

Computational linguistics and automatic processing of Arabic morphology

Dr. Akila Zemouri ¹, Dr. Rachida Abed ², Dr. Hamrani Leyla ³, Dr. Fatima Abderrahmane ⁴,
Dr. Hadjer Hadjcherif ⁵

¹ University of Nour Al-Bachir, Al-Bayadh (Algeria)

¹ Linguistics and Literature Laboratory. Email: a.zemouri@cu-elbayadh.dz

² University of Khemis Miliana, Algeria

Email: r.abed@univ-dbk.m.dz; Orcid: 0009-0004-7644-3059

³ University of Hassiba Ben Bouali, Chlef (Algeria). Email : Leyla.hamrani@univ-alger2.dz

⁴ University of Hassiba Ben Bouali, Chlef (Algeria). Email: f.abderrahmane@univ-chlef.dz

⁵ University of Hassiba Ben Bouali, Chlef (Algeria). Email : hadjerhadjcherif19@gmail.com

Received: 05/01/2025 ; Accepted: 01/03/2025 ; Published: 27/04/2025

Abstract:

Computational linguistics is rapidly gaining importance with modern technological developments, where natural language meets computing to develop tools and techniques that contribute to the automatic understanding and processing of languages. The Arabic language is of particular interest due to its prominent status as a religious and cultural language. Therefore, the role of automatic processing of Arabic morphology is to analyze and generate linguistic weights, forms, and conjugations in an advanced manner. In this presentation, we will explore how linguistics and computing intersect in the field of automatic Arabic morphology processing, and how these technologies can contribute to the development of tools that facilitate the effective and accurate understanding and analysis of the Arabic language.

Keywords: computational linguistics, computer, automatic processing, morphology.

Introduction:

Arabic is one of the greatest languages due to God's blessing of it as the language of the Holy Quran. To promote its growth and

development, linguistic studies overlap with scientific techniques for its automatic processing.

Computers are the pinnacle of modern technology, so it was only natural that language and computers would come together. Language is the embodiment of human mental activity, while computers mimic mental abilities by translating language through symbols. Among the automated language processors, we find the automated processing of Arabic morphology, as it is the most computable linguistic level, by building rules to create a database through which Arabic words can be projected for processing and generated automatically. What is computational linguistics? How is Arabic morphology processed automatically?

1. Definition of linguistics:

Linguistics has been known by several names, including linguistics, philology, linguistics, linguistics, linguistics, and linguistics. Since its emergence at the beginning of the twentieth century, it has been able to study language in all its fields.

Linguistics is defined as "the scientific and objective study of the human tongue through the tongues of each society. It is the study of the human tongue, characterized by science and objectivity."¹ Linguistics seeks to create a general linguistic theory through which all human languages can be studied and described. The first task in linguistics begins with describing all the special phenomena that can be observed. The establishment of linguistics depends on collecting data from diverse and comprehensive observations of the different forms of the tongue as much as possible."²

Linguistics is a comprehensive, language-focused scientific study that looks at how languages are constructed, organized, and used by humans. Linguistics includes the study of linguistic structures such as morphology (change in words), syntax (sentence structure), and semantics (meanings), as well as the study of phonetics (sounds) and semiotics (symbols and symbolism).

In addition, linguistics also studies the evolution of languages over time and the influence of social and cultural factors on them. It also studies standard linguistics and local dialects, and how they change and evolve across regions and generations.

A modern trend in linguistics is the study of interactions between different languages, whether within a single domain, such as translation and linguistic interaction between two languages, or across multiple cultures,

such as communication between multiple languages in multilingual societies.

In short, linguistics is a comprehensive and in-depth study of languages and their role in human life, addressing multiple aspects including linguistic structures, historical development, and cultural and social interactions.

1.Computer definition:

The computer is considered one of the advanced technological means that has effectively contributed to growth and development in various sectors. Its definitions have varied and multiplied, including:

Sami Kazem Hassan Al-Ramahi defines it: "It is a programmed machine consisting of a physical entity (Hardware)) and a software store (software) that gives it the flexibility to deal with tasks that are formulated in the form of programs consisting of instructions that operate the electronic calculator to process data very quickly, efficiently, and with reliable results"³

Riyad Al-Sayed also defines it as "a programmable electronic device characterized by the ability to store, process, and communicate with humans and other devices and components using various media, and one of its most important functions is processing data to obtain information"⁴

If a computer is an electronic device or machine used in various areas of life, it receives data and then processes it and outputs

¹ Ahmed Hassani, *Studies in Linguistics*, Office of University Publications, Central Square, Ben Aknoun, Algeria, ed., 1994, p. 14.

²See: Jean Perrault, *Linguistics*, translated by Al-Hawas Masoudi and Miftah bin Arous, Dar Al-Afaq, 2001, p. 8.

³ Rania Muhammad Qasim Ali, *The Impact of Computers on Children's Social Relations*, Alexandria Book Center, ed., 2009, p. 63.

