**ALTRALANG Journal**
Volume: 03 Issue: 01 / July 2021 pp. 138-145
e-ISSN: 2710-8619 p-ISSN: 2710-7922

# Arabic Language and Computers. Application of Computational Linguistics to serve the Arabic Language

**Salim MEZHOUD**[1]
[1]*University Center Abdelhafid Boussouf, Mila, Algeria*
s.mezhoud@centre-univ-mila.dz

**Received:** 16/01/2021, **Accepted:** 30/06/2021, **Published:** 31/07/2021

**ABSTRACT**: *The success of modern software for natural language processing impresses. Programs for orthography and grammar correction, information retrieval from document databases, and translation from one natural language into another, among others, are sold worldwide in millions of copies nowadays.*

*The relationship of the Arabic language to the computer in the process by which the learner acquires the capacity to perceive and comprehend language (in other words, gain the ability to be aware of language and to understand it), as well as to produce and use words and sentences to automated communication.*

**KEYWORDS**: Computational linguistics, Language, Machine translation, Processing, Software.

**RÉSUMÉ :** *Les ordinateurs sont devenus les outils les plus importants de l'activité linguistique, car les études de linguistique informatique reposent sur des programmes informatiques pour les systèmes de langage humain, en normalisant et en simulant le système cérébral humain pour des systèmes de travail pour ordinateurs. mécanisme. Cet article vise à clarifier la relation de la langue aux ordinateurs, à découvrir l'influence de la linguistique computationnelle sur le développement de la langue arabe et à généraliser son travail automatisé dans divers domaines. Basé sur l'approche descriptive, l'article explique comment tirer parti des capacités des programmes informatiques dans l'analyse et le traitement de la langue arabe pour comprendre diverses sciences et connaissances, et pour pratiquer l'enseignement et l'apprentissage.*

**MOTS-CLÉS** : Langage, Linguistique computationnelle, Logiciels, Traduction automatique, Traitement

ALTRALANG Journal
Volume 03 Issue 01 / July 2021

**الملخص**: أصبح الحاسوب أهم أدوات النشاط اللغوي، إذ تعتمد دراسات اللسانيات الحاسوبية على برامج جاسوبية لأنظمة اللغات البشرية، من خلال تقييس ومحاكاة نظام عمل الدماغ البشري لنظم عمل الحواسيب الآلية، فهي فرع تطبيقي حديث، يستغل التكنولوجيا المتطورة من أجل بلورة برامج وأنظمة لمعالجة اللغات الطبيعية، معالجة آلية .يهدف هذا البحث انطلاقا من المنهج الوصفي إلى توضيح علاقة اللغة بالحاسوب والكشف عن تأثير اللسانيات الحاسوبية في تطوير اللغة العربية، وتعميم العمل الآلي بها في مختلف المجالات، ويوضح طريقة الاستفادة من قدرات برامج الحاسوب في تحليل اللغة العربية ومعالجتها لفهم العلوم والمعارف المتنوعة، وممارسة التعليم والتعلّم.

**الكلمات المفتاحية**: برمجيات، ترجمة آلية، لسانيات حاسوبية، لغة، معالجة.

## Introduction:

The modern era world is characterized by the widespread use of natural language processing technologies that are part of the global process of the digitalization of society. As billions of users in this world send and retrieve information, give voice and written commands, and use many symbols. Although they do not realize the importance of their interaction, they are actually contributing to the development of the use of algorithms and software applications for processing natural languages texts.

Computational linguistics deals with the study of computer systems that are dedicated to the analysis and generation of natural language units (Grishman, 1986, p. 4)

In Arabic Linguistic Studies, the term "natural language processing" (NLP), is used to define the concept of computational linguistics, but special attention should be paid to the term “applied linguistics,” whose accepted Arabic scientific meaning differs from its Anglo–American, or generally Western interpretation. So Until recently, applied linguistics was understood as a language teaching methodology, because these developments in applied linguistics were dedicated to language study, especially of English as a foreign or second language.

At Now, the field of applied linguistics has become more and more broad, as it includes treatment of aphasia problems, speech disorders and translation problems.

To the extent that language is a mirror of mind, a computational understanding of language also provides insight into thinking and intelligence. And since language is our most natural and most versatile means of communication, linguistically competent computers would

ALTRALANG Journal
Volume 03 Issue 01 / July 2021

greatly facilitate our interaction with machines and software of all sorts, and put at our fingertips, in ways that truly meet our needs, the vast textual and other resources of the internet.

