Contributions Of Arab Institutions To The Computerization Of The Arabic Language "Prospects And Challenges"

Dr. Mahfoudi Salima Mohamed Cherif Messaadia University - Souk Ahras (Algeria).
s.mahfoudi@univ-soukahras.dz

Abstract:

The world today is experiencing a technological renaissance at various levels, which has led to fundamental innovations in lifestyles in various fields, as it has imposed a change in methods and means of work in various institutions and sectors, which has produced a new type of management known as digital management, and its impact has been reflected in the field of language, where attention has been paid to Many linguists and technology specialists have sought to exploit the great potential provided by the computer and its various applications in processing language levels effectively, quickly, and with minimal effort.

Language computerization has become one of the most important researches undertaken by Arab specialists in order to join the technological civilization. Therefore, this intervention came to reveal the most important Arab institutions that were interested in this aspect, in addition to the works of the most prominent Arab linguists who participated in computerizing the Arabic language.

Keywords: language, computing, institutions, technology, software.

1. Introduction:

Computer It played a fundamental role in disseminating scientific facts, making it impose its mechanisms and software in various fields. From this standpoint, specialists devoted themselves to studying the principles of this science and its techniques that enable them to process human languages in a technical, automated manner. And Why There is no doubt that language have benefited Of use Soft ware the computer Variety after showing the possibility Adapting its systems to suit the specificity of the Arabic language at all its linguistic levels, including phonetic, morphological, grammatical, lexical, and semantic. The researchers were able to Arab linguists assist computers in processing vocabulary, correcting grammatical, morphological, and spelling errors automatically, and developing electronic dictionaries.

Through these efforts, belief increases in the ability of the computer to overcome linguistic obstacles, such as processing the semantic level of prose and poetic texts and

formulating them mathematically according to the techniques provided by the computer. Hence, through this intervention, we are trying to raise the following general problem: What addition have Arab institutions and Arab linguists made in the field of computerization of the Arabic language?

2. Definition of computer It is an electronic device that can be programmed to enter and process data, store it, retrieve it, or show it to the user in another way. If we look at a device that has the ability to perform multiple operations in a few seconds, and a computer, from a comprehensive look, we find that its function goes beyond processing the entered data, so it can transfer it to another computer, that is, exchange information. With computersThe other and that fromduring Information Network.

3. Institutions Concerned with computerization of the Arabic language

The Arab world knows many people from Scientific institutions Concerned with language computing, which It works to process the Arabic language automatically at all levels. Among these institutions, we mention, but are not limited to, the following:

1.3: Natural Language Processing Research Team, School of Computer Science, Leeds:
NLP - School of Computing at the University of Leeds
An academic institution specializing in computer science inThe University of Leeds is ranked among the top ten academic institutions specializing in computer science. The research team pays special attention to Arabic language techniques and linguistic resources. And it is concentrated his research on language modeling and building Arabic linguistic codes. It is based at the University of Leeds - United Kingdom [1].

2.3: the organization Arab League for Education, Culture and Science (ALECSO) Science Management:
A number of Arab academics and experts work on language computing projects at the Department of Science through work teams and research groups. A group of institutions and bodies concerned with computing the Arabic language also cooperate with the organization. It should be noted that the Arab Organization for Education, Culture and Science is keen to provide the completed Arabic language techniques in a free or open source form, seeking to make them available to researchers on the one hand, and working to develop them and address their shortcomings on the other hand. It is based in Tunisia[2]

3.3 King Abdulaziz City for Science and Technology (Computer Research Institute):
established The Institute was founded in 1992. The Institute includes four scientific departments: the Department of Phonetics and Linguistics, the Department of Scientific Computing, the Department of Systems and Networks, and the Department of Software Engineering and Innovative Systems. The Department of Phonetics and Linguistics is

concerned with preparing research and studies and providing solutions to the problems of computing the Arabic language. It also provides consultations and holds conferences and specialized scientific meetings to follow up on developments in the fields of language technology and research. Its headquarters are: King Abdulaziz City for Science and Technology - Riyadh - Kingdom of Saudi Arabia Yes[3]

4.3: institution: LDC Linguistic Data Consortium

She a research institution, in which a group of universities and institutions concerned with linguistic resources contribute, in addition to a number of governmental research laboratories. It was established in 1992 with support from the Advanced Research Projects Agency (ARPA). The LDC Foundation pays attention to building linguistic blogs, written and spoken linguistic databases, and linguistic dictionaries for research and development purposes. It is based at the University of Pennsylvania - United States of America[4]

