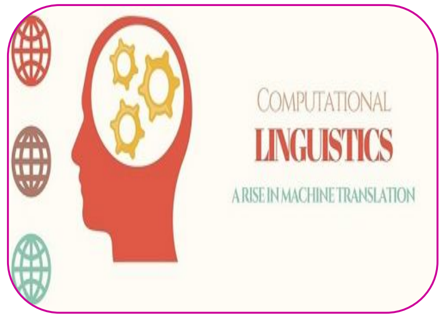
ISSN: 2249-894X IMPACT FACTOR : 3.8014 (UIF)

REVIEW OF RESEARCH

VOLUME - 6 | ISSUE - 9 | JUNE - 2017

COMPUTATIONAL LINGUISTICS: AN OVERVIEW

Dr. N. N. Londhe
Associate Professor and Head , Postgraduate
Department of English and Research Center
DBF Dayanand College of Arts & Science, Solapur .



ABSTRACT

Computational linguistics, a branch of linguistics, is often a combination of studies in computer science and programming, particularly statistics, language structures, and natural language processing. Combined together, these fields most often lead to the development of systems that can recognize speech and perform some task based on that speech. Examples include speech recognition software, such as Apple's Siri feature, spell check tools, speech synthesis programs, which are often used to demonstrate pronunciation or help the disabled, and machine translation programs and websites, such as Google Translate.

KEYWORDS: *computer, natural language, processing etc.*

1. INTRODUCTION

Computational Linguistics, a recently developed interdisciplinary branch of linguistics, deals with the interface between natural languages and the use of computers in processing them. Originated in 1950s in the United States of America, it focuses its attention on the use of computers to automatically translate texts from foreign languages into English. This branch of linguistics deals with mainly the processing of natural languages. The term *Computational Linguistics* is coined by David Hays. It is assumed that since computers can make arithmetic calculations faster and more accurate than humans, they can used for processing the natural languages.

Natural language processing is concerned with the use of computers for processing natural language text or speech. Machine translation (the automatic translation of text or speech from one language to another) began with the very earliest computers. Natural language allows computers to interact with humans using natural language, for example, to query databases. Coupled with speech recognition and speech synthesis, these capabilities will become more important with the growing popularity of portable computers that lack keyboards and large display screens. Other applications include spell and grammar checking and document summarization. Applications outside of natural language include compilers, which translate source code into lower-level machine code, and computer vision.

Most natural language processing systems are based on formal grammars. The development and study of formal grammars is known as computational linguistics. A *grammar* is a description of a language; usually it identifies the sentences of the language and provides descriptions of them, for example, by defining the phrases of a sentence, their interrelationships, and perhaps also aspects of their meanings. *Parsing* is the process of recovering a sentence's description from its words, while *generation* is the process of translating a meaning or some other part of a sentence's description into a grammatical or well-formed sentence. Parsing and generation are major research topics in their own right. Evidently, human use of language involves some kind of parsing and generation process, as do many natural language processing applications. For example, a machine translation program may parse an input language sentence into a (partial) representation of its meaning, and then generate an output language sentence from that representation.

Noam Chomsky may called the precursor of Modern computational linguistics because it began with Chomsky (1957), and was initially dominated by the study of his 'transformational' grammars. These grammars involved two levels of analyses, a 'deep structure' meant to capture more-or-less simply the meaning of a sentence, and a 'surface structure' which reflects the actual way in which the sentence was constructed. The deep structure is a clause in the active voice, 'Sandy saw Sam,' whereas the surface structure involves the more complex passive voice, 'Sam was seen by Sandy.'

The Transformational Generative (TG) grammars are computationally complex, and in the 1980s several linguists came to the conclusion that much simpler kinds of grammars could describe most syntactic phenomena, developing Generalized Phrase-Structure Grammars and Unification-based Grammars. These grammars generate surface structures directly; there is no separate deep structure and therefore no transformations. These kinds of grammars can provide very detailed syntactic and semantic analyses of sentences, but even today there are no comprehensive grammars of this kind that fully accommodate English or any other natural language.

Natural language processing using handcrafted grammars suffers from two major drawbacks. First, the syntactic coverage offered by any available grammar is incomplete, reflecting both our lack of understanding of even relatively frequently occuring syntactic constructions and the organizational difficulty of manually constructing any artifact as complex as a grammar of a natural language. Second, such grammars almost always permit a large number of *spurious ambiguities*, that is, parses which are permitted by the rules of syntax but have unusual or unlikely semantic interpretations. For example, in the sentence 'I saw the boat with the telescope,' the prepositional phrase 'with the telescope' is most easily interpreted as the instrument used in seeing, while in 'I saw the policeman with the rifle,' the prepositional phrase usually receives a different interpretation in which the policeman has the rifle. Note that the corresponding alternative interpretation is marginally accessible for each of these sentences: in the first sentence one can imagine that the telescope is on the boat, and in the second, that the rifle (say, with a viewing scope) was used to view the policeman.

In effect, there is a dilemma of coverage. A grammar rich enough to accommodate natural language, including rare and sometimes even 'ungrammatical' constructions, fails to distinguish natural from unnatural interpretations. But a grammar sufficiently restricted so as to exclude what is unnatural fails to accommodate the scope of real language. These observations led, in the 1980s, to a growing interest in stochastic approaches to natural language, particularly to speech. Stochastic grammars became the basis of speech recognition systems by outperforming the best of the systems based on deterministic handcrafted grammars. Largely inspired by these successes, computational linguists began applying stochastic approaches to other natural language processing applications. Usually, the architecture of such a stochastic model is specified manually, while the model's parameters are estimated from a *training corpus*, that is, a large representative sample of sentences.

As explained in the body of this article, stochastic approaches replace the binary distinctions (grammatical vs. ungrammatical) of nonstochastic approaches with probability distributions. This provides a way of dealing with the two drawbacks of nonstochastic approaches. Ill-formed alternatives can be characterized as extremely low probability rather than ruled out as impossible, so even ungrammatical strings can be provided with an interpretation. Similarly, a stochastic model of possible interpretations of a sentence provides a method for distinguishing more plausible interpretations from less plausible ones.

REFERENCES

- Clark, A., C. Fox, and S. Lappin (eds), 2010, *The Handbook of Computational Linguistics and Natural Language Processing*, Chichester, UK: Wiley Blackwell.
- Clarke, Daoud, 2012, "A context-theoretic framework for compositionality in distributional semantics," *Computational Linguistics*, 38(1): 41–71.