**Definition of computational linguistics:**

Computational linguistics is an interdisciplinary field concerned with the computational modelling of natural language, as well as the study of appropriate computational approaches to linguistic questions. In general, computational linguistics draws upon linguistics, computer science, artificial intelligence, math, logic, philosophy, cognitive science cognitive psychology, psycholinguistics, anthropology and neuroscience, and others. Traditionally, computational linguistics emerged as an area of artificial intelligence performed by computer scientists who had specialized in the application of computers to the processing of a natural language. With the formation of the Association for Computational Linguistics, and the establishment of independent conference series, the field consolidated during the 1970s and 1980s. The term "computational linguistics" is now a days (2020) taken to be a near-synonym of natural language processing (NLP) and (human) language technology. These terms put a stronger emphasis on aspects of practical applications rather than theoretical inquiry and since the 2000s. In practice, they have largely replaced the term "computational linguistics" in the NLP/ACL community, although they specifically refer to the sub-field of applied computational linguistics.(Tim, 2020, p. 2)

If we say that computational linguistics was initially aimed at the study of natural languages, then natural language must be defined.

In neuropsychology, linguistics, and the philosophy of language, a natural language or ordinary language is any language that has evolved naturally in humans through use and repetition without conscious planning or premeditation. Natural languages can take different forms, such as speech or signing. They are distinguished from constructed and formal languages such as those used to program computers or to study logic. (Lyons, 1991, p. 68)

Computational Linguistics is a field of linguistics that deals with making computers understand human language. Some of the biggest sub-fields of computational linguistics are:

- Speech Recognition, which is a computer program that listens to people talk and writes down what they said

ALTRALANG Journal
Volume 03 Issue 01 / July 2021

- Speech Synthesis, which is a computer program that takes writing and reads it out loud
- Machine Translation, which is a computer program that turns one language into a different one
- Dialog Systems, which is a computer program that talks back-and-forth with a human to help them do something.

**Beginnings and development of computational linguistics:**

Computational Linguistics, or Natural Language Processing (NLP), is not a new field. As early as 1946, attempts have been undertaken to use computers to process natural language. These attempts concentrated mainly on Machine Translation and, due to the political situation at the time, almost exclusively on the translation from Russian into English. Considerable resources were dedicated to this task, both in the U.S.A. and in Great Britain, during the fifties and sixties. Other countries, mainly in continental Europe, joined the enterprise, and the first systems ("SYSTRAN") became operational at the end of this period. However, the limited performance of these systems made it clear that the underlying theoretical difficulties of the task had been grossly underestimated, and in the following years and decades much effort was spent on basic research in formal linguistics. Today, a number of Machine Translation systems are available commercially although there still is no system that produces fully automatic high-quality translations (and probably there will not be for some time). Human intervention in the form of pre- and/or post-editing is still required in all cases. Another application that has become commercially viable in the last years is the analysis and synthesis of spoken language, i.e., speech understanding and speech generation. Potential applications go from help for the handicapped (e.g., text-to-speech systems for the blind) to telephony based information systems (e.g., inquiry systems for train or plane connections, telebanking) and further on to office dictation systems (as offered by several vendors). Several text-to-speech systems are commercially available, and are in daily use in many places. The difficulties of speech understanding are much greater than those for speech generation yet some of the speech understanding systems are also entering the marketplace. An application that will become at least as important as those already mentioned is the creation, administration, and presentation of texts by computer. Even reliable access to written texts is a major bottleneck in science and commerce. The amount of textual

information is enormous (and growing incessantly), and the traditional, word-based, information retrieval methods are getting increasingly insufficient as either precision or recall is always low (i.e., you get either a large number of irrelevant documents together with the relevant ones, or else you fail to get a large number of the relevant ones in the collection). Linguistically based retrieval methods, taking into account the meaning of sentences as encoded in the syntactic structure of natural language, promise to be a way out of this quandary. However, the creation of texts is also becoming a problem. Manuals of complex technical systems (airplanes, computers etc.) are constantly out of date as the systems themselves are upgraded ever faster. Writing manuals by hand is thus getting ever more expensive and unreliable, and if manuals have to be maintained in different languages, manual production becomes increasingly unmanageable. If different versions of the manuals have to be written (for service users, for technicians, for auditors etc.), things get out of hand altogether. The automatic creation of manuals from a common knowledge base, in different languages and for different types of readers is a possible solution of this cluster of problems. The creation of natural language texts has always been a bit of "poor cousin" in the field of Computational Linguistics. The situation described is about to change this in a fundamental manner. Another topic that might come to the forefront of research in Computational Linguistics is the presentation of textual information. Traditionally, text generation systems have created standard, i.e., linear, text. If the amount of text is large, and/or if different types of readers must be addressed, hypertext is a better medium of presentation. The automatic creation of hypertext from an underlying knowledge base calls for an extension of this traditional approach (Martin, 2015, pp. 1-2).

