition, Handwriting recognition, Human action recognition, Sign Language Translation, Protein Homology, Detection, and etc.
LSTM handles this kind of situation and enables RNN to preserve memory throughout the whole learning process RNN by adding additional interactions per module (or cell) [44].
LSTM architecture consists of memory blocks, which are all recurrently connected to each other. Each memory block contains at least one self-connected memory cell [45]. Any next state receives two states, the cell state, which allows the data to flow forward without change these data and the hidden state. LSTMs include sigmoid gates, which used to add or remove data from the cell state, a gate contains weights like the layer or series of matrix operations [44]. It also uses gates to control the memorizing process as the following: first identify and remove the information is not required from the cell, then storing information from the new input in the cell state as well as to update the cell state. Finally, the output values is based on the output cell state [44].

K. Connectionist temporal classification (CTC)

People talk with very different rates of speed which makes training an ASR system a lot more difficult. That is why the alignment between characters in the transcript and audio is always unknown [40]. One way of solving this problem is to manually align all characters to their location in the audio. The major downside is that it's very time consuming when dealing with large datasets. Another option is to use connectionist temporal classification (CTC) which has become a very popular among RNNs [47].

Graves et al [48] proposed CTC to allow for training an acoustic model without the need for frame-level alignments between the acoustics and the transcripts. The acoustic model training using CTC as



the loss function is an end-to-end

Figure 6: CTC structure [39]

training, which does not need to align the data in advance, but only needs an input sequence and an output sequence to be trained. Structure as shown in Fig. 6 [39].
CTC has applications like speech, handwriting recognition, and Scene Text Recognition. CTC is a type of neural network output and associated scoring function. It is used with RNNs to handle sequential problems. CTC sums over the probability of all possible alignments between the input and the output [49]. In this way, we do not need to align the data one by one, the CTC represents the spike in the whole speech as blank (which has no predicted value in any frame) [39]. Assuming that an input has a length greater than the actual word's length, one option for solving the problem is to collapse all repeating characters as shown in Fig. 7.

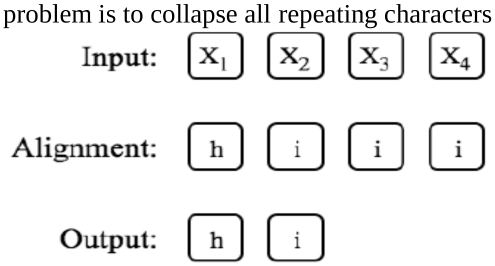


Figure 7: Merging repeated characters

L. Attention model

Attention Identifies encoded frames that are relevant to producing current output. The first use of attention-based Encoder-Decoder Models in the context of neural machine translation. The purpose of this model was to solve the problem of RNN-based sequence to sequence (Seq2Seq) model. Seq2Seq is represent end-to-end machine translation that need to encode input texts into a fixed-length vector and decode the target result [39]. The attention-based Encoder-Decoder is used to encode the input data into a sequence of vectors, the decoder based the attention mechanism makes each vector in this sequence has different weights (different length). Then the output is the sum of sequence's weights and the output of previous sequences [39]. Because the signal in speech have a different length, the attention-based Encoder-Decoder is suitable for speech recognition tasks for the following reasons: 1) speech recognition, like translation task, need a sequence-to-sequence approach to recognize the output sequence from the input sequence, 2) the encoder–decoder mechanism based attention method can implicitly find the soft alignment between input and output sequences, which solves a signal length problem for speech recognition, 3) Encoding result is not depend on a single fixed-length vector, the model can still have a good effect on long input sequence, so it is also possible for model to deal with different length in speech input [40].
Attention-based end-to-end model includes three parts: encoder, aligner, and decoder. In particular, its aligner part uses attention mechanism as shown in Fig. 8.

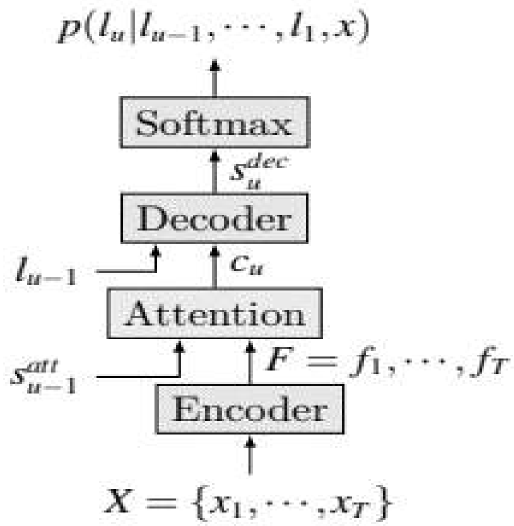


Figure 8: Attention model structure [39, 40]

3 Related Work

The aim of speech recognition is to enable machines to accept sounds and act based on it. Automatic speech recognition is the ability for a machine to recognize "receive and interpret" the speech and convert it into readable form or text and performing an action based on the instructions defined by the human [50].

We will focus on Arabic ASR using HMM, neural network-based, deep learning, recurrent neural network, and hybrid approaches. We will discuss the literature in terms of Arabic ASR classification, stages, and techniques.

