RESEARCH TITLE

THE CONCEPT OF COMPUTATIONAL LINGUISTICS

Zainab Abd Al-Razaq M. Al-Asfoor[1], Weaam Hussain Ali[2]

[1] Master Degree in Linguistics, A Lecturer in Karbala University/ College of Education, Department of English language, Zainab.a@s.uokerbala.edu.iq

[2] Bachelor Degree in Computer Sciences, A Programmer in Karbala University/ College of Education for Pure Sciences, Weaamhussein30@gmail.com

Abstract

The field of computational linguistics has developed in recent years. It has become an essential area of industrial growth. Computational linguistics involves both sides engineering and scientific. The engineering side is regularly named natural language processing. This component includes creating computer programs that perform beneficial language-related tasks, such as machine translation, question-answering, and summarization. Natural language processing, like many engineering fields, is based on the findings of numerous scientific sides (Johnson M. Oct. 2011). This study tends to give a brief idea about Computational Linguistics specifically its definitions, development, and applications.

1- Introduction

Most readers may wonder about the reason behind making a study in computational linguistics (Sometimes the writer uses the abbreviation CL). The researcher of this study seeks to dig deeper into this field of linguistics, especially since it is a modern field. Our imagination is captured by the success of modern natural language processing software. Millions of copies of various programs, including those for grammar and orthography correction, translation from one natural language into another, and information recovery from document databases, are being sold all over the world. Despite that, we should admit that these programs "still lack real intelligence" (Bolshakov I. and Gelbukh A., 2004, p.5). Although there have been attempts to solve this issue for nearly 50 years, till now there is an ambitious goal to develop software for language production and understanding. Such development would offer tools for sufficient automatic translation and man-machine communication in unlimited natural language. Bolshakov I. and Gelbukh A. explain that developers of novel software need to apply the techniques and findings of essential science, thus linguistics rather than other solutions are appropriate to solve problems of translation. In other words, the problems of translation and automatic text understanding will not be resolved by increasing computer speed, improving programming tools, or developing a large number of toy systems for language.

2- The Meaning of Computational Linguistics

The scientific investigation of human language from a computational perspective is known as computational linguistics. Grishman R. defines CL as "the study of computer systems for understanding and generating natural language" (1999, p.4). Richards and Schmidt, in their dictionary (Longman Dictionary, 2002), describe CL as "the scientific study of language from a computational perspective". Bussman H. defines CL, in his dictionary (Rutledge Dictionary of Language and Linguistics, 2006) as a field of study that combines linguistics and (applied) computer science that is focused on how computers understand natural languages "on all levels of linguistic description" (ibid). Gelbukh and Balshakov (2004, p.25) consider CL as a "synonym for automatic processing of natural language". They indicate that its main objective is "the construction of computer programs to process words and texts in natural language" (ibid). Trask, R.L. describes CL as "the use of computers to perform various tasks involving language" (key Concepts in language and linguistics, 1999). He adds that digital computers have made approaches to solving the practical and descriptive problems of language that were previously addressed insufficiently. Mitkov R. states that CL is "an interdisciplinary field concerned with the processing of language by computers" (2003, p.ix). He continues by saying that CL has advanced conceptually and grown exponentially due to the creation of formal and computational models of language (ibid).

3- The Development of Computational Linguistics

Through the 1950s, efforts were made in the United States to employ computers to mechanically translate articles from foreign languages into English, precisely Russian scientific magazines (Hutchins J., Sep. 1999). It was believed that computers would soon be able to understand language as they are far faster and more accurate than

humans at arithmetic (systematic) computations (Barach A. B., 1975). It found that earlier techniques such as glottochronology and lexicostatistics are inaccurate and premature. Thus, recent studies that use biological study concepts, particularly gene mapping, have demonstrated to yield more advanced analytical techniques, besides trustworthy outcomes (T. Crowley and C. Bowern). Artificial processing of human languages was realized to be much more sophisticated than had been initially believed when machine translational initially failed to provide correct translations. The new area of research focused on creating software and algorithms for intelligently analyzing language data, which was given the name CL. David Hays is a member of the "International Committee on Computational Linguistics" (ICCL), and the "Association for Computational Linguistics" (ACL). He is the first scholar who used the phrase "computational linguistics". Wolfgang Saxon wrote an article on (July 28, 1995) published in The New York Times under the title "David G. Hays,66, a Developer Of Language Study by Computer". The article covered several topics about David G. Hays. Saxon W. mentions that David G. Hays is a social scientist, and was a pioneer in the field of computational linguistics. The article indicated that Similar to artificial intelligence, computer-aided linguistics conducts language research by using computer systems, mathematical models, and computational techniques known as algorithms (Wolfgang Saxon, 1995). It is used in "translation, documentation, and lexicography and in the study of stylistics and content analysis" (ibid).

