THE SCIENCE CALLED *COMPUTATIONAL LINGUISTICS*

Cătălin Dehelean

Introduction

The main objective of this article is to provide the reader with some basic idea of Computational Linguistics. It intends to show that in order to describe how natural language works it is necessary to process a huge amount of information, and to do so one needs specialized electronic tools. The ones truly interested in Computational Linguistics must master both these tools and have linguistic training. This seemingly ambivalent situation will prove itself useful in the management of databases, translation software, specialized dictionaries etc.

Definition

Computational Linguistics is a science dedicated to the study of the natural language machine processing. It is a mix of Informatics and Linguistics, yet its beginnings go back well into the nineteen-fifties. During this half a century it has successfully established itself, nationally and internationally, as it has developed, based on Informatics and Linguistics, new methods of studying the spoken and written natural language machine processing.

Purpose

The influence of Computational Linguistics (CL) on our daily life in the "information society" we live is increasing. It is

BDD-A12868 © 2007 Editura Mirton; Editura Amphora
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:54:55 UTC)

almost impossible not to come into contact with the products of this relatively new science, whether surfing the internet or just using a computer. An eight grader writing his essay uses morphological processes (Spelling check), grammatical analysis (Grammar check), maybe some statistical information about the written text (Frequency of occurrence check) or Lexicography (Thesaurus). If a search on the Internet also occurs, the field of methods is enlarged to comprise information indexing and even online translation.

But even if the computer is not being used, one is likely to encounter the applications of Computational Linguistics when reading the computer-translated instructions for the new toaster or when calling the bank and being asked for the account number by a friendly machine.

These examples show how important Computational Linguistics has become during the last few years: it opens up new sources of information, it makes the interaction with machines a lot easier and helps overcome the language barriers.

Classification

The subject matter of Computational Linguistics is the Computer Natural Language Processing (deferral to Programming languages, for instance), as far as the written language (Text) and spoken language (Speech) are concerned. Computational Linguistics is basically and historically a synthesis of Information and Linguistic methods and knowledge.

This characterization is deliberately general, so as to cover a combination of the different perspectives on Computational Linguistics.

Computational Linguistics is a part of Linguistics (like Sociolinguistics, Psycholinguistics, etc.) and, theoretically speaking, is concerned with the computationally relevant aspects

BDD-A12868 © 2007 Editura Mirton; Editura Amphora
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:54:55 UTC)

of the study of language and language processing, independent from its actual realization on the computer. The development of the grammatical formalization is an example of this perspective on Computational Linguistics.

Computational Linguistics is a discipline for the development of linguistically relevant programs and the processing of linguistic data (Linguistic Data Processing). This perspective that can be traced back to the early days of informatics has experienced a Renaissance due to increasingly important empirical research by means of comprehensive language data corpora.

Computational Linguistics is a realization of natural language phenomena on the computer („natural language processing"). The research of many of these phenomena has a long tradition within the Philosophy of Language, especially within the language-oriented formalist logic. Since language as part of a cognitive system can be understood as knowledge of language and as nonlinguistic science, thinking processes and action planning being linked closely together, Artificial Intelligence and Cognitive Science are particularly interested in the research of the modeling of these phenomena. Computational Linguistics is therefore intrinsically linked to the formal and/or cognitive sciences.

Computational Linguistics is a practice-oriented engineeringly conceived development of language software („Language Technology")

This list of different perspectives shows, in principle, the differences in perceiving Computational Linguistics. Computational Linguistics is a sum and synthesis of its different values.

Hereby the four cornerstones of Computational linguistics are laid down: the development of the methods to operationalize the natural language phenomena; the build-up and the

BDD-A12868 © 2007 Editura Mirton; Editura Amphora
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:54:55 UTC)

administration of large re-usable linguistic data corpora, which can be used for empirical development and evaluation purposes; the development of realistic applications, which prove the relevance of Computational Linguistics for modern information technology and, at the same time, which reflects once more its technological evolution; and the creation of real-life evaluation machines which should enable this evolution. To these we add the fact that the specialized basis of Computational Linguistics is anchored partly in what it inherits from its mother discipline and partly in what it borrows form other neighboring disciplines.

Heritage and loans

Computational Linguistics inherits from Linguistics the object of investigation and, at the same time, basic vocabulary of linguistic terms and differentiations. The structuring of the method range of Computational Linguistics is therefore oriented as far as possible towards the established branches of Linguistics: Phonology, Semantics and Pragmatics, which build the structural description of the natural language utterance.

Computational Linguistics is not just a recipient of linguistic theories and issues, but it can also be an incentive for the gain of insight and for the formulation of new appendages within Linguistics.

A successful example of interdisciplinary work between Linguistics and Computational Linguistics is the development of Optimality Theory arguments. Originally generated by the union of Artificial Neural Networks and Universal Grammar, the Optimality Theory with its rule – oriented arguments has in the meantime become a real competing model for the description of phonological issues. Furthermore, it will be increasingly used to describe phenomena on other levels, e.g. Morphology and Syntax.

