

Computational Linguistics

Matthew N. O. Sadiku¹, Chandra M. M. Kotteti,² and Janet O. Sadiku³

¹Department of Electrical & Computer Engineering

Prairie View A&M University, Prairie View, TX USA

²School of Computer Science and Information Systems

Northwest Missouri State University, Maryville, MO 64468

³Juliana King University, Houston, TX, USA

ABSTRACT

Linguistics is concerned with rules that are followed by languages as a system. Computational linguistics (CL) combines the power of machine learning and human language. As a subfield of linguistics, CL is concerned with the computational description of rules that languages follow. It is what powers anything in a machine or device that has to do with language—speaking, writing, reading, and listening. It is often linked with natural language processing (NLP), which is the use of computers to identify structures in natural language. The boundary between NLP and CL is not so clear-cut. This paper is a primer on computational linguistics.

Key Words: Computation, Computational linguistics, Linguistics.

1. INTRODUCTION

Everywhere, every day, everybody uses language. There is no human society which does not employ a language. Computational linguistics (CL) is based on the idea that machines can be computed to understand, learn, or output languages. It is often linked with natural language processing (NLP) and natural language generation (NLG), which are its subsets of. Natural language processing is the application of processing language that enables a computer program to understand human language as it is written or spoken. The NLP engine is programmed to naturally interpret written commands using simple rules. NLG is the use of computers to generate language. Computers that are linguistically competent help facilitate human interaction with machines and software.

Computational linguistics first emerged in the 1950s as machine translation, with the United States military attempted to develop programs meant to automatically translate Russian text into English and Chinese to English using computers. The idea was that computers could make systematic calculations faster and more accurately than a person. Since its inception in the 1950s, computational linguistics has gone through several iterations. Research on computational linguistics has benefited greatly from the work on artificial intelligence. It is the process of deciphering what customers are asking and prompting AI to deliver accurate responses to their questions based on internal data. Modern CL relies on a variety of tools including AI, ML, deep learning, and cognitive computing.

Computational linguistics is an interdisciplinary field that applies computer science (and the use of algorithms) to analyze and comprehend written and spoken language. As shown in Figure 1, computational linguistics combines

linguistics, computer science, artificial intelligence (AI), mathematics, psychology, engineering, cognitive science, and philosophy [1].

2. BECOMING A COMPUTATIONAL LINGUIST

Computational linguists transform unstructured and structured data into something useful. To get started in the journey of becoming a computational linguist, you need to get a degree in computer science, computer engineering, linguistics, or a related field. This can be a bachelor's degree, a master's degree, a doctorate degrees, or an associate degree. You should build computer science skills (such as coding, data structures, databases, data visualization, and machine learning) before studying linguistics. In order to program the algorithms used in computational linguistics, you need to learn a programming language. Python is a good one to start with because it is one of the most commonly used. You should be able to design and create your own applications for retrieving, processing, and visualizing data. Learn to develop programs to gather, clean, analyze, and visualize data using Python. It is also helpful to develop your skills in math and statistics. Once you feel comfortable with your skills and knowledge of computational linguistics, you are ready to apply for job positions in the field. Unfortunately, only big tech companies like Amazon, Apple, and Grammarly can hire computational linguistics [2].

Common business goals of computational linguistics include the following [3]:

- Create grammatical and semantic frameworks for characterizing languages.
- Translate text from one language to another.
- Offer text and information retrieval that relates to a specific topic.
- Analyze text or spoken language for context, sentiment or other affective qualities.
- Answer questions, including those that require inference and descriptive or discursive answers.
- Summarize text.
- Build dialogue agents capable of completing complex tasks such as making a purchase, planning a trip or scheduling maintenance.
- Create chatbots capable of passing the Turing Test.
- Explore and identify the learning characteristics and processing techniques that constitute both the statistical and structural elements of a language.