To translate from one language into another, one has to understand the system of the target language, including grammar, syntax, morphology, semantics, lexicon, and even something of pragmatics. As a result, what began as an attempt to translate across languages turned into a field of study devoted to learning and understanding how to represent and process natural languages by using computers. (Liz Liddy etal)

4- The Uses of Computational Linguistics

Trask R. L. (in his Key Concepts IN Language and Linguistics,1999) explains that there there are several tasks for computers in linguistics. The first one is saving a corpus of written and spoken texts, and the second one is the process of creating concordances. **The Machine-Readable corpus** can be used to get information about the frequency of certain constructions, forms, or words. Thus, the users can get factual statistics regarding actual language use that otherwise would not be available.
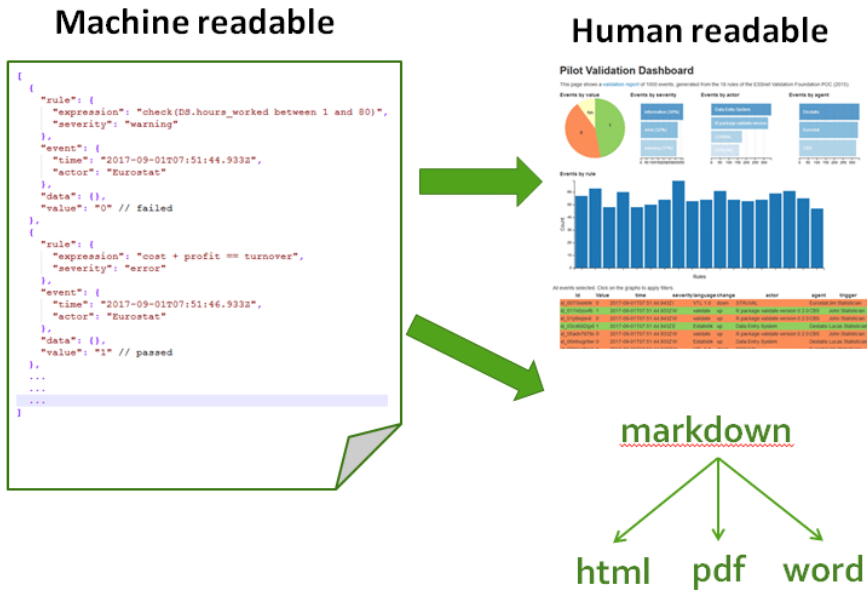
Figure (1): Sample for Machine-readable corpus

On another hand, **Concordance i**s the listing of words found in the body of writing. It provides researchers with an easy way to find all the passages related to the themes they are interested in.

Figure (2): Sample for Concordance

Trask R.L argues that the combination of these two approaches is sometimes useful to identify an author from another by using the statistical examination (1999, p.33). Such examination may identify a style of one author from other, by using words or forms which match a specific style.

There are other applications of CL, which are man-machine interfaces, information retrieval, and machine translation. Trask R. L. defines **machine translation** as "the development of computer programs which can take a text written in one language and convert it into a different language" (ibid). The work on "machine translation" started in the late 1950s with some difficulties occupied. The appearance of some difficulties and problems with machine translation stimulated research in computational linguistics and linguistics. Grishman R. (1999, p.4) states that "extensive work was done in early 1960, but a lack of success, and in particular a realization that fully-automatic high-quality translation would not be possible without fundamental work on text's understanding, led to a cutback in funding". Besides, very