5.3: group Columbia Arabic Dialect Modeling Center, Columbia University:

Group We formed a research team at the Center for Computer Learning Systems at Columbia University. Means the team With the support of Arabic dialects, and based on contemporary Arabic language standards, the group adopts a project for automatic speech recognition. The project aims - in general - to determine the differences between Arabic dialects at the level of sounds, structures, and lexicon. The group's interest also extended to automatic translation from Arabic to English and dealing with Arabic dialects. Headquarters: Columbia University - New York - United States. [5]

6.3-a team Research in natural language processing at Stanford University:

University Stanford is a private American university that was opened on October 1, 1891. The Stanford Natural Language Processing Team consists of the Departments of Linguistics and Computer Science, whose members work together on algorithms that allow computers to process and understand human languages. Headquarters: California - United States of America[6]

7.3 institute Qatar Computing Research Center

is a non-profit research institute, established as part of the Qatar Foundation for Education, Science and Community Development in 2010. Its specializations include Arabic language technologies, high-performance computing, and bioinformatics. Headquarters: Doha, Qatar.[7]

8.3 Association Egyptian Language Engineering

Created In 1996 at the Faculty of Engineering at Ain Shams University, it is concerned with language computerization in general and the Arabic language in particular. The association works - in accordance with its stated goals - to bring together academics, researchers, technicians, producers and users from different places and different

languages, which helps to exchange information in order to improve the technology of communication between humans and machines in natural language. Aspects of caring for Arabic language technologies include following up on technical development in the fields of automated processing, and opening channels for research and academic cooperation with scientific societies concerned with language computerization at the local, national and international levels. Headquarters: Ain Shams University - Cairo - Egypt[8]

9.4 center Arabic language techniques

Institution a non-profit organization. The center was established in 2009 with the support of technical companies and research and academic institutions. It aims to provide linguistic resources for those concerned with computing the Arabic language, including researchers and developers. It also aims to develop human resources through training in the fields of Arabic language engineering and...[9]Pursuit To find specialized scientific degrees in universities and academic institutions. The center is concerned with providing technical services to people with special needs through developing appropriate technologies .for their needs ,It is also concerned with developing techniques to assist non-native speakers in learning the Arabic language. Headquarters: - Egypt[10]

4 Efforts of some Arab scholars in the field of computational linguistics:

Arabic scholars have worked hard in the field of computerizing the Arabic language, and among the contributions made are the efforts he made" Nabil Ali" wrote a book. Arabic Language and Computers 1988 AD, where he called for the adoption of informatics and mathematical logic in dealing with levels, including the morphological level, where he says: "The automation of morphological processes for the Arabic language is a basic input and a common denominator for most of its automated systems. It also indicates that the extent of our success in Arabizing information and knowledge systems depends primarily on what we can achieve on the morphological side. On the technical level, processing Arabic morphology automatically is a requirement." It is essential and indispensable for automating dictionaries, retrieving information, and analyzing the content of texts."[11]

This scientific author worked on establishing the rules and mechanisms of computational linguistics, which benefited researchers and those interested in computing technologies Ahmed Ghazal (Director of the Institute of Studies and Research for Arabization in Morocco) was also credited with developing an Arabic linguistic model that works on electronic computers with two alphabetic systems. He called this linguistic model, which is "questionable standard Arabic," the Arabic code. Amm-Shaa), this researcher has tried to explain the principles of this system in relation to the historical development of Arabic calligraphy and writing and how to adapt Arabic drawing to contemporary electronic accounting technology.[12]

As for the world, "Nohad Moussa"He wrote a book called:"Arabic towards a new description in light of computational linguistics 2000, and He is a pioneer in this field because his book is considered the first building block in the field of computerization of the Arabic language and he: "An attempt to move from describing Arabic to describing it,

in light of the general thesis of computational linguistics. Practical experience shows that there is a big difference between describing the language, abstracting its examples, and controlling its rulings, where this description is directed to humans, and when it is designed to be deposited in the computer. Perhaps this is what he carried "Nihad Musa" established the difference between these two sciences, as he named what works for humans "Description" and what works for the computer "Characterization The explanation of the difference between them is that the description of Arabic is the rules that Arab scholars have learned from actual linguistic performance, and it is partly based on the fact that the recipient contributes actively to the communicative event, in addition to the elements that humans recognize through intuition, common sense, and multiple clues. Whereas The computer lacks this pure human element. The describer must compensate for this deficiency in order to reach the level of human knowledge of language with the computer.[13]