Yu and Deng [43] presented an architecture for Arabic ASR, consists of four stages: 1) preprocessing stage. 2) Feature extraction stage. 3) Pronunciation dictionary, Language model, and Acoustic model (decoding). 4) Post-processing results were the best hypothesis is produced. The mechanism of work as following: First, speech signals of utterance used as input in the preprocessing stage. Then, the output is processed speech signals used as input to feature extraction stage where the features are represent by vectors as output. After that, the vectors use as input in the next stage, the decoding stage. The decoding stage is working along with a pronunciation dictionary. Finally, the n-best hypothesis, the output of the pronunciation dictionary stage, is used as input to post-processing. As a result, the best hypothesis is produced from this work operation.

Turab, Khatatneh, and Odeh in [51] proposed system for the phoneme recognition as ASR system. Many techniques are used in that paper: firstly Gaussian Low Pass filtering algorithm along with the neural network in the pre-processing stage to have an enhancement the result. Then, catching a signal, sampling, quantization and setting energy is done in the phoneme recognition sage. After that, a neural network is used to improve the results is achieved. When applying the Gaussian Low Pass filter in voice signals, the enhanced impact in results hence, due to noise reduction. After that, in the training phase, the neural network has been used to train the system in order to recognize the speech signal.

Ahmed and Ghabayen in [4] suggested three approached to enhance the AASR. The first approach is the punctuation modeling, where they used a decision tree with variant pronunciation generation. Then, a hybrid approach is proposed to acclimate the native acoustic model with another native acoustic model. Finally, processed text was used to enhance and improve the language model. In the results, the word error rate (WER) is reduced by 1%, 1.2% and 1.9% for the pronunciation model, the acoustic modeling, and the language model respectively.

Emami and Mangu [52], scrutinized usage of the neural network to recognize the Arabic speech by a distributed word representation. Consequently, the neural network makes robust generalization able to fight the data sparseness problem in the best way. They were investigated many factors such as: n-gram order parameter experiment, output vocabulary, the method of normalization, model size and parameters. The evaluation used the Arabic news broadcast, and conversation broadcast. In the result, some enhancement has been achieved using optimized neural network model over 4-gram up to 3.8% relative WER.

Kirchhof, Bilmes and Stolcke in [53] proposed system for conversational Arabic speech recognition. They utilized a language model for morphology. This proposed system enhances the results for two different test sets, this result is 1.8% and 1.5% respectively.

The authors in [54] developed system for an Arabic broadcast news transcription system. The speaker-independent large vocabulary is used in this system. The developed system uses five-state HMM for triphone acoustic models, with 8 and 16 Gaussian mixture distributions. Arabic broadcast news corpus is used in experiments and performs 10.14 as WER.

Hyassat and Abu Zitar in [55] presented system for the holy Quran corpus using language model and lexicon and WER improved by 46.182%.

In [56] Elmahdy and Mohamed developed a dialectal Arabic speech recognition system using a new multilingual approach. This multilingual approach includes several acoustic models using HMM. The news broadcast speech corpus of modern standard Arabic speech and Egyptian colloquial Arabic are used for training and testing. The accuracy reached 99.34%.

Selouani, Sid Ahmed and Malika Boudraa in [57] used MSA continues speech corpus using Algerian Arabic Speech Database (ALGASD) and HHM. As a result the accuracy is achieved 91.65%.

The authors in [58] compared between two Arabic ASR using different techniques. These techniques are the traditional multi-layer perceptron and general regression neural network (GRNN) algorithm. The proposed system involves two components: first, pre-processing component which includes segmental normalization and feature extraction process, second, a classification component which uses neural networks based on nonparametric density estimation. From the results it is clear that the GRNN gives better results and faster performance than the feedforward backpropagation in the recognition rate using Arabic digits corpus.

The authors in [59] investigated the use of a recurrent neural network for Arabic digits related speech recognition. They developed system based on a multi-speaker mode and a speaker-independent mode with accuracy 99.5% and 94.5% correct digit respectively.

In [2] the authors proposed a novel approach for Arabic isolated speech recognition system. The system is implemented using modular recurrent Elman neural networks (MRENN). They conclude that the result almost reaches to the same result by the last traditional HMM-based speech recognition approaches.

The authors in [60] aimed to develop an Arabic ASR through the differences in the 29 letters of the Arabic alphabet. They proposed a system based on a fully-connected recurrent neural network with a backpropagation through time learning algorithm. The investigation has been done twice: first, to prove hidden layer makes the learning of complex classification tasks more efficient, then the comparison of the LPCCC and MFCC. The results overall the LPCCC outperform the MFCC performance by 0.7%.

The authors in [61] introduced three different system structures based on biologically inspired methods for Arabic ASR. The system consists of two parts: features extraction using Mel Frequency Cepstral Coefficients algorithm (MFCC), and adapt the dataset normalization to use for train and test the three different systems. The performances of these systems were 47.52%, 44.58% and 46.63% frame recognition for single MLP identification system, category-based phonemes recognition system and individual Phoneme classifier system respectively.

In [62], authors present the enhancement of the Arabic ASR performance in mobiles communication system. The proposed approach involves two modules: 1) Front-End represent the features extraction stage using MFCC-MT (Multitaper Frequency Cepstral Coefficients features) and Gabor features GF-MFCC. 2) Back-end is represent the recognizer stage, in this stage the authors were investigated different ASR techniques such as CHMM (Continues Hidden Markov Models), DNN (Deep Neural Network) and HMMDNN hybrid. They focused on HMMDNN and claimed that it can get consistently almost 8% of clean speech, 13% of AMR-NB coder and 8.5% of DSR coders.

El-Desoky et al. [63] proposed a novel approach for Egyptian Arabic ASR. They mix the best features of the morpheme-based LMs and feature-rich modeling with the DNN-LMs. On the other hand, make the mixture of words and morphemes along with their features. The result shows that the WER is reduced according to compared to the traditional word-based LMs.

