

An Introduction to Computational Linguistics Advantages&Disadvantages

Eman M. Muslim
University of Baghdad
College of Education for Women

Many decades passed when the electronic age first began. We have to quicken steps forward before being so far from it.

Abstract

The current study tends to show the growth of computational linguistics and the motivations of its existence, the field which it refers to, its theoretical and applied components, its relation to natural language understanding, advantages and disadvantages of natural language processing, some of the most important applications, with reference to its role in teaching and learning processes and the problems facing them.

Introduction

Computational linguistics is a field ranges between linguistics and computer science with the help of psychology and logic. It uses computers to simplify treating with linguistic issues. It is considered as a branch of computer science as well as of linguistics. However; it should be cooperation between computer science and linguistics. (McGuigan, 2006)

The aim of the study is to highlight on the field of computational linguistics and its skills concentrating on linguistics with the reference to computer uses concerning teaching and the role of today teachers towards the students of the electronic age.

The study is supposed to be very important for both teachers and students who are interested in language studies. And it is also important to draw attention to the danger of being electronically dead. It is an

invitation to start discussion with the specialists of computer science in order to get the utmost benefit concerning linguistic subjects.

The growth of C. L.:

Martin Kay (2003) says that computational linguistics perhaps first began in 1949 concerning machine translation. The first conference on machine translation was in 1952. The first journal of "Mechanical Translation" was in 1954. But; the phrase "computational linguistics" began to be used in 1965 when it appeared as a sub-title of the journal "Mechanical Translation and Computational Linguistics". It was written in very small type. In 1974, the name of the journal "Mechanical Translation and Computational Linguistics" was changed into "The American Journal of Computational Linguistics". And in 1980, it became "Computational Linguistics" which is still alive.

Two essential motivations were beyond the activities of computational linguistics:

1. Theoretical: It came from the thinking that adopting computational aims would cause important progress in linguistics.
2. Technological: It came from the desire to produce a technology to serve the practical needs for translation, information extraction, grammar checking---etc.

None of these ventures can be achieved by the adoption of linguistic methods alone. (ibid)

What does the field of C. L. refer to?

Computational linguistics which is still considered a new field connects linguistics with computer science. Researchers of C.L. are divided into two teams: One of them apply their experience in computer science to linguistics showing people what they should know in order to understand a natural language, how they find this knowledge and how they should use this knowledge. The second apply their experience in

linguistics to computer science so that computers can understand everyday human language and translation. They also deal with computers linguistically under the name of natural language processing NLP. (Wintner, 2004)

In general; C.L. could be seen as a synonym of automatic processing of natural languages which is concerned with constructing computer programs to process words and texts in natural languages. Yet; the lack of programming knowledge is still a problem. Professional linguists cannot explain the data structures which the program should use. Specialists in C.L. with a good background in linguistics can productively work in different tasks such as computing tables and dictionaries or working in an interdisciplinary team or finding out any other new idea or approach which can be very useful to understand the literature on the subject. (Bolshakov, 2004)

Components of C. L.:

Computational linguistics is ranging between cognitive sciences and artificial intelligence. It has theoretical and applied components. The theoretical component depends on theoretical linguistics and cognitive science. Since the linguistic theories have nowadays become so complex, linguists decided to simplify them by employing computers. Thus; they began to cooperate with the computational programmers with the aid of cognitive psychology in order to develop computational models for the formal linguistic theories. The applied component develops practical models of human language depending on artificial intelligence. This is also to be put under the terms "Language Engineering" or "(Human) Language Technology". The function of applied computational linguistics is to find a program which can improve the interaction between human and machine so that human and computers can communicate easily. (Uszkoreit, 2000)

Natural Language Understanding:

Linguists complain against lacking in artificial intelligence that they are forced to search a replacement which is known as "Natural Language Processing". And since natural language understanding is a synonym to computational linguistics, it is important to know its essential aims. It is to:

- generate and produce contents in any natural language in any domain,
- support multilingual services.

There are two lines of research concerning natural language understanding. One is towards text-based applications and the other is towards dialogue-based applications. Text-based applications imply the processing of all kinds of written texts such as books, articles, messages, magazines---etc. in such a way that they can be easily read by users. Thus; researchers are continuously developing different means of information access to such texts as:

- finding relevant documents on the desired subject from a database of texts, e. g. finding relevant books in a library,
- taking information out of some kind of text on a certain topic, e. g. building a database of all adequate on-line information written in the news on a certain day,
- translating documents from one language to another.
- changing long texts into short summaries.

The following are the applications of dialogue-based ones:

- Question-answering system,
- Automated customer service over the telephone,
- Tutoring systems in which the student can interact with the machine. (Bolshakov; 2004)

Natural language processing and drawbacks:

One of the most important problems that investigators suffer is the misunderstanding of machines for natural languages. Natural language processing is complex because there will be misunderstanding in the application of computational programs concerning:

1. **phonology and phonetics** which is concerned with pronunciation. The problem of computational programs concerning this field is that some words have the same pronunciation with different meanings such as "weak" and "week". Computers cannot differentiate between the two words.