As for Abdul Rahman Haj Saleh, his project is based on achieving what he called the Arabic repertoire, which is the establishment of an automatic bank of Arabic language words, relying on "A large blog of the real use of the Arabic language is a necessary condition for first achieving the objectivity of linguistic research and achieving the effectiveness of linguistic development and generation by giving the developed words the opportunity to actually enter into use and for people to accept them."[14],

Haj Saleh believes that the linguistic repertoire serves the Arab researcher in his research and studies and is similar to the achievements of the ancient Arabs when they wrote down the words of the Arabic language. They were interested in "listening, that is, tracking and browsing the words of the Arabs, writing them down in the form of poetry collections and extensive blogs of their proselytized speech, and monitoring the use of the Arabs wherever they were." They searched for the extent of this usage, and who are the ones who frequently use this or that word in this form and with this meaning, and whether it is frequent or not.[15]

The contributions of Abdul Rahman Haj Saleh had an eloquent impact in revealing the logical thinking followed by Al-Khalil bin Ahmed Al-Farhidi in his various linguistic achievements, as he alerted those working in the field of linguistics that the language's characteristics enable it to enter the field of computing with ease, which opened the way for Arab researchers. Likewise, the Algerians should pay attention to this aspect and move forward in order to advance Arab research

5-Arabic language and computers
The beginning was with the international company "Sakhr" for its owner Abdul Rahman Al-Sharekh Where he put a working computer It works with a system The Japanese MSX, which was written in Arabic, made Sakhr computers operate in an authentic Arabic environment.

The second experiment was launched within the "Alice" company" ALIS", which was sent by Mr. Bashir Helmy, an Algerian born in Canada, which attempted to design an Arabic DOS system compatible with the MS-DOS system developed by the American software

company Microsoft Microsoft for its owner Bill Gates, before the two companies agreed to include the Arabic version in the list of versions available in many languages in the MS operating system. -DOS.

As for the third experience, it was through which some Arab experts tried to provide the UNIX systemU NIX in Arabic, in line with what they noticed of the increasing importance of this system, and the extent of its exploitation, whether on small, medium, or large computers..**[16]**

We loved it after that "We go beyond focusing on the issue of computer thinking to move deeper into developing the use of the Arabic language and its information capabilities"[17].

6- Level Y Linguistic analysis

1.6 Analysis Phonological
Set at this stage road Pronunciation of the word, taking into account letters that are not pronounced or that are pronounced without their original origin (such as assimilated letters and letters in which two consonants meet). The sciences of Tajweed are the basic means that help to know the rules of pronunciation of the Holy Qur'an, while noting the contrast between them and the rules of pronunciation of classical Arabic in general. In Arabic in particular, we face a difficulty in not using short vowels, and this makes achieving the automatic pronunciation mechanism for Arabic words a difficult task compared to its language counterparts whose writing takes into account short and long vowels, such as many Latin languages.[18]

2.6- Morphological analysis
Decompose The word to the elements RHathe basic ; According to the etymological nature of the Arabic language, an Arabic word usually has more than one possibility, but we can choose the most appropriate possibility based on the context. This selection is made using statistical methods that also help in identifying the parts of speech and the descriptions related to them.[19]

3.6- Grammatical analysis
You are awaren The function of the word in the sentence in terms of its grammatical position, which helps to understand the meaning (connotation); however, the existing connection between grammatical analysis and meaning makes the matter more difficult in the Arabic language, due to its flexibility in arranging the components of the sentence (such as the predicate precedes the subject or the subject follows the On the other hand, the role of meaning cannot be ignored in the descriptive analysis in determining the grammatical function of the word. An example of this is the word (worker). From the descriptive point of view, it is (a participle) of the verb (work), but from the grammatical point of view it is suitable as a subject or object. Or an adjective or something else

4.6AnalysisSemantic

Semantic analysis is considered one of the most important and difficult stages, as semantics has many levels The lowest Determining the meaning of the word in its context (such as determining the meaning of the word "eye" on a predator, a spy, or a water well). It should be noted that semantic analysis has many sub-levels, such as: Resolving the semantic ambiguity of the word. For example, the word "eye" comes in the meaning of a predator. , or a water well, or a spy...etc. As a sub-discussion of the meaning of the word: identifying the names of objects (an example of which is the word "Cairo", which is a feminine adjective, or a flag for the city of Cairo.