few CL projects in the United States are focused on machine translation, in contrast with Japan and Europe.
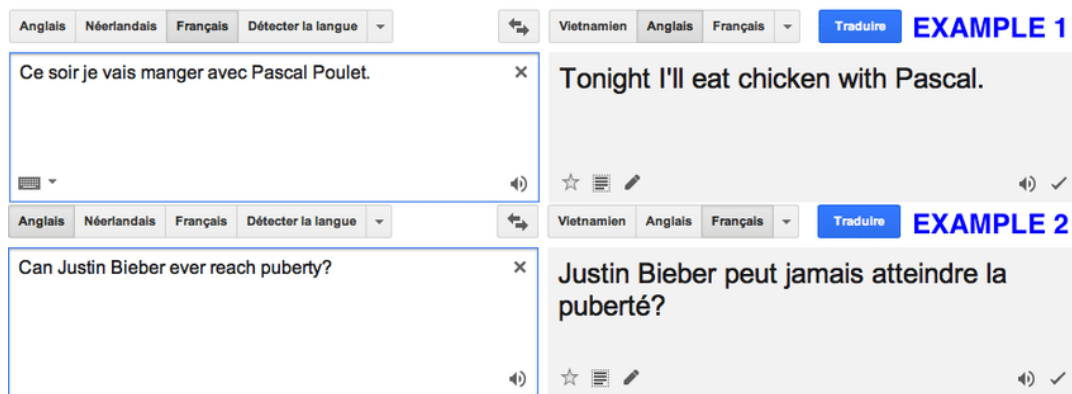

<div align="center">Figure (3): Sample for Machine Translation</div>

Information Retrieval is another application mentioned by Grishman R. Grishman states that "the system was to extract the relevant text from a corpus and either display the text or use the text to answer the query directly" (Grishman R.; 1999, p.4). He mentions the reason behind using this system which Because of the complexity of the text in most domains of interest whether scientific or technical reports.
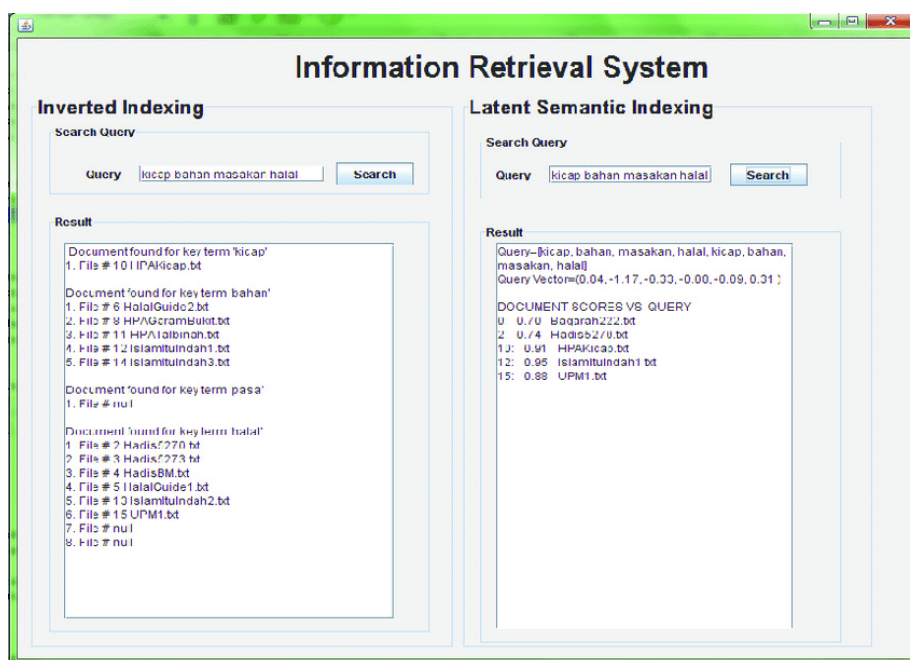

<div align="center">Figure (4): Sample for Information Retrieval</div>

Grishman R. mentions a third application which is **Man-machine** interfaces. He differentiates it from the previous applications, by mentioning its advantages which are:

> "*First the input to such systems is typically simpler (both syntactically and semantically) than the texts to be processed for machine translation or information retrieval. Second, the interactive nature of the application allows the system to be useable even if it occasionally rejects an input*"

(Grishman R.,1999, p.5)

Hausser R. mentions Man-machine communication as a type of machine which provides "a keyboard for language input and a screen for language output" (1994, p. 15-16). In other words, this machine permits the in-put and out-put for random or arbitrary language. He refers to two types: restricted and non-restricted. He explains that the non-restricted may be considered as the minimum standard for successful computational linguistics" (Hausser R;1994, p. 15-16). The restricted form refers to the communication between the user and a computer, such as a workstation or a PC.