3. APPLICATIONS OF COMPUTATIONAL LINGUISTICS

There are many applications of computational linguistics in the real world. Typical areas of application include the following [2,4]:

- **Machine translation:** Using AI to translate from one language to another, such as from Chinese to English. Google Translate is a good example.
- **Chatbots:** Software programs that simulate human conversation via spoken or written language, usually for customer service purposes. Current chatbots are the descendants of Weizenbaum's ELIZA and are typically used for entertainment. Many companies, such as Amazon and Verizon, have live chat available alongside

phone and email options. Companionable dialogue agents have so far relied rather heavily on chatbot techniques, i.e., authored input patterns and corresponding outputs.

- **Knowledge extraction:** Creating knowledge from unstructured and structured text sources. An example is Wikipedia, which is the product of random editors.
- **Natural language interface:** These types of tools allow humans to interact with our devices' operating systems using spoken words. Examples include Siri and Alexa.
- **Sentiment analysis:** This refers to the detection of positive or negative attitudes on the part of authors of articles or blogs towards commercial products, films, organizations, persons, ideologies, etc. This is a type of NLP that identifies emotional tone in text or spoken language. It has become a very active area of applied computational linguistics, because of its potential importance for product marketing and ranking, social network analysis, political and intelligence analysis, classification of personality types or disorders based on writing samples, etc. Grammarly is an example of sentiment analysis.
- **Text mining:** This is the process of extracting useful information from massive amounts of unstructured textual data.

4. BENEFITS

Computational linguistics is the study of the computational processing, understanding, and generation of human languages. Work in computational linguistics, often emboldened by the increasing power and availability of general-purpose computers, is often motivated from a scientific perspective to provide a computational explanation for a particular linguistic phenomenon. Computational linguistics is very important because human knowledge is expressed in language. It enables progress in fields such as customer service, scientific research, AI tools, etc. Just as language is a mirror of mind, a computational understanding of language provides insight into thinking and intelligence.

Robots have been used to test linguistic theories. They are beginning to be equipped with web services, question answering abilities, chatbot techniques, tutoring functions, and so on. The keen public interest in intelligent robots and their enormous economic potential (for household help, eldercare, medicine, education, entertainment, agriculture, industry, search and rescue, military missions, space exploration, etc.) will surely continue to energize the drive towards greater robotic intelligence and linguistic competence [4].

5. CHALLENGES

AI leadership had in recent years become a two-horse race between China and the US. The publish or perish rat race is more intense than ever before. There has been a trend for publications to report better and better numbers, but less and less insight. Priorities are different in different places. Funding has been relatively strong over many decades, perhaps because budgets for defense are relatively large in America. Governments in Europe are trying to build a community, whereas in America, they are trying to rock the boat. There is more emphasis in Europe on inclusiveness and diversity. Investment in research in Europe is still a fraction of the public and private investment in other regions of the world. In Chinese industry, there is an emphasis on bold initiatives and milestones. America should invest more in AI because of investments in China [5].

Language is a complex topic to study. CL tends to treat language as a closed system. Data privacy is a concern everywhere, especially in Europe. Data privacy seems less respected in Asia, especially from China. The press in Asia is tracking statistics such as government investments and numbers of publication by region. The overall picture that emerges is that large-scale resources of knowledge for language, whether lexical or about the world, still remain

too sparse and too imprecise. Translations remain error-prone, but their quality is usually sufficient for readers to grasp the general drift of the source contents.

6. CONCLUSION

Computational linguistics is now an active sub-discipline in applied linguistics. It is a field of data science that powers chatbots, search engines, and more. Applications of computational linguistics techniques range from those minimally dependent on linguistic structure and meaning to those that attain some level of competence in comprehending and using language. At Amazon, computational linguists and language engineers work on Alexa. At Apple, computational linguists and speech engineers develop Siri.