**Machine Translation:**

Machine translation is a sub-field of computational linguistics that investigates the use of software to translate text or speech from one natural language to another.

In the 1950s, machine translation became a reality in research, although references to the subject can be found as early as the 17th century. The Georgetown experiment, which involved successful fully automatic translation of more than sixty Russian sentences into English in 1954, was one of the earliest recorded projects (Gordin, 2015, p. 8). Researchers of the Georgetown experiment asserted their belief that machine translation

would be a solved problem within three to five years. In the Soviet Union, similar experiments were performed shortly after (Madsen, 2009, p. 11) .

Consequently, the success of the experiment ushered in an era of significant funding for machine translation research in the United States. The achieved progress was much slower than expected; in 1966, the ALPAC report found that ten years of research had not fulfilled the expectations of the Georgetown experiment and resulted in dramatically reduced funding. Interest grew in statistical models for machine translation, which became more common and also less expensive in the 1980s as available computational power increased. Although there exists no autonomous system of "fully automatic high quality translation of unrestricted text", there are many programs now available that are capable of providing useful output within strict constraints. Several of these programs are available online; Google Translate and SYSTRAN system that powers Alta Vista's BabelFish (Bar-Hillel, 1964, p. 174).

Some researchers and philosophers believed that digital computers would achieve linguistic universality by overcoming the differences between languages and within the same language per se. For example, mathematician Warren Weaver in1949, explained his vision and hope for linguistic universalism in a note that became a catalyst for research Machine translation in United States of America (Hutchins, 2000, p. 17)

Warren Weaver expected that computers would solve the problem of the infinite diversity of languages by defining a global infrastructure upon which all human languages are built, and he predicted that computing machines would be able to translate between all languages, to build a bridge between the different forms of human communication. Common Human Communication, established by Weaver by a universal set of rules by which all languages must operate (Weaver, 1955, p. 23),

Anyone who has experienced modern translation applications over the Internet will know that Weaver's dream is on its way to fulfilment, as computer programs now can recognize human speech, but the goal of a single language that brings together all human languages into a single global infrastructure, remains elusive for the time being. While the promise of machine translation had long helped research into speech recognition technologies, by the 1970s speech science had begun to abandon the search for a universal, undiscovered Weaver language.

ALTRALANG Journal
Volume 03 Issue 01 / July 2021

Despite the amazing human ambition to talk to the machine, we still use speech recognition techniques in specific applications in electronic shopping, the automated voice assistant that leads us to issue a statement, or the mobile phone applications that direct us through a series of specific actions to the target What is meant, however, is that we do not always need to speak to a machine. In many social and psychological situations, we need to speak to people like us.

**Conclusion:**

Computational linguistics is an important science that helps in the civilized response to the rapid technological development in various fields, and thus it can be employed in the service of the Arabic language and its sciences.

I suggest the following:

- All basic types of Arabic words (noun, verb, and letter) must be electronic coding, as well as subclasses such as pronouns, nouns, sign nouns, etc., according to the phonemic and lexical system of the Arabic language.
- Solutions should be found to the capitalization problem, so that we can distinguish between nouns and adjectives in machine translation, towards the word (happy) or (Saeed).

A large amount of fully vocalized Arabic text that can be processed should be stored in a language-text bank in the computer's memory.

The history of Arabic literature from the pre-Islamic era to the current era should be digitized, as should the digitization of linguistic sciences.

## References:


- Bar-Hillel. (1964). *Language and information.* Massachusetts: Addison-Wesley.
- Gordin, M. D. (2015). *Scientific Babel: How Science Was Done Before and After Global English.* Chicago: University of Chicago Press.
- Grishman, R. (1986). *Computational Linguistics: An Introduction.* Cambridge: Cambridge University.
- Hutchins, J. (2000). *Warren Weaver and the Launching of MT, Early Years in Machine Translation* . Amsterdam: Brief Biographical Note.
- Lyons, J. (1991). *Natural Language and Universal Grammar. New York: Cambridge University Press. pp. 68–70.* New York: Cambridge University Press.
- Madsen, M. W. (2009). *The Limits of Machine Translation (Thesis). University of Copenhagen. p. 11*. Copenhagen: University of Copenhagen.
- Martin, V. (2015, November 22). *adweb aclwiki*. Consulté le December 2020, 31, sur Frequently_asked_questions_about_Computational_Linguistics: https://aclweb.org/aclwiki/Frequently_asked_questions_about_Computational_Linguistics
- Tim, B. (2020, August 17). *ACL Member Portal*. Consulté le December 20, 2020, sur he Association for Computational Linguistics Member Portal: http://www.aclweb.org
- Weaver, W. (1955). *Machine Translation of Languages, Fourteen Essays* . Cambridge: Massachusetts Institute of Technology.