An AASR system was introduced by AlHanai. Et al. [64] with a 1,200-h speech corpus. The developed system used Deep neural network (DNN), DNN structure involving many techniques: Feed-forward, Convolutional, TimeDelay, Recurrent Long Short-Term Memory (LSTM), Highway LSTM (H-LSTM) and

Grid LSTM (GLSTM). The evaluation using corpus shows the best result with 18.3% WER using trained GLSTM models.

In [65] the authors presented description of comparison for some of the state-of-the-art speech recognition techniques details. The corpus that contain 50-h of transcription audio from a news channel “Al-jazeerah” is used to train all different approaches. The hybrid DNN/HMM approach with the MPE (Minimum Phone Error) criterion are trained by sequential DNN achieves the best result. The results were obtained 17.86%, 29.85%, and 25.6% WER for broadcast news reports, a broadcast conversations, and overall respectively.

The Arabic news speech recognition system is developed in [66], the authors of this system presented the KALDI recipe to build it. The broadcast news system using 200 hours GALE data used to trained model. The building of text normalization, vowelization, and language model details are described. The results using the building prototype broadcast news system are 15.81% WER on Broadcast Report (BR) and 32.21% WER on Broadcast Conversation (BC) with a combined WER of 26.95%.

This paper is description of ASR system in in the framework of the 2016 Multi-Genre Broadcast (MGB-2) Challenge in the Arabic language [67]. The main idea of this work was to take the GMM derived features for training a DNN and combined with the use of time-delay neural networks for acoustic models, the combined approach used to automatically phonetize Arabic words. The final system was a combination of a five systems where the result obtained succeeded the best single LIUM ASR system with a 9% of WER.

Ettaouil et al. [68] introduced the Arabic digit system using a hybrid model ANN/HMM. The core of this work is to determine the optimal codebook generated using Self Organizing Maps (SOM) and the optimal class using optimal neural network. From the results, the dictionary size affects the classification. The codebook vectors are with size 34, 36, and 48 they had a recognition rate 84%, 85%, and 86% respectively.

Wahyuni in [69] recognize spoken Arabic letters challenge issue. The proposed methods is built using Mel-Frequency Cepstral Coefficients (MFCC) based on feature extraction and ANN to distinguish between the pronounce of three different letters (sa, sya, and tsa). The accuracy for this work obtains with an average accuracy is 92.42%.

AbdAlmisreb et al. [70] presented the DNN with three hidden layers, 500 Maxout units with 2 neurons for the unit and used Mel-Frequency Cepstral Coefficients (MFCC) for feature extraction. This approach was trained and tested over a corpus which consisted of 20 Malay speakers of consonant Arabic phonemes recorded. The training set consisted of 5 waveforms and the tested set contained 15 waveforms. The result show that the Maxout based deep structure gave better performance with lowest error rate than other deep networks such as Restricted Boltzmann Machine (RBM), Deep Belief Network (DBN), Convolutional Neural Network (CNN), the conventional feedforward neural network (NN) and Convolutional Auto-Encoder (CAE) which had error rate between 2800 and 3000 (numbers).

4 END-TO-END ARABIC ASR

End-to-end approach has different techniques which can be applied on AASR in a simple way. In this paper we focus on the End-to-end methods for Arabic ASR.

Zerari, Naima, et al [71] present general end-to-end approach to recognize the Arabic spoken digit spelling of an isolated Arabic word. This approach used Mel Frequency Cepstral Coefficients (MFCC) for extraction the relevant features from the natural speech signal, these features are presented by a deep neural network able to deal with the non-uniformity of the sequences length of the speech utterances. It firstly used recurrent Long Short-Term Memory (LSTM) with three gates to decode sequences of these features as fixed vector, then a multilayer perceptron network will receive this vector to perform the classification.

Spoken Arabic Digit dataset is used for training and testing, this dataset consists of 8800 tokens by 88 individual (44 males and 44 females) Arabic native speakers. The performance of this approach is achieved 98.77% of global f-measure.

Authors in [72] have implemented and evaluated character-based modeling in a state-of-the-art speech recognition systems for Arabic using end-to-end ASR. This system used Kaldi toolkit to present an effective character-based ASR and evaluate the models based on words, characters, and statistical morphs. In this system, the word, morph, and character *n*-gram models are trained using a projection layer, an LSTM layer, and a highway layer. Recurrent Neural Network Language Models (RNNLMs) also is used for model training. An acoustic models utilize the MGB-2 corpus for training, which consists of 1,200 hours of broadcast data from multiple genres and even dialects.

Ahmed, Abdelrahman, et al. [73] presents the first end-to-end recipe for an Arabic speech-to-text transcription system based on BDRNNs. This system includes a character based decoder which is used for search to avoid using a word lexicon. The Connectionist Temporal Classification (CTC) is also used which, objective function is used to maximize the output character sequences given the acoustic features as an input. It consists of: 1-a BDRNNs acoustic model, 2- a language model and 3- a character based decoder. In addition, the training and decoding process are based on Arabic grapheme. The objective function used to train BDRNNs is CTC that removes the need for pre-segmented acoustic observations. The evaluation for the test set will be performed on both the word and character levels in order to validate the results with other word based models. The recipe was evaluated using 1200 hours corpus of the Aljazeera multi-Genre broadcast programs. On the development set, the WER is 12.03% for non-overlapped speech. The morph-based models yield the best recognition results for both well-resourced and lower-resourced tasks, but the character-based models are close to their performance in the lower-resource tasks, outperforming the word-based models. Character-based models are especially good at predicting novel word forms that were not seen in the training data. In the results, the word model outperforms the character-based model for the full dataset, while for the under-resourced scenario the character-based model improves over the word model by 6%.