⁴ Rania Al-Sayed, *Introduction to Computer Science*, Dar Al-Hamid for Publishing and Distribution, Jordan, 1st ed., 2000, p. 24.

it in the form of information after storing it for a long time. Among its advantages is the tremendous speed in processing data and its storage capacity. It is also distinguished by the accuracy in processing data.

2. Computational linguistics:

Computational linguistics is one of the applied branches that is concerned with benefiting from computer data in studying various linguistic issues such as: monitoring linguistic phenomena according to their levels, phonetic, morphological, grammatical, rhetorical, and prosodic, performing statistical operations, dictionary making and machine translation, and teaching languages.¹

Computational linguistics is one of the newest branches of linguistics that processes language with computers. This science is divided into two halves. The first half belongs to general linguistics, whose subject is language, and the second half belongs to computational linguistics, whose subject is translating language into symbols, or preparing language to be a language for communication and dialogue with computers. The study of the Arabic language using computational linguistics is one of the latest linguistic trends in contemporary Arabic linguistics.

3. Automated language processing:

Computer specialists and linguists have turned to processing levels of the Arabic language, as automated language processing is concerned with studying the computational

aspects of the language and the various problems facing this processing.

Definition of processing:

Processing, according to computational linguistics, is the automated application of a set of language texts by changing and transforming them, and creating something new based on them. All of this is done using techniques and tools from linguistics, automated media, and modeling (modélisation). When processing, a distinction must be made between describing knowledge, which is the function of linguistics, and expressing this knowledge in models using effective techniques and strategies derived from computer science, which is the function of computational linguistics.²

Mechanism definition:

Automated operations are those performed by machines and their counterparts are operations performed by humans. The computer is the one that performs the mathematical operations, as it must be developed in order to process information of a linguistic nature. Whereas the automatic processing of language is the sequence of mathematical movements performed by the machine according to a chronological sequence, that is, the automatic processing program is the sequence of mathematical movements performed by the machine according to a chronological sequence, that is, the automatic processing program (programme

¹ Abdul Qadir Abdul Jalil, *Modern Linguistics*, Dar Al-Safa, Jordan, 1st ed., 2002, p. 181.

² See: Fares Shasha, *Automated Processing of the Arabic Language, Creating a Linguistic, Morphological,*

and Grammatical Model of the Arabic Verb, Master's Thesis in Library Science and Documentation, University of Algiers, Faculty of Social and Human Sciences, Algeria, 2008, p. 13.

Automatique) can be complete (Total) or partial (Contraintes), as:¹

1) Total: The computer does everything.

2) Partial: Man intervenes in some stages.

Processing anything linguistic in a machine faces many limitations in describing the linguistic texts themselves, so the components of the texts must be modeled in a clear and consistent manner.

Language definition:

Language has great value in human life, as it is a means of communication and interaction between people, through which each individual expresses his innermost thoughts and feelings. Ibn Jinni says about this: "As for its limits, they are sounds through which each people expresses their goals."²

4. Automated processing of Arabic exchange:

Arabic morphology: It is a branch of linguistics concerned with studying the structures and formation of words in the Arabic language, including weights, roots, and inflections. Morphology is considered an essential part of the Arabic linguistic system that contributes to understanding the structure of words and their formation based on certain patterns. "Morphology is a science by which the conditions of word structures are known, which are not inflected or constructed."³

Morphology is a science derived by Muslim scholars who created rules for studying the

internal structure of the Arabic word in terms of formulating it to convey meanings, or in terms of searching for its incidental conditions.⁴

Conjugation is the balance of Arabic, and with it the origin of the word is transformed into several structures, Ibn Asfour says: "Conjugation is divided into two parts: one of them is making the word in different forms for types of meanings, such as hit and hit, hit, hit, hit, and hit...Hence the difference in the form of the noun due to the meanings that plague it, from diminutive and broken...This type of conjugation is usually mentioned by grammarians along with something that is not a conjugation...The other of the two types of morphology is: changing a word from its origin, without that change indicating a meaning that occurs in the word, such as their change from "qawal" to "qalal."⁵ So the science of morphology is that science that studies the changes that occur in a word; meaning that morphology is the knowledge of the rulings of a word because of the addition, validity, and explanation of its letters.

Automated processing of Arabic exchange:

By automated processing of Arabic morphology, we mean computer-based processing based on its various programming styles. This is achieved by developing models for automated morphology processing. A number of systems for this analysis have emerged, the most important of which is the

¹ Ibid., same page

² Ibn Jinni, Characteristics, edited by: Muhammad Ali Al-Najjar, Alam Al-Kutub, Beirut, Scientific Library, vol. 1, p. 23.

³ Mohsen Ali Attia, The Arabic Language: Its Levels and Applications, Dar Al-Manahij for Publishing and Distribution, Amman, Jordan, ed. (1429 AH/2009 AD), p. 51.

⁴ See: Hatem Saleh, The Guarantor Exchange, Dar Al-Hikma Press, Baghdad, 1st ed., 1991, p. 55.

