Pegem Journal of Education and Instruction, Vol. 15, No. 4, 2025 (pp. 4942-4947)

RESEARCH ARTICLE WWW.PEGEGOG.NET

The problem of computing lexical meaning in the context of the project The linguistic repertoire of Abdul Rahman al-Haj Saleh

Dr. Radia Ben Ariba[1], Dr. Hakima Fekaouni [2], Dr. Fatima Abderrahmane [3]

[1] University of Hassiba Ben Bouali, Chlef (Algeria)
Email: r.benarbia@univ-chlef.dz
[2] University of Hassiba Ben Bouali, Chlef (Algeria)
Email : Hakimafekaounidz@gmail.com
[3] University of Hassiba Ben Bouali, Chlef (Algeria)
Email: f.abderrahmane@univ-chlef.dz

Received: 23/03/2025 ; Accepted : 12/07/2025 ; Published : 11/10/2025

Abstract :

The enormous capabilities of modern computers and their superior ability to investigate linguistic systems have opened up vast horizons for the use of computers in this field in areas such as linguistic statistics, linguistic analysis and composition, machine comprehension of texts, text analysis, machine translation, and machine processing of dictionaries.

The Arabic Lexicon project, adopted by the Algerian scholar Professor Abdel Rahman Al-Haj Saleh (may God have mercy on him), aims to create an automated database of ancient Arabic texts, particularly those related to Arab cultural heritage, as well as modern texts, such as contemporary Arab intellectual works and the most important international scientific works in Arabic, on a website .

The automated lexicon is based on processing a single word selected by the user and displays the most important roots available for that word in the encyclopaedia under study, allowing the user to select the appropriate root for the search.

The automated processing of Arabic semantics also requires a large amount of information about various aspects of the language, and the best way to organise this information is in databases that include rules relating to writing and morphology, separating the root of the word from its prefixes and suffixes, and linking the root to known morphological patterns. This is followed by grammatical databases that require the division of affixes into their components.

Keywords: Lexical semantics – Machine processing – Semantic networks – Language resources – Abdul Rahman Al-Haj Saleh.

Text of the presentation:

Machine language processing focuses on studying the computational aspects of language and the linguistic

and computational problems that this processing faces, whether the language is spoken or written.

Building an Arabic language processing system is a complex and difficult task, due to the difficulty of integrating phonological, morphological, syntactic and semantic knowledge into this system.

1- Definition of machine processing:

Processing: the automatic application to a set of language texts, by modifying and transforming them, and creating something new based on them, using techniques and tools from linguistics, computer science and modelling. A distinction must be made between the description of knowledge, which is the function of linguistics, and the expression of this knowledge in models using effective techniques and strategies derived from computer science, which is the function of computational linguistics.

Automated processes are those carried out by machines, as opposed to those carried out by humans. Computers are the machines used to process language. They were invented to perform calculations, and had to be developed to process information related to the nature of language, since automated processing is a sequence of computational movements performed by the machine in chronological order, meaning that the automated processing programme (Progamme Automatique) can be total (Total) or partial (Contraintes), where:

- Total: the computer does everything.

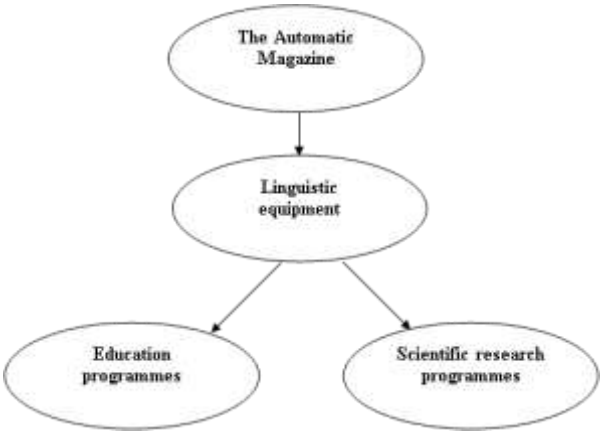
- Partial: humans intervene at certain stages.