In [74], the authors are developed framework for multilingual speech recognition on low-resource languages. This work used sequence-to-sequence attention-based models and a single transformer for recognition. Sub-words are utilized as the multilingual modeling unit without need for any pronunciation lexicon. The ASR transformer is chosen to be the basic architecture of sequence-to-sequence attention-based model. They employed the language-specific softmax layer instead of softmax layer to solve the problem of few training data on low-resource languages. . The CALLHOME datasets with 6 languages: Mandarin (MA), English (EN), Japanese (JA), Arabic (AR), German (GE) and Spanish (SP) are used in experiment. The result on this Arabic dataset was 13.5% average WER.

In [75] an end-to-end speech recognition model analyzed in terms of Arabic phonetic and graphemic representations as well as different articulatory features. In this work traditional ASR systems and separate components for phoneme or grapheme modeling were employed to run forced-alignment and get annotated data to be used by the classifier.

The lexicon also was used the phonetic lexicon for the phoneme system and the word list with 1:1 word-to-character mapping for grapheme system. Convolutional layers and 5 recurrent long short-term memory (LSTM) are used for encoding and trained with CTC. The system consists of two steps: 1) training the E2E model in the normal fashion, on pairs of utterances and transcriptions, 2) running the trained model on a dataset with frame-level annotations. The steps are used to analyze the representation quality in the end-to-end ASR model. MGB-2 corpus which comprises 1,200 hours of broadcast videos from the Aljazeera Arabic TV channel used for training and testing. Experiments on corpus yield 12:94% and 10:60% WER for Phonemes and Graphemes respectively.

Zerari, Naima, et al. [1] proposed to recognize a set of isolated Arabic utterances issued from two ASR applications, namely: a) TV spoken command recognition, b) spoken digit recognition. The proposed system involves, first, extracting pertinent features from the natural speech signal using Mel Frequency Cepstral Coefficients MFCC (static and dynamic features), the Filter Banks (FB) coefficients, and, next, the extracted features are padded in order to deal with the non-uniformity of the sequences length. Then, a deep architecture represented by a recurrent LSTM or GRU (Gated Recurrent Unit) architectures are used to encode the sequences of MFCC/FB features as a fixed size vector that will be introduced to a Multi-Layer Perceptron network (MLP) to perform the classification (recognition).

Based on recurrent neural networks to process sequences of variable lengths of (1) MFCCs, (2) FBs and (3) delta-delta features of the different spoken digits/commands was presented. The extracted features using the different techniques are, first, encoded as a fixed size vector by a recurrent LSTM/GRU neural network, next, a standard Multi-Layer Perceptron is used to classify the spoken digits/commands with the obtained vector as input. Two datasets are used in this work: the first dataset is the spoken Arabic digit A number of 88 (44 males and 44 females) Arabic native speakers were asked to utter all digits ten times. Accordingly, the database consists of 8800 tokens for spoken digit recognition, the second dataset, speaker-independent mode is considered, where one hundred of Arabic native speakers were participated (50 males comprising 37 adults and 13 kids whereas 50 females including 31 adults and 19 kids) for TV spoken command recognition. The accuracies achieved are **98.77% and 96% for** spoken digit recognition and TV spoken command recognition respectively.

5 END-TO-END ARABIC ASR SERVICES

A. API Services

Google developed Cloud Speech-to-Text service application [35] used to convert Arabic speech or audio file to text using a deep-learning neural network algorithm. Cloud Speech-to-Text service allows for its translator system to directly accept the spoken word to be converted to text then translated. The service offers an API for developers with multiple recognition features. Microsoft Speech API [36] is developed by Microsoft. It used deep neural networks to build speech recognition systems. IBM cloud provide Watson service API for speech to text recognition [37] support modern standard Arabic language until now there is not any work use this API with Arabic.

B. Toolkits

Kaldi [76]. It is a toolkit for speech recognition using deep neural network and support Arabic language. Ali et al. [66] shows the usage of Kaldi to build Arabic broadcast news speech recognition system. They use all Kaldi conventional models.

Miao et al. [77] present the Eesen toolkit for end-to-end ASR, this tool is open source and used sequence-to-sequence learning. The deep recurrent neural networks (RNNs) are used as the acoustic models, and the Long Short-Term Memory (LSTM) units the RNN building blocks. The weighted finite-state transducers (WFSTs) are used for decoding. The CTC labels, lexicons and language models components are encoded into WFSTs.

ESPnet [78] is an open source toolkit is provide a neural end-to-end platform for ASR and other speech processing. It emploes Chainer, PyTorch and dynamic neural network as a main deep learning engine. For data processing, feature extraction/format, and recipes this tool used Kaldi ASR toolkit style.

ESPRESSO [79] is a novel neural end-to-end system for ASR ESPRESSO is built upon the popular NMT framework FAIRSEQ2 , and the flexible deep learning framework PyTorch. By extending FAIRSEQ, ESPRESSO inherits its excellent extensibility: new modules can easily be plugged into the system by extending standard PyTorch interfaces. Additionally, we gain ability to perform distributed training over large data sets for ASR. They also present the first fully parallelized decoder for end-to-end ASR, with look-ahead word-based language model fusion, tightly integrated with the existing sets of optimized inference algorithms (e.g. beam search) inherited from FAIRSEQ and tailored to the scenario of speech recognition.