⁵ Ibn Asfour Al-Ishbili, Al-Mumti' fi Al-Tasrif, edited by: Fakhr Al-Din Qabawa, Dar Al-Ma'rifah, Beirut, Lebanon, 1st edition, vol. 1, (1407 AH/1987 AD), p. 31.

multi-phase morphological analyzer from Sakhr International.¹

The members of the Egyptian Sakhr Company focus on the grammatical morphology processor and the etymological processor, without neglecting the aspects of parsing and diacritics.

A. Grammatical morphology processor:

This processor disassembles the word and strips it of adhesives and suffixes, thus going to the origin of the word so that it becomes suitable for processing, such as the word: religion.²

In this word, it is suggested that there are several possibilities, including:

1. Father + Yen Muthanna Al-Masculine
2. Father + Yen masculine plural of Salem
3. And + religion from the verb dan
4. And + religion from the verb dan

B. Derivative processor:

It extracts the root of the word and the morphological form that the morphological processor has broken down grammatically.³

Word root weight morphological probability

Religion is a religion that acts (by soothing the eye) and religion

Din Dan Faal (with a fatha on the 'ayn) and religion

The father of a son is active (with a kasra on the 'ayn) and the religion (with a kasra on the dal)

The father of a son is active (with a kasra on the 'ayn) and the religion (with a sukoon on the 'ya)

When analyzing morphology automatically, we divide the word into suffixes, which in turn include the previous and subsequent ones. Then we search for the root of the word, the state it is in, and finally the formation of the word. Such as: the word "He will support them".

Primary adhesives = precedents

Supports the trunk, present tense, does

Nasr root of the word: past tense verb.

They are final suffixes = suffixes. If "Yansar" is classified as a verb according to its state and automatic description.

Example 2: The word "he will let them in"

They are primary adhesives and are therefore considered precedents

She enters the stem, it is a present tense verb, it does...

The verb "entered" is the root of the word: it is a past tense verb

They are a pronoun for final suffixes = suffixes

"enters" is classified into verbs according to their state and automatic description.

¹ See: Ali Nabil, The Arabic Language and the Computer, Dar Al-Ta'rib, Beirut, 1st ed., 1988, p. 308.

² Hosni Muzaffar Al-Razou, Ways to Develop an Automated Exchange Analyst, an article published on the Internet, 2008, p. 4

³ See: ibid., same page.

Conclusion:

- Computational linguistics is a modern science that uses computers to convert texts into digital computer languages for analysis and translation into other languages.
- Automated language processing is concerned with studying the computational aspects of language and the various problems facing this processing. It is one of the means that facilitates deep and luxurious texts, especially in the age of the atom and the computer.
- The morphological aspect is concerned with how the word is constructed, its root, and the additions that follow it. Through the automated morphological analyzer, we can study and analyze the Arabic text automatically.
- A morphological analyzer is a program that studies the structure of a word and converts it to its original root.

Through these points, we realize that computational linguistics represents a vital field that combines linguistics and technology, contributing to enhancing our understanding of the structure and use of languages in a deep and comprehensive manner in our contemporary societies.

Approved references:

1. Ibn Jinni, Characteristics, edited by: Muhammad Ali Al-Najjar, Alam Al-Kutub, Beirut, Scientific Library, Part 1.
2. Ibn Asfour Al-Ishbili, Al-Mumti' fi Al-Tasrif, edited by: Fakhr Al-Din Qabawa, Dar Al-Ma'rifah, Beirut, Lebanon, 1st ed., vol. 1, (1407 AH/1987 AD).
3. Ahmed Hassani, Studies in Linguistics, University Publications Office, Ben Aknoun Central Square, Algeria, DT, 1994.
4. Hatem Saleh, Guarantor Exchange, Dar Al-Hikma Press, Baghdad, 1st ed., 1991 AD.
5. Ali Nabil, The Arabic Language and the Computer, Dar Al-Ta'rib, Beirut, 1st ed., 1988.
6. Abdul Qadir Abdul Jalil, Modern Linguistics, Dar Al-Safa, Jordan, 1st edition, 2002 AD.
7. Mohsen Ali Attia, The Arabic Language: Its Levels and Applications, Dar Al-Manahij for Publishing and Distribution, Amman, Jordan, ed. (1429 AH/2009 AD).
8. Rania Muhammad Qasim Ali, The Impact of Computers on Children's Social Relations, Alexandria Book Center, ed., 2009.
9. Jean Perrault, Linguistics, translated by Al-Hawas Masoudi and Miftah bin Arous, Dar Al-Afaq, ed., 2001 AD.
10. Fares Shasha, Automated Processing of the Arabic Language, Creating a Linguistic, Morphological, and Grammatical Model of the Arabic Verb, Master's Thesis in Library Science and Documentation, University of Algiers, Faculty of Social and Human Sciences, Algeria, 2008.
11. Riyadh Al-Sayed, Introduction to Computer Science, Al-Hamid Publishing and Distribution House, Jordan, 1st ed., 2000.
12. Hosni Muzaffar Al-Razou, Ways to Develop an Automated Exchange Analyst, an article published on the Internet, 2008.