BDD-A12868 © 2007 Editura Mirton; Editura Amphora
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:54:55 UTC)

The application and evaluation of linguistic theories is another task for Computational Linguistics. Firstly, the application of theories to real, occurring data supplies a picture of its coverage and completeness and can even ensure, to some extent, its use outside strictly theoretically-oriented circles. The factor mentioned here is the implementation of a system of Structural Analysis, which is based on the principle and parameter approaches.

In the end, some branches of Linguistics are more involved than others in the computer processing of their material. Corpus Linguistics, which deals with the research of the linguistic connections by examining Corpora (samples), has been only during the last years able to research a realistically large amount of data with huge coverage (often about one billion words) by using computers.

Informatics shares with Computational Linguistics, in essence, the knowledge about data structures as well as the application of efficient methods. Besides the obvious connection between the research and the realization of the natural language systems and Informatics (Systems Analysis, Modeling, Algorithmics, Implementation Theory), the aspects of Theoretical Informatics (Computability Theory, Computational Complexity Theory, and the field of Formal Language) also play an important role in this matter.

The aspects of the question of how Language, Thought and Action relate to each other derive from Philosophy (especially from the Philosophy of Language and from Logic); Language in itself can't be considered just an independent phenomenon, but as a phenomenon related to non-verbal circumstances, both in the world as such and (in a narrow meaning of the word) in the companionship of a speaker of a language. Formal Logic is a one of the central means of Computational Linguistics for the precise description of Natural Language Phenomena.

BDD-A12868 © 2007 Editura Mirton; Editura Amphora
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:54:55 UTC)

Computational Linguistics owes Artificial Intelligence a series of important procedures (e.g. search technique and planning process). They will be found, for instance, in Speech Recognition, Grammar Analysis and Data Generation.

Computational Linguistics and Artificial Intelligence

A very intricate relationship has been established between Artificial Intelligence and Computational Linguistics. It is safe to say that certain subsystems which make up systems can be recognized as parts of Artificial intelligence. Certain methods and ideas can be found in the theories which constitute the basis of Computational Linguistics. On the other hand, certain linguistic software systems make good use of a whole array of methods which are the indisputable domain of Artificial Intelligence, as part of Computer Science. A computational linguist might easily be found working with complex structures, algorithms, matrices, etc.

Such a situation has generated the idea that Computational Linguistics is not autonomous but a part of Artificial Intelligence. This conception, although based on the very close relationship between the two sciences is, as on might say, a misconception, for just like Artificial Intelligence, Computational Linguistics has, in fact, its own systems of theories, highly distinct purposes and scopes, its own methods of investigation, etc., which precisely position it among the other sciences. Moreover, the application of systems belonging strictly to Artificial Intelligence in order to process natural language has not yielded desirable results.

It seems that only a continuation of the collaboration between these two different sciences will ensure of their evolutions. Their relationship of good vicinity basically implies understanding each other methods and using them

BDD-A12868 © 2007 Editura Mirton; Editura Amphora
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:54:55 UTC)

consciously whenever needed, rather than spending a lot of time and energy in developing a parallel version of what already exists. The readers are thus encouraged to study both Computational Linguistics and Artificial Intelligence.

Conclusion

This undertaking is not intended to be more than a pinpoint towards Computational Linguistics, as the latter is truly vast. This part of linguistics has had the unfortunate faith that it was studied more by people from the field of computer science rather than by linguists. An impediment in this direction was the fact that basic knowledge of programming is required in order to study it. Therefore it is only relatively recently that linguists have started to interact with programmers, in turn being themselves recognized by the programmers as their much-needed partners in developing this field.

References

1. Fellbaum, Christiane (editor). **Title:** *Wordnet: An Electronic Lexical Database*. **Publication Details:** MIT Press. **Publication Year:** 1998

2. McEnery, Tony and Wilson, Andrew. **Title:** *Corpus Linguistics*. **Publication Details:** Edinburgh University Press. **Publication Year:** 1996.

3. Mitchell, Tom M. **Title:** *Machine Learning.* **Publication Details:** McGraw-Hill. **Publication Year:** 1997

4. Pustejovsky, James. **Title:** *The Generative Lexicon*. **Publication Details:** Bradford Books. **Publication Year:** 1998

5. Radford, Andrew. **Title:** *Linguistics: An Introduction*. **Publication Details:** Cambridge University Press. **Publication Year:** 1999

6. Russell, Stuart J. and Norvig, Peter. **Title:** *Artificial Intelligence: A Modern Approach*. **Publication Details:** Prentice Hall. **Publication Year:** 1994

7. Wang, Paul P. **Title:** *Computing with Words*. **Publication Details:** John Wiley and Sons. **Publication Year:** 2001

BDD-A12868 © 2007 Editura Mirton; Editura Amphora
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:54:55 UTC)

Ştiinţa numită *Lingvistică Computaţională*

Rezumat

Lingvistica computaţională s-a format în secolului al XX-lea, din lingvistică şi informatică. Obiectul Lingvisticii computaţionale este studiul limbajului natural.

Aplicaţii ale acesteia sunt corectorul ortografic, corectorul gramatical, dicţionarul electronic, traucerea on-line.

Relaţia dintre Lingvistica computaţională şi Inteligenţa artificială este complexă. Studiul Lingvisticii computaţionale necesită cunoştinţe de informatică.

BDD-A12868 © 2007 Editura Mirton; Editura Amphora
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:54:55 UTC)