6 CONCLUSIONS

This paper introduced a review on Arabic speech recognition using end-to-end technology. It mentioned the importance of Arabic speech recognition and the limitation of Arabic language. The literature covered the newest papers in Arabic speech recognition based serval techniques and end-to-end approach include Modern Standard Arabic (MSA) and Dialectal Arabic.

Furthermore, the end-to-end model is an important research direction of speech recognition. It uses the deep learning technique and include two parts: attention model and CTC to solve the data alignment size of signal vector issues. Attention method also used to encoding and decoding.

REFERENCES

[1] Zerari, Naima, et al. "Bidirectional deep architecture for Arabic speech recognition." *Open Computer Science* 9.1 (2019): 92-102.

[2] El Choubassi, M.M., El Khoury, H.E., Alagha, C.E.J., Skaf, J.A., Al-Alaoui, M.A.: Arabic speech recognition using recurrent neural networks. In: Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology (IEEE Cat. No. 03EX795), Darmstadt, Germany, pp. 543–547 (2004)

[3] Lipka, M., Hackett, C.: Why Muslims are the world's fastest-growing religious group. Pew Research Center (2017). http://www.pewresearch.org/fact-tank/2017/04/06/why-muslimsare-the-worlds-fastest-growing-religious-group/. Accessed 14 Nov 2018

[4] Ahmed, Basem HA, and Ayman S. Ghabayen. "Arabic automatic speech recognition enhancement." *2017 Palestinian International Conference on Information and Communication Technology (PICICT)*. IEEE, 2017.

[5] Afify, Mohamed, et al. "On the use of morphological analysis for dialectal Arabic speech recognition." *Ninth International Conference on Spoken Language Processing*. 2006.

[6] Alsayadi, Hamzah A., and Abeer M. ElKorany. "Integrating semantic features for enhancing arabic named entity recognition." *Int. J. Adv. Comput. Sci. Appl.(IJACSA)* 7.3 (2016): 2016.

[7] Algihab, Wajdan, et al. "Arabic Speech Recognition with Deep Learning: A Review." *International Conference on Human-Computer Interaction*. Springer, Cham, 2019.

[8] Lippmann, Richard P. "Review of neural networks for speech recognition." *Neural computation* 1.1 (1989): 1-38.

[9] Rabiner L. R., Juang B. H., Fundamentals of speech recognition, PTR Prentice Hall Englewood Cliffs, 1993

[10] Jelinek, Frederick. *Statistical methods for speech recognition*. MIT press, 1997.

[11] Ordowski, Mark, et al. "A public domain speech-to-text system." *Sixth European Conference on Speech Communication and Technology*. 1999.

[12] Huang X., Alleva F., Hon W., Hwang M., and Rosenfeld R., "The SPHINX-II Speech Recognition System: An Overview," Computer Journal of Computer Speech and Language, vol. 7, no. 2, pp. 137-148, 1993.

[13] Satori, Hassan, et al. "Investigation Arabic Speech Recognition Using CMU Sphinx System." *International Arab Journal of Information Technology (IAJIT)* 6.2 (2009).

[14] Ali, Ahmed, Hamdy Mubarak, and Stephan Vogel. "Advances in dialectal arabic speech recognition: A study using twitter to improve egyptian asr." *International Workshop on Spoken Language Translation (IWSLT 2014)*. 2014.

[15] Huang X., Acero A., and Hon H., Spoken Language Processing: A Guide to Theory, Algorithm and System Design, Prentice Hall, 2001.

[16] Hiyassat H., Nedhal Y., and Asem S., "Automatic Speech Recognition System Requirement Using Z Notation," in Proceedings of of AMSE' 05, France, pp. 514-523, 2005.

[17] Huang D., Automatic Speech Recognition: The Development of the SPHINX System, Kluwer Academic Publishers, 1989.

[18] F. Diehl, M. J. F. Gales, M. Tomalin, and P. C. Woodland, "Morphological decomposition in Arabic ASR systems," *Comput. Speech Lang.*, vol. 26, no. 4, pp. 229–243, Aug. 2012.

[19] D. Rybach, S. Hahn, C. Gollan, R. Schluter, and H. Ney, "Advances in Arabic broadcast news transcription at RWTH," in *2007 IEEE Workshop on Automatic Speech Recognition & Understanding (ASRU)*, 2007, pp. 449–454.

[20] L. Mangu, H.-K. Kuo, S. Chu, B. Kingsbury, G. Saon, H. Soltau, and F. Biadsy, "The IBM 2011 GALE Arabic speech transcription system," in *2011 IEEE Workshop on Automatic Speech Recognition & Understanding*, 2011, pp. 272–277.

[21] B. Kingsbury, H. Soltau, G. Saon, S. Chu, H.-K. Kuo, L. Mangu, S. Ravuri, N. Morgan, and A. Janin, "The IBM 2009 GALE Arabic speech transcription system," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 4672–4675.

[22] Juang, Biing-Hwang, and Lawrence R. Rabiner. "Automatic speech recognition–a brief history of the technology development." *Georgia Institute of Technology. Atlanta Rutgers University and the University of California. Santa Barbara* 1 (2005): 67.

[23] Anusuya M. A., Katti S. K., Speech recognition by machine, a review, arXiv preprint arXiv:1001.2267, 2010.