2- Areas of research in natural language processing:

Areas of research in natural language processing are divided into three main areas:

First: a common area, which refers to linguistic hardware.

Second: research programmes.



Third: educational programmes

3-The objective of Arabic language processing:

The importance of language processing is evident in its many important applications, such as:

- Learning Arabic.
- Teaching Arabic to non-native speakers.

- Computer-assisted machine translation from and into Arabic.
- Detecting and correcting linguistic errors using spell checkers and grammar checkers.
- Automatic recognition of human speech and its structure.
- Automatic reading of written texts.
- Interacting with machines in written language.
- Interacting with machines in natural language.
- Automatic indexing of texts.
- Formatting of unstructured or partially structured texts.

4- The Arabic Language Resource Project:

4-1- Project objectives:

The Arabic Resource Project aims to create an automated database of ancient Arabic texts, particularly those relating to Arab cultural heritage, as well as modern texts, such as contemporary Arab intellectual works and the most important international scientific works in Arabic, on a website. This automated text database has two important aspects: ( )

First:

The linguistic aspect: it is a 'diwan of the Arabs' because it represents the actual use of the Arabic language, both ancient and modern, through millions of literary, scientific, technical and other texts (or what is known in our time as a linguistic database).

Second:

The cultural aspect: including scientific and educational aspects, because its content brings together, through texts, all information related to scientific, technical, historical, social and other fields. It can be consulted to obtain any information contained in computerised Arabic texts, and researchers can refer to the original text if the Arabic text has been translated.

The enormous capabilities of modern computers and their superior abilities to investigate linguistic systems have opened up vast horizons for the use of computers in this field in areas such as: linguistic statistics, linguistic analysis and synthesis, machine comprehension of texts, text analysis, machine translation, and machine processing of dictionaries.

4-2 -Objectives of the linguistic resource:

- **The resource is an automated information bank:**

This is the most important objective of the resource project, as it greatly expands the scope of use of classical Arabic. Abdul Rahman Al-Haj Saleh says: The most important thing is that this usage, which will be stored in text form in computer memory, represents the finest examples of Arabic usage over the past fifteen centuries, covering the entire Arab world in the best way possible."

- **Generating different specialised dictionaries:**

The repository collects a huge and endless number of old and modern words, which form a general and specialised dictionary:

- The comprehensive dictionary of Arabic words in use: It contains Arabic words in use and their meanings.

-Comprehensive dictionary of commonly used Arabic words: Contains commonly used Arabic words and their meanings.

-Dictionary of commonly used scientific and technical terms: Contains Arabic words, even those used in only one country, accompanied by their English and French translations. Words for which no translation exists are listed on their own. This dictionary is then divided into specialised dictionaries according to subject and field of research. Concepts can be added to this repository at any time, i.e. it is possible to create new entries and add new information .

-Dictionary of Foreign and Coined Words

- Dictionary of Geographical Terms

-Historical Dictionary of the Arabic Language

and other specialised dictionaries.

5- Characteristics and importance of the linguistic repository:

-Its comprehensiveness and inclusiveness, covering the actual use of the Arabic language in all Arab countries and spanning from the pre-Islamic era to the present day.

- Its representation of this usage through the inclusion of all important texts, both written and spoken, in literature, civilisation, religion, science, general culture, the arts, and everyday life.

-Its reliance on electronic devices such as computers, mobile phones and others that can store an infinite number of sentences and retrieve information at high speed.

-The ability to ask thousands of questions remotely, simultaneously across the world, display them on screen, and print them using printers, lasers, and other devices in a short time.

6- Functions of linguistic resources:

-Gathering information about Arabic words, whether common or technical

-Gathering information about word types

-Gathering information about the role of poetic devices, rhymes, alliteration, and so on.