Jaw Verbal confusion. Examples of this include: "The girl loved her sister very much, so when she met her she ran to her." The point of confusion is that we find it difficult to know who ran to whom. Is it the girl to her sister? Mother

Sister To the girl? We may call it "removing pronoun ambiguity."

Jaw Rhetorical ambiguity: For example, "I saw lions in battle," when the intent is: I saw brave soldiers.

the chapter Between subjects. In many cases, the article deals with more than one topic, and we do not necessarily find sub-headings to separate the partial topics. Yes; Each topic can have separate paragraphs,

But The number of paragraphs for each topic is different and they are not separated by a subtitle. How can we separate partial topics? This separation is necessary when retrieving useful information; If it is possible to indicate the exact location of useful information in the article or book, this will save researchers a lot of time.[20]

5.6- Lexical analysis

The researcher identifies the components of the linguistic lexicon, and at this level he is concerned with phonological analysis, while the second level is concerned with the morphological aspect in order to extraction The basic units of the lexicon [buildings]; Syntactic analysis is concerned with inferring the contexts of vocabulary and their functional meanings. It is also concerned with semantic analysis in inferring lexical meanings.[21]

6- Method of processing the Arabic language Computationally

The computer is a deaf machine that does not care about the context that surrounds the circumstances of speech, despite the attempt to subjugate it. Language is an endless use, and this is what makes it difficult to understand structures and sentences. The computer only deals with the precise and precise, which requires, if we want to computerize the language, that we reveal the minute details of the structure of the language and the information. Complete with its vocabulary, and this is possible, even if difficult[22]

Levels of computer analysis

a. Letter

Trilingual operating systems have been developed(Arabic, English, French) Input and output peripheral units dealing with the Arabic language were also developed, such as keyboards and programs for automatically reading Arabic texts, writing from right to left, and choosing between Arabic numbers used in the Arab Levant and the Arab Maghreb. Printers and operating systems were also Arabized. And software. [23] It should be noted thatOperating systems as a whole became available using the Arabic letter, and this was helped by the expansion of the Arab information market, which made Microsoft allow for continuous circulation of several micro-information systems that take into account the peculiarities of the Arabic language, so that it became easy to use the Arabic language in the fields of innovation, creativity and communication through Artificial intelligence and adapting information imagination and its technologies to the needs of using the Arabic language.[24]

B. Morphological analysis

and he link The words of the text are characterized by the primary morphological elements that enter into their formation, as well as by the grammatical values without regard to their location. In the analysis, we move from the word to its original root. That is, the computer processes Arabic words that are partially, completely, or not formed, and describes the changes that occur to them. An increase, a decrease, an accusative, a substitution, an assimilation, or an inversion, where "the type, its morphological balance, its antecedent (or antecedents) and suffix (or suffixes) are determined, its grammatical status, and its connotation... If the word to be analyzed contains letters other than It is problematic. The computer determined the possible movements for it based on stored data. It is known that the word's lack of form makes it multi-form, and therefore has multiple meanings, as long as it is independent of the context of the text[25].

C. Word analysis:

The computer processes words by extracting the original root of the word and processing the problem
Or completely, or unproblematic, and determines its type, morphological balance, antecedent, suffix, grammatical status, and connotations.... If the word to be analyzed contains non-problematic letters, the computer sets the possible movements for it based on the statistics of the matching of the movements with the letters in preparation for its analysis. . It is known that the lack of a form in a word makes it have multiple forms, and hence meanings. As long as it is independent of the context of the text.
The word "rose," for example, can have the following possibilities:
Rose = (received_ received_ received_.....) verbs
= (Ward_ Ward....) names
= (and + response - and + response) conjunction + verbs
= (and+ radd - wa+ radd) conjunction + nouns[26]

7. Conclusion

Automated processing of the Arabic language has been able to provide a lot of accurate information that will help us join the pace of civilization and keep pace with developments in the field of linguistic research.