2. **morphology** which is concerned with the inner structure of words in their written (graphemic) form and spoken (phonemic) form. It has two essential functions:

a. Inflection: It is related to the grammatical function of words of the same part of speech; e. g. the paradigm of the verb

- play for the present simple; 1st and 2nd persons,
- plays for the present simple; 3rd person,
- played for the past simple and the past participle,
- playing for the present participle.

b. Derivation: It is related to the production of new words of different parts of speech; e. g.

- nation ----- (a noun)
- national----- (an adjective)
- nationalize----- (a verb)

A morphological analyzer should be intelligent enough to know and extract the base forms from inserted documents in computers. The applications which are achieved in this respect are:

- a: hyphenation (segmenting words into their morphs),
- b: spelling correction,

c: stemming which reduces the related words as possible.

The problem of such computational programs is the input which should be very broad. Other forms of application are parsing and generating natural language utterances in written or spoken form and machine translation. (Troost, 2006)

3. **syntax** which is concerned with the structure of sentences. Sometimes; word order of some kinds of structure causes misleading such as the following examples:

(1) I saw her *with* a telescope.

The word "*with*" may either be an adjunct with the verb "*saw*" or with the pronoun "*her*".

(2) The article covers the rights of women *and* childhood.

The conjunction "*and*" is understood as conjoining the two nouns "*women*" and "*childhood*". But; it may also be understood as conjoining the phrase "*the rights of women*" with the noun "*childhood*", and this is incorrect. This is considered as one of the problems faced in translation by computer.

4. **semantics** which deals with the meanings of words, phrases and sentences. But; since a word may have several meanings like the word "*covers*" which means "*to hide*", "*to spread over*" or "*to deal with*", it would also be a problem in translation by computer.

5. **pragmatics** which deals with the meanings of utterance depending on the context. Many times the meanings of the sentence words are clear but the interpretation depends on its context. For example;

(3) We are waiting.

The sentence may bear any one of the different kinds of interpretation according to its context: a. an ordinary fact, b. a promise and c. a threat.

Computational translation cannot distinguish between those kinds of interpretation which describe ironic (speaking the opposite of intention)

or metaphoric (showing expressions which are literally not exist) phenomena as in the following examples;

(4) You are clever enough to achieve this. (Ironic)

(5) He looks wood-minded. (Metaphoric)

The problems mentioned above show they can neither be solved by computer science alone nor by linguistics alone. (Wintner, 2004)

Some of the Most Important Applications of C. L.:

1. Automatic hyphenation: McIntosh (1990) defines the hyphen: " it is that small horizontal bar which is used either to join two elements of compound words (the link-hyphen), or to signal that a word is being split at the end of a line of printing (the beak-hyphen)". This program is the system which splits long words properly which cannot fit within the accepted margin of the line. At the beginning; they depended on simple algorithms in splitting long words such as putting a hyphen after the third, fifth or seventh character in any word. But; this resulted in "idiot breaks", e.g. the word "photographic" would be split into either "pho" and "tographic", "photo" and "graphic" or "photogr" and "aphic". In order to improve typing texts; Microsoft Word used the menu item "Hyphenation". Programs of such type need linguistic information about morphemic structure of words, vowel and consonant letters.

2. Spell checking: It is the process of finding out and correcting error occurred in typing a certain text. Millions of users get benefit from this program. The spell checker points to the errors, then gives some suggested words from which the user can choose the one suits the context e.g. if the word "present" incorrectly written "presen" , the spelling checker would give the following alternatives "present", "preset" and "pressmen" from which the user can choose the suitable one .

3. Grammar checking: It is the process of finding out and correcting grammatical errors taking into consideration either the whole sentence or

the adjacent words such as subject agreement with verbs, adjectives adverbs, prepositions and so forth. Grammar checkers are supposed to be so useful to solve such problems but there are still only simple and commercial ones. A useful grammar checker should bear a complete syntactic analysis (parsing) system of a text in order to be an active assistant to the user.

Grammar checkers have experienced much progress especially the one which is included in Microsoft Word. Yet; it is still not perfect though it is somewhat helpful. However; it is the responsibility of the user to be sure of what he / she is writing because sometimes the grammar checker gives alarms where no error or suggest an unreasonable correction.

4. Style checking: Each literary category has its own style of writing. In official writing; one should choose constructions far from slang language. The style checker provides the user with the correct choice. It also parses the text automatically in order to find out the wrong syntactic constructions.

5. References to words and word combinations: The user can access to a set of words which are semantically related to a certain one. This is achieved by autonomous on-line dictionaries and other which are built in. The user can get benefit from such references to choose the most appropriate word for his / her text.