- Collecting information about cultural or practical concepts (searching for Arabic words to cover scientific concepts.(

The automatic processing of Arabic semantics requires a large amount of information about various aspects of the language, and the best

way to organise this information is in databases that include rules relating to writing and morphology, separating the root of the word from its prefixes and suffixes, and linking the root to known morphological patterns. This is followed by grammatical databases that require the division of affixes into their components.

7-Automated processing of the Arabic lexicon:

Automated processing of the Arabic lexicon occupies a distinguished position in the fields of dealing with the Arabic language from an informational perspective, and is considered a fundamental and common denominator in the management of most of its automated systems concerned with the creation of information and knowledge systems.

The automated processing of the Arabic lexicon refers to the use of advanced computing systems based on software algorithms that utilise Arabic lexical logic in processing Arabic words by extracting the basic elements of the word structure and directly determining its lexical features.

The automated lexicon processes a single word selected by the user and displays the most important roots available for that word in the encyclopaedia under study, allowing the user to select the appropriate root for the search.

It should be noted that the multi-stage automated lexical processor is characterised by an advanced mechanism in terms of pure programming and lexical and linguistic processing, which has earned it a prominent position in the field of computerised encyclopaedia research. Given the absence of automated lexical analysers in the Arab world that match its level of linguistic and programming sophistication, we will attempt to study its mechanism of operation to gain a better understanding of its lexical, linguistic and programming structure.

8- Semantic databases:

A database is nothing more than a collection of general ideas that need further refinement and expansion. The codification of meanings and semantics is a difficult task for modern languages, let alone Arabic, which has a wide range of meanings and semantics. Semantic databases include:

8-1 The root database:

Roots are interrelated in terms of meaning. For example, roots such as: came, arrived, came, arrived, managed, and so on, as well as: entered, entered, accepted... etc. These groups are interconnected as they relate to other interconnected groups, as they relate to other groups in terms of human movement back and forth or going out and coming in.( )

The roots alone are mostly combined to form the origins of many meanings, and therefore relying on them gives a general but inaccurate meaning .( )

8-2 Rule of representation in semantic networks:

Semantic networks are geometric networks that represent concepts and connect them with lines representing the types of relationships that exist between these elements. These networks became popular in the early 1960s in various fields, most notably psychology and library science, and then evolved to represent small texts at Cambridge University (). With the emergence of cognitive psychology and knowledge engineering, these networks evolved as scientists attempted to mimic the human mind. Automatic semantic processing requires automatic sentence processing, as it affects the meaning of the sentence and determines its semantic meaning within the context.

9-Marginalisation:

See: Automated Processing of the Arabic Language, Creation of a Morphological Model for Arabic Verbs, Fares Shasha, Master's Thesis in Library Science and Documentation, University of Algiers, Faculty of Social Sciences and Humanities, Algiers, 2008, p. 13.

2 See: Automated Processing of the Arabic Language, Fares Chacha, p. 13.

3 See: Research and Studies in Arabic Linguistics, Abdelrahman Haj Saleh, Algeria, 2007, vol. 2, p. 153.

4. The Arabic Language Resource Project and its Scientific and Educational Dimensions, Abdel Rahman Haj Saleh, Al-Adab Magazine, Algeria, Issue 3, p. 7.

5. See: Previous reference, pp. 7-8.

6. See: Ibid., pp. 9-10.6.

7. See: Previous reference, pp. 11-12.7-

8. See: Towards Addressing Semantics in the Arabic Language through Databases, A Preliminary Study of the Text of the Holy Qur'an, Muhammad Zaki Al-Khader, University of Jordan, 17th National Conference on Computers, King Abdulaziz University, Madinah, April 2004, p. 415.

9. See: Ibid., p. 429.

10. See: Statistics of Arabic Verbs in the Computerised Dictionary, Marwan Al-Bawab et al., Lebanon Library, 1996, p. 32.

L’intelligence Artificielle et la longue, Sabah, Gérard, Edition hermès, Paris, 1988, p. 152.