It has been shown that languages can be learned with a combination of simple input presented incrementally as the child develops better memory and longer attention span. Researchers have created a system which not only predicts future linguistic evolution but also gives insight into the evolutionary history of modern-day languages [6]. More information about computational linguistics can be found in the books in [7-31] and the following related journals:

- *Computational Linguistics*
- *International Journal of Computational Linguistics*

REFERENCES

[1] <https://image2.slideserve.com/5312589/slide9-1.jpg>

[2] “What is computational linguistics? insights and career guide,” November 2023,

<https://www.coursera.org/articles/computational-linguistics>

[3] K. Yasar, A. S. Gillis, and C. Bernstein, “Computational linguistics (CL),”

<https://www.techtarget.com/searchenterpriseai/definition/computational-linguistics-CL>

[4] “Computational linguistics,” February 2014,

<https://plato.stanford.edu/entries/computational-linguistics/>

[5] K. Church and M. Liberman, “The Future of Computational Linguistics: On Beyond Alchemy,” *Frontier of Artificial Intelligence*, vol. 4, April 2021.

[6] “Computational linguistics,” *Wikipedia*, the free encyclopedia,

https://en.wikipedia.org/wiki/Computational_linguistics

[7] A. Przepiorkowski et al. (eds.) , *Computational Linguistics: Application*. Springer, 2012.

[8] R. Grishman, *Computational Linguistics: An Introduction*. Cambridge University Press, 1986.

[9] S. Abney, *Semisupervised Learning for Computational Linguistics*. CRC Press, 2007.

[10] B. McGillivray, *Methods in Latin computational linguistics*. Paris: Brill, 2013.

[11] G. Ferrari, *State of the Art in Computational Linguistics*. na, 2003.

- [12] A. Gliozzo and C. Strapparava, *Semantic Domains in Computational Linguistics*. Springer Science & Business Media, 2009.
- [13] F. Lamarche, A. Lecomte, and G. Perrier (eds.), *Logical Aspects of Computational Linguistics*. Springer, 1999.
- [14] C. Souter and E. Atwell (eds.), *Corpus-based Computational Linguistics*. Rodopi, 1993.
- [15] C. L. Paris, W. R. Swartout, and W. C. Mann (eds.), *Natural Language Generation in Artificial Intelligence And Computational Linguistics*. Springer Science & Business Media, 2013
- [16] B. Srinivasa-Desikan, *Natural Language Processing and Computational Linguistics: A Practical Guide to Text Analysis With Python, Gensim, spaCy, and Keras*. Packt Publishing Ltd, 2018.
- [17] J. Heinz et al., *Grammatical Inference for Computational Linguistics*. Morgan & Claypool, 2016.
- [18] M. Z. Kurdi, *Natural Language Processing and Computational Linguistics: Speech, Morphology and Syntax*. John Wiley & Sons, vol. 1, 2016.
- [19] M. Z. Kurdi, *Natural Language Processing and Computational Linguistics 2: Semantics, Discourse and Applications*. John Wiley & Sons, vol.2. 2017.
- [20] G. Hirst, *Views of Text Meaning in Computational Linguistics: Past, Present, and Future*. na, 2007.
- [21] A. Zampolli, N. Calzolari, and M. Palmer (eds.), *Current Issues in Computational Linguistics: In Honour of Don Walker*. Springer Science & Business Media, 1994.
- [22] G. Sidorov, *Syntactic n-grams in Computational Linguistics*. Cham, Switzerland: Springer International Publishing, 2019.
- [23] D. Nouvel, M. Ehrmann, and S. Rosset, *Named Entities for Computational Linguistics*. John Wiley & Sons, 2016.
- [24] R. Mitkov (ed.), *The Oxford Handbook of Computational Linguistics*. Oxford University Press, 2nd edition, 2014.
- [25] A. Clark, C. Fox, and S. Lappin (eds.), *The Handbook of Computational Linguistics and Natural Language Processing*. Blackwell Publishing, 2010.
- [26] *Computational Linguistics: Concepts, Methodologies, Tools, and Applications*. Information Science Reference, 2014.

[27] C. F. Alexander and S. Lappin (eds.), *The Handbook of Computational Linguistics and Natural Language Processing*. John Wiley & Sons, 2012

[28] R