6. Information retrieval: It is the program which is designed to search for relevant information which might be in various kinds of documents. Different methods of research are used because of the great desire to search for scientific articles which exist in documents. (Wikipedia, 2006)

The Role of C. L. in Language Teaching and Learning Processes:

Computer-assisted language learning is a means of teaching and learning processes. It is referred to by CALL. It has been used for more than forty years. Lee (2000) divided the period of using CALL into three

stages. Each stage shows the level of technology and the pedagogical theories of its age:

- a. behaviorist CALL which first began in the 1960s and 1970s basically depended on repetitive language drills.
- b. communicative CALL which emerged in the 1970s and 1980s, they focused on generating original utterances instead of the old means using repetition drills.
- c. integrative CALL which is the recent stage has moved away from a cognitive view of communicative language teaching to a socio-cognitive view in which a meaningful authentic context of real language use is emphasized . This stage also emphasizes on the integration of the four skills of language learning (listening, speaking, writing, and reading) and the integration of technology.

The following description for CALL is quoted from Wikipedia, (2006): "Typical CALL programs present a stimulus to which the learner must respond. The stimulus may be presented in any combination of text, still images, sound, and motion video. The learner responds by typing at the keyboard, pointing and clicking with the mouse, or speaking into a microphone. The computer offers feedback, indicating whether the learner's response is right or wrong". Such programs can either be gained with CDs or by the internet.

Problems Facing Using Computer –assisted Language Learning:

The problems are categorized as the following as Lee (2000) states:

1. Financial: It is the most important problem especially in the poor countries. In spite of achieving higher results with less time, poor countries cannot supply their schools and universities with adequate number of computers to get the desired benefit.

2. Availability of computer hardware and software: The rapid changes in technology force to choose the best quality of computer hardware and software. This is also difficult for the poor counties.
3. Technical and theoretical knowledge: There is a shortage in technical knowledge that many instructors do not know how to use the current technology. Not only is this but many instructors who are fascinated by the new technology depend basically on technology and overlook the theoretical plans of methodology to integrate the knowledge for their students which may cause negative affect for both the teachers and the learners.
4. Acceptance of technology: Many instructors feel that the new technology threatens their future because it requires constant preparation and contact with the rapid change so that they feel in continuous challenge which needs time and adherence.

Conclusion

Computational linguistics first began in 1949 interesting in translation. It is extended to include various activities that serve computer users in general and those who are interested in language teaching and learning in particular. It works to achieve two purposes: One is to teach language by computers through the internet or CDs. Two is to make programs that linguists can get benefit of like dictionaries and translation and checking programs. This can neither be dealt by the Department of Linguistics alone nor by the Department of Computer alone. There should be a close relationship between the two departments. Though of the great advantages of computational linguistics, there are some disadvantages.

Recommendations

The researcher recommends the following items:

1. Developing the teachers' attitudes towards the usage of computers in the teaching process.
2. Making training sessions for the teachers.
3. Supplying computer labs for the students of the different levels in each department.
4. Choosing suitable CDs for the students of the different levels which suit the integration of each level.
5. Opening discussion with the specialists in order to teach computational linguistics as a subject.

References

1. Bolshakov, Igor & Gelbukh, A.: (2004) Computational Linguistics: Models , Resources, Applications. Mexico, Center for Computing Research.
2. Chapelle, Carol A.: (2001) Computer Applications in Second Language Acquisition: Foundations for teaching, testing and research. Cambridge University Press, Cambridge, UK.
3. ICT4LT Module 1.4 Information and Communications Technology for Language Teachers: (2006) Introduction to Computer-Assisted Language Learning (CALL).
[http://209.85.129.104/search?q=cache:208MFW0DWvcJ:www.ict4lt.org/en/en_mod1-4.htm+3.+ICT4LT+Module+1.4:+Introduction+to+Computer-Assisted+Language+Learning+\(CALL\)&hl=en&ct=clnk&cd=1](http://209.85.129.104/search?q=cache:208MFW0DWvcJ:www.ict4lt.org/en/en_mod1-4.htm+3.+ICT4LT+Module+1.4:+Introduction+to+Computer-Assisted+Language+Learning+(CALL)&hl=en&ct=clnk&cd=1)
4. Kay, Martin: (2003) in Ruslan Mitkov, The Oxford Handbook of Computational Linguistics. U.S.A. Oxford University Press.

5. Lee, Kuang-wu: (2000) English Teachers' Barriers to the Use of Computer-Assisted Language Learning. In *The Internet TESL Journal*, Vol. VI, No. 12, December 2000.
6. McGuigan, Brendan: (2006) What is Computational Linguistics. <http://www.wisegeek.com/what-is-computational-linguistics.htm>
7. McIntosh, Ronald C.: (1990) Hyphenation. <http://www.hyphenologist.co.uk/book/BOOK-ED3.HTM>
8. Trost, Herald: (2006) Computational Morphology. <http://www.ai.univie.ac.at/~harald/handbook.html>
9. Uszkoreit, Hans: (2000) What is Computational Linguistics. http://www.coli.uni-saarland.de/~hansu/what_is_cl.html.
10. Wikipedia, *the free encyclopedia*: (2006) Computer-Assisted Language Learning. http://en.wikipedia.org/wiki/Computer-assisted_language_learning.
11. Wintner, S.: (2004) Hebrew Computational Linguistics: Past and Future. University of Haifa, Haifa.