[24] K. Kirchhoff, J. Bilmes, J. Henderson, R. Schwartz, M. Noamany, P. Schone, G. Ji, S. Das, M. Egan, F. He, D. Vergyri, D. Liu, and N. Duta. 2002. Novel approaches to Arabic speech recognition - final report from the JHU summer workshop 2002. Technical report, Johns Hopkins University.

[25] Vergyri, Dimitra, and Katrin Kirchhoff. "Automatic diacritization of Arabic for acoustic modeling in speech recognition." *Proceedings of the workshop on computational approaches to Arabic script-based languages*. Association for Computational Linguistics, 2004.

[26] Belinkov, Yonatan, Ahmed Ali, and James Glass. "Analyzing Phonetic and Graphemic Representations in End-to-End Automatic Speech Recognition." *arXiv preprint arXiv:1907.04224* (2019).

[27] Hori, Takaaki, Jaejin Cho, and Shinji Watanabe. "End-to-end speech recognition with word-based RNN language models." *2018 IEEE Spoken Language Technology Workshop (SLT)*. IEEE, 2018.

[28] Y. Belinkov and J. Glass, "Analysis methods in neural language processing: A survey," *Transactions of the Association for Computational Linguistics (TACL)*, 2019.

[29] T. Nagamine, M. L. Seltzer, and N. Mesgarani, "Exploring how deep neural networks form phonemic categories," in *Interspeech*, 2015.

[30] Nagamine, Tasha, Michael L. Seltzer, and Nima Mesgarani. "On the Role of Nonlinear Transformations in Deep Neural Network Acoustic Models." *Interspeech*. 2016.

[31] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton, “Speech recognition with deep recurrent neural networks,” in *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on*. IEEE, 2013, pp. 6645–6649.

[32] Jan K Chorowski, Dzmitry Bahdanau, Dmitriy Serdyuk, Kyunghyun Cho, and Yoshua Bengio, “Attention-based models for speech recognition,” in *Advances in Neural Information Processing Systems (NIPS)*, 2015, pp. 577–585.

[33] Alex Graves and Navdeep Jaitly, “Towards end-to-end speech recognition with recurrent neural networks,” in *International Conference on Machine Learning (ICML)*, 2014, pp. 1764– 1772.

[34] Suyoun Kim, Takaaki Hori, and Shinji Watanabe, “Joint CTC attention based end-to-end speech recognition using multitask learning,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 4835–4839.

[35] Takaaki Hori, Shinji Watanabe, and John R. Hershey, “Joint CTC/attention decoding for end-to-end speech recognition,” in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL): Human Language Technologies: long papers*, 2017.

[36] Caglar Gulcehre, Orhan Firat, Kelvin Xu, Kyunghyun Cho, Loic Barrault, Huei-Chi Lin, Fethi Bougares, Holger Schwenk, and Yoshua Bengio, “On using monolingual corpora in neural machine translation,” *arXiv preprint arXiv:1503.03535*, 2015.

[37] Takaaki Hori, Shinji Watanabe, Yu Zhang, and William Chan, “Advances in joint CTC-attention based end-to-end speech recognition with a deep CNN encoder and RNN-LM,” in *INTERSPEECH*, 2017.

[38] Anjuli Kannan, Yonghui Wu, Patrick Nguyen, Tara N Sainath, Zhifeng Chen, and Rohit Prabhavalkar, “An analysis of incorporating an external language model into a sequence-to-sequence model,” *arXiv preprint arXiv:1712.01996*, 2017.

[39] Wang, Song, and Guanyu Li. "Overview of end-to-end speech recognition." *Journal of Physics: Conference Series*. Vol. 1187. No. 5. IOP Publishing, 2019.

[40] Wang, Dong, Xiaodong Wang, and Shaohe Lv. "An Overview of End-to-End Automatic Speech Recognition." *Symmetry* 11.8 (2019): 1018.

[41] Miao, Y., & Metze, F. (2017). End-to-End Architectures for Speech Recognition. In *New Era for Robust Speech Recognition* (pp. 299-323). Springer, Cham.

[42] N. Bostrom. Superintelligence: Paths, Dangers, Strategies. Oxford University Press, 1st edition, 2014.

[43] Yu, D., Deng, L.: Automatic Speech Recognition: A Deep Learning Approach, pp. 13–21. Springer, London (2015). https://doi.org/10.1007/978-1-4471-5779-3.

[44] Le, Xuan-Hien, et al. "Application of long short-term memory (LSTM) neural network for flood forecasting." *Water* 11.7 (2019): 1387.

[45] D. Britz, “WILDML,” 17 September 2015. [Online]. Available: http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1- introduction-to-rnns/. [Accessed 21 March 2018].

[46] Olah, C. Understanding LSTM Networks. Available online: http://colah.github.io/posts/2015-08-Understanding-LSTMs/ (accessed on 20 June 2020).

[47] A. Hannun, “Sequence Modeling with CTC,” *Distill*, 2017.

[48] Graves et al. ”Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks.”(ICML 2006).

[49] A. Graves, Supervised Sequence Labelling with Recurrent Neural Networks, Berlin: Springer-Verlag Berlin Heidelberg, 2012.

[50] Nasereddin, H.H.O., Omari, A.A.R.: Classification techniques for automatic speech recognition (ASR) algorithms used with real time speech translation. In: 2017 Computing Conference, London, pp. 200–207 (2017)

[51] Turab, N., Khatatneh, K., Odeh, A.: A novel Arabic Speech Recognition method using neural networks and Gaussian Filtering. (IJEECS) Int. J. Electr. Electron. Comput. Syst. 19 (01) (2014)

[52] Emami, A., Mangu, L.: Empirical study of neural network language models for Arabic speech recognition. In: 2007 IEEE Workshop on Automatic Speech Recognition & Understanding (ASRU), The Westin Miyako Kyoto, pp. 147–152 (2007)

[53] Kirchhoff, K., Vergyri, D., Bilmes, J., Duh, K., Stolcke, A.: Morphology-based language modeling for conversational Arabic speech recognition. Comput. Speech Lang. 20(4), 589–608 (2006)

[54] Alghamdi, M., Elshafei, M., Al-Muhtaseb, H.: Arabic broadcast news transcription system. Int. J. Speech Technol. 10(4), 183–195 (2007)

[55] Hyassat, H., Abu Zitar, R.: Arabic speech recognition using SPHINX engine. Int. J. Speech Technol. 9(3–4), 133–150 (2006)

[56] Elmahdy, M., et al.: Modern standard Arabic based multilingual approach for dialectal Arabic speech recognition. In: Eighth International Symposium on Natural Language Processing, SNLP 2009. IEEE (2009)

[57] Selouani, S.A., Boudraa, M.: Algerian Arabic speech database (ALGASD): corpus design and automatic speech recognition application. Arab. J. Sci. Eng. 35(2C), 15 (2010)

[58] Amrouche, A., Rouvaen, J.M.: Arabic isolated word recognition using general regression neural network. In: 2003 46th Midwest Symposium on Circuits and Systems, Cairo, Egypt, vol. 2, pp. 689–692 (2003)

[59] Alotaibi, Y.A.: Spoken Arabic digits recognizer using recurrent neural networks. In: Proceedings of the Fourth IEEE International Symposium on Signal Processing and Information Technology, Rome, Italy, pp. 195–199 (2004)

[60] Ahmad, A.M., Ismail, S., Samaon, D.F.: Recurrent neural network with backpropagation through time for speech recognition. In: IEEE International Symposium on Communications and Information Technology, ISCIT 2004, Sapporo, Japan, vol. 1, pp. 98–102 (2004)

[61] Hmad, N., Allen, T.: Biologically inspired continuous Arabic speech recognition. In: Bramer, M., Petridis, M. (eds.) SGAI 2012, pp. 245–258. Springer, London (2012). https://doi.org/10.1007/978-1-4471-4739-8_20

[62] Bouchakour, L., Debyeche, M.: Improving continuous Arabic speech recognition over mobile networks DSR and NSR using MFCCs features transformed, 12, 8 (2018)

[63] El-Desoky Mousa, A., Kuo, H.-K.J., Mangu, L., Soltau, H.: Morpheme-based feature-rich language models using deep neural networks for LVCSR of Egyptian Arabic. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, pp. 8435–8439 (2013)

[64] AlHanai, T., Hsu, W.-N., Glass, J.: Development of the MIT ASR system for the 2016 Arabic multi-genre broadcast challenge. In: 2016 IEEE Spoken Language Technology Workshop (SLT), San Diego, CA, pp. 299–304 (2016)

[65] Cardinal, P., et al.: Recent advances in ASR applied to an Arabic transcription system for AlJazeera, p. 5

[66] Ali, Ahmed, et al. "A complete KALDI recipe for building Arabic speech recognition systems." *2014 IEEE spoken language technology workshop (SLT)*. IEEE, 2014.

[67] Tomashenko, N., Vythelingum, K., Rousseau, A., Esteve, Y.: LIUM ASR systems for the 2016 multi-genre broadcast Arabic challenge. In: 2016 IEEE Spoken Language Technology Workshop (SLT), San Diego, CA, pp. 285–291 (2016)

[68] Ettaouil, M., Lazaar, M., En-Naimani, Z.: A hybrid ANN/HMM models for arabic speech recognition using optimal codebook. In: 2013 8th International Conference on Intelligent Systems: Theories and Applications (SITA), Rabat, Morocco, pp. 1–5 (2013)

[69] Wahyuni, E.S.: Arabic speech recognition using MFCC feature extraction and ANN classification. In: 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, pp. 22–25 (2017)

[70] AbdAlmisreb, A., Abidin, A.F., Tahir, N.: Maxout based deep neural networks for Arabic phonemes recognition, p. 6 (2015)

[71] Zerari, Naima, et al. "Bi-directional recurrent end-to-end neural network classifier for spoken Arab digit recognition." *2018 2nd International Conference on Natural Language and Speech Processing (ICNLSP)*. IEEE, 2018.

[72] Smit, Peter, et al. "Character-based units for unlimited vocabulary continuous speech recognition." *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. IEEE, 2017.

[73] Ahmed, Abdelrahman, et al. "End-to-End Lexicon Free Arabic Speech Recognition Using Recurrent Neural Networks." *Computational Linguistics, Speech And Image Processing For Arabic Language* 4 (2018): 231.

[74] Zhou, Shiyu, Shuang Xu, and Bo Xu. "Multilingual end-to-end speech recognition with a single transformer on low-resource languages." *arXiv preprint arXiv:1806.05059* (2018).

[75] Belinkov, Yonatan, Ahmed Ali, and James Glass. "Analyzing Phonetic and Graphemic Representations in End-to-End Automatic Speech Recognition." *arXiv preprint arXiv:1907.04224* (2019).

[76] KALDI. http://kaldi-asr.org/. Accessed 18 Feb 2019

[77] Miao, Yajie, Mohammad Gowayyed, and Florian Metze. "EESEN: End-to-end speech recognition using deep RNN models and WFST-based decoding." *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*. IEEE, 2015.

[78] Watanabe, Shinji, et al. "Espnet: End-to-end speech processing toolkit." *arXiv preprint arXiv:1804.00015* (2018).

[79] Wang, Yiming, et al. "Espresso: A Fast End-to-end Neural Speech Recognition Toolkit." *arXiv preprint arXiv:1909.08723* (2019).

BIOGRAPHY

Dr. Abdelaziz A. Abdelhamid is assistant professor in Faculty of Computer and Information Sciences. He got his MSc degree from Faculty of computer and information sciences in 2009, and got his PhD degree from Faculty of Engineering, The university of Auckland in 2014. He has more than 25 publications in international conferences and journals. His research interests are speech recognition, speech synthesis, image processing, and deep learning.

Hamzah A. Alsayadi is PhD student in Faculty of Computer and Information Sciences. He got his BSc degree from Faculty of sciences in 2007, and held an MSc degree from Faculty of Computer and Information Technology, Cairo university in 2016. His research interests are speech recognition, speech synthesis, Neutral language processing, and deep learning.

Dr. Islam Hegazy is an Assistant Professor in the Faculty of Computer and Information Sciences, Department of Computer Science. Dr. Islam has more than 18 years experience in research and teaching in many fields in the fields of Computer Science. He holds a PhD degree from the University of Calgary, AB, Canada. He did his M.Sc. and B.Sc. degrees in the Faculty of Computer and Information Sciences, Ain Shams University. His research interests focus on network, security, and cloud computing. Dr. Islam acquired several managerial skills as the Director of the Scientific Computing Center, Ain Shams University, and the coordinator of the Software Engineering Credit Hours program, Faculty of Computer and Information Sciences.

Prof. Zaki Taha is a Professor in the Faculty of Computer and Information Sciences, Department of Computer Science. Prof. Zaki has more than 30 years experience in research and teaching in many fields in the fields of Computer Science. He holds a PhD from Communication Department, Facility of Engineering, Ain Shams University 1997. He did his M.Sc. in Computer Science (Speech Processing), Communication and computer department, Faculty of Engineering, Ain Shams University, 1992 and B.Sc. degrees in in Electronic Engineering (Communication Engineering), Military Technical College, 1976. His research interests focus on Speech Processing and Computer Security. Prof. zaki held several positions in Military IS Center, Ministry of Defense, Military Academy, Armed Forces and Faculty of Computer and Information Sciences.

TRANSLATED ABSTRACT

أستعراض الطريقة الشاملة للتعرف على الكلام باللغة العربية

عبدالعزيز عبدالمنعم *،** ، حمزة احمد الصيادي * ، إسلام حجازي * ، زكي طه فايد

*قسم علوم الحاسوب – كلية الحاسبات و المعلومات – جامعة عين شمس – مصر

[1]Abdelaziz@cis.asu.edu.eg

[2]hamzah.sayadi@cis.asu.edu.eg

[3]islheg@cis.asu.edu.eg

[4]zaki.taha@cis.asu.edu.eg

**قسم علوم الحاسوب – كلية الحاسبات و تقنية المعلومات – جامعة شقراء – المملكة العربية السعودية

ملخص:

يعد التعرف التلقائي على الكلام مجالًا مهمًا بسبب تطبيقاته الكثيرة التي يمكن تطويرها لمساعدة البشر على تحسين مهام حياتهم اليومية. على الرغم من تاريخ التعرف على الكلام الطويل ، إلا أنه لا يزال مجالًا بحثيًا مهما وممتعًا بشكل عام وفي اللغة العربية بشكل خاص. اللغة العربية هي واحدة من أكثر اللغات انتشارًا. ومع ذلك ، لا تزال الاعمال البحثية الموجودة حاليا محدودة بسبب الاختلافات الكثيرة في تراكيب الجمل و الحركات الاعرابية المعقدة. لذلك ، هذا العمل يسلط الضوء على أحدث التقنيات المستخدمة في التعرف على الكلام باللغة العربية لمساعدة الباحثين المهتمين بالبحث في التعرف على الكلام باللغة العربية. هناك العديد من تقنيات التعلم الآلي المطبقة في بناء أنظمة التعرف على الكلام. أستمرت تقنيات (HMMs) والنماذج المختلطة الجاوسية (GMMs) لفترة طويلة أفضل طرق للتعامل مع انظمة التعرف على الكلام. ومع ذلك ، في العقد الماضي ، برزت نماذج الشبكة العصبية الهجينة (DNN) المختلطة مع (HMM) وطرق التعلم العميق و الطريقة الشاملة (end-to-end) كتقنيات جديدة لتحسين أداء هذه الانظمة. تتميز الطريقة الشاملة (end-to-end) بأنها أحدث منهجية في هذا المجال ويمثل المحور الرئيسي لهذا العمل. لذلك ، يناقش هذا العمل أحدث إنجازات البحث في مجال التعرف على الكلام باللغة العربية من منظور المنهجية الشاملة (end-to-end). بالإضافة إلى ذلك ، يتم شرح بعض الخدمات وعدد من المكتبات البرمجية المتوفرة حاليًا اللازمة لبناء نماذج شاملة (end-to-end) للتعرف على الكلام.

الكلمات المفتاحية: التعرف الآلى على الكلام، التعلم العميق، **LSTM; CTC; RNN; HMM .**