Speech Synthesis is another application for computational linguistics, which means "converting written input into an intelligible imitation of human speech". One of the speech synthesis techniques is "providing a 'voice' to disabled people who cannot speak" (Key Concepts in Language and Linguistics; 1999, p.33)
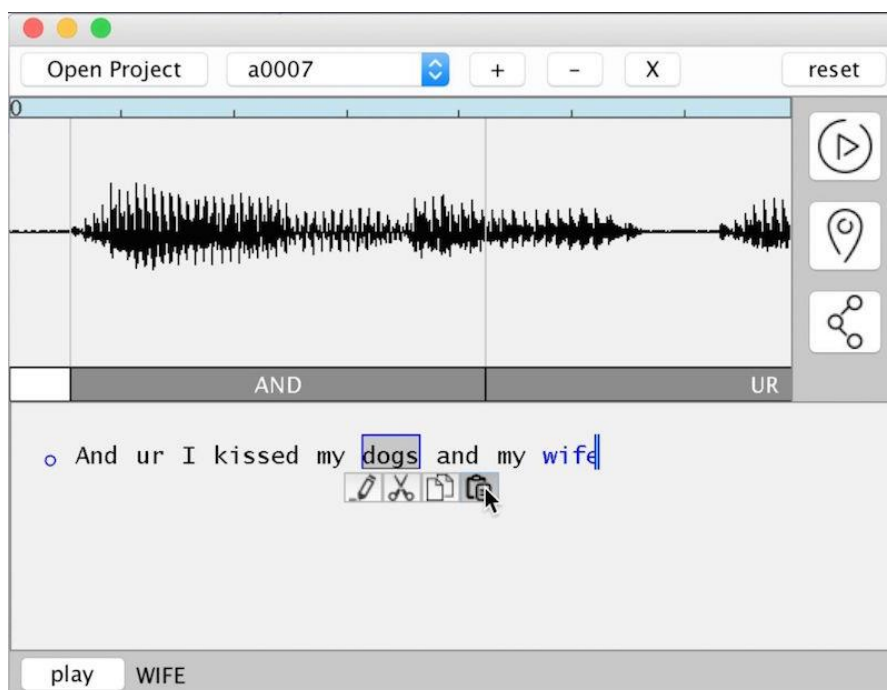

Figure (5): Sample for Speech Synthesis

Trask R.L. refers to another application which is **Computer-Assisted language**. The computer-assisted language indicates that a student learns a foreign language mostly through interaction with computer software that sets tasks, assigns grades, and modifies its behavior based on the learner's level of performance.

Figure (6): Sample for Computer-Assisted language.

Thus, the majority of effort in computational linguistics is focused on natural language processing, which involves creating systems that can accept typed input (and occasionally, speech as well), interpret it, and give the appropriate response.

Conclusion

Nowadays, it's important to concentrate on the computational linguistics field, since the technology is in rapid development. From the previous uses of computational linguistics, we notice it is an important field to the users. Lerner of the foreign language may depend on translation machines heavily to develop their target language. Besides that, the translator may have linguistics knowledge from the scientific side of the target language. Such knowledge involves phonology, vocabulary, syntax, semantics, and pragmatics. Sound recordings programs are important in developing pronunciations for foreign language learners. Recently, some programs appeared for the syntactic side, such are Grammarly, which is used to correct syntactic errors in writing the English language.

References

Bolshakov, I. A., & Gelbukh, A. (2004). *Computational linguistics models, resources, applications*. Ciencia de la computación.

Bussmann, H., Kazzazi, K., & Trauth, G. (2006). *Routledge dictionary of language and linguistics*. Routledge.

Clark, A., Fox, C., & Lappin, S. (Eds.). (2012). *The handbook of computational linguistics and natural language processing* (Vol. 118). John Wiley & Sons.

Grishman, R. (1999). *Computational linguistics: an introduction*. Cambridge University Press.

Institut für Germanistik. Pp. xii+ 534. Journal of Linguistics, 37(3), 593-625.

Mitkov, R. (Ed.). (2022). *The Oxford handbook of computational linguistics*. Oxford University Press.

Retrospect and prospect in computer-based translation, by John Hutchins, Sep. 1999 (University of East Anglia, Norwich UK)

Richards, J. C., & Schmidt, R. W. (2002). *Longman dictionary of language teaching and applied linguistics*. Routledge.

Roland Hausser, (1994). Foundations of computational linguistics: man-machine communication in natural language. Friedrich-Alexander-Universität Erlangen-Nürnberg

Saxon, W. (1995). David G. Hays, 66, a Developer Of Language Study by Computer. *The New York Times*. (Available at: https://aclanthology.org/www.mt-archive.info/NYT-1995-Saxon.pdf)

T. Crowley., C. Bowern. An Introduction to Historical Linguistics. Auckland, N.Z.: Oxford UP, 1992. Print.

Trask, R. L. (1999). *Key concepts in language and linguistics*. Psychology Press.