It has become clear through the achievements made by Arab institutions and researchers that the Arabic language is one of the most suitable languages for computing due to its diverse characteristics, whether at the level of sound, lexicon, or morphology.

Despite the many obstacles, the determination to adapt information technology in its various aspects was behind a number of experiments .in Operating systems, although most of these experiments were directed at operating systems for family computers and personal computers.

8.. Research footnotes

[1]. Abdul Rahman Al-Hajj Saleh, Research and Studies in Arabic Linguistics, Part 1, Movm Publishing, Algeria 2007, p. 85.

[2]. Mohsen Rashwan; Al-Moataz Billah Al-Saeed, Introduction to the Computerization of the Arabic Language, Dar Wujooh Al-Nisr and Distribution, Kingdom of Saudi Arabia - Riyadh, 1st edition, 2019, p. 51.

RashwanImproverSame, p. 52..[3]

[4]. Mohsen Rashwan, previous reference, p. 52.

[5]. Mohsen Rashwan himself, p. 55

[6]. Mohsen Rashwan himself, p. 56

[77]. Mohsen Rashwan himself, p. 56

[8]. Mohsen Rashwan; Al-Mu'tazz Billah Al-Saeed, Introduction to Computerizing the Arabic Language itself, p. 57

[9]. Mohsen Rashwan; Al-Mu'tazz Billah Al-Saeed, Introduction to the Computerization of the Arabic Language itself, pp. 57, 58

[10]. Abdul Rahman Al-Hajj Saleh, Research and Studies in Arabic Linguistics, Part 1, Movm Publishing, Algeria 2007, p. 85.

[11]. A Reading of the Contributions of Arabic Computational Linguistics: Horizons and Bets, Saeed Fahim, Mouloud Mammeri University, Center for Scientific and Technical Research for the Development of the Arabic Language, Dirasat Journal of the University of Laghouat, issue 36, September 2015, p. 134.

[12]Ibid., p. 135

[13]Saeed Fahim, A Reading of Arabic Computational Linguistic Contributions, Horizons and Bets, same, p. 135.

[14]. Abdul Rahman Al-Haj Saleh: Computerization of Arab heritage and intellectual production in a computerized Arab repertoire as a national project, Research and Studies in Arabic Linguistics, p. 151.

14- Ibid.,p. 152

[16]. Muhammad bin Ahmed, the Arabic language, computer systems, and software, in the book "Using the Arabic Language," p. 125, from the Arab Heritage Magazine, p. 101.

17. Abdullah Abu Haif, The Arabic Language and the Challenges of Globalization, in the book, "The Arabic Language Conference in the Face of the Challenges of Globalization," pp. 466-498.

18. Mohsen Rashwan; Al-Moataz Billah Al-Saeed, Introduction to the Computerization of the Arabic Language, Dar Wajoh Al-Nisr and Distribution, Kingdom of Saudi Arabia - Riyadh, 1st edition, 2019, p. 18

19 Ibid., p. 19

20. Mohsen Rashwan; Al-Mu'tazz Billah Al-Saeed, previous reference, p. 21

21. Marwan Al-Bawab, Muhammad Hassan Al-Tayyan, method of processing the Arabic language in informatics(Word - Sentence), Using the Arabic Language in Informatics, Arab Educational, Cultural and Scientific Organization, Tunisia, 1996, p. 25.

22. Maroush Dhiyab Zaghdouda, The relationship of the Arabic language to computers, Journal of the College of Arts and Human Sciences, No. 3, 2009, p. 118

23. Abdullah Abu Haif, The Future of the Arabic Language, Computerization of the Arabic Dictionary and its Linguistic and Technical Problems as a Model, Arab Heritage Magazine, Arab Writers Union, Syria, No. 93, 94, p., 2004,

24. Abdulaziz bin Abdullah Al Mahyoubi to morphological analysis, an introduction to computational linguistics, p. 50

25. See: Marwan Al-Bawab, Muhammad Hassan Al-Tayyan, Method of Processing the Arabic Language in Informatics (Word - Sentence), Using the Arabic Language in Informatics, Arab Organization for Education, Culture and Science, Tunisia, 1996, p. 36.

26. Abdulaziz bin Abdullah Al-Muhayoubi for morphological analysis, an introduction to computational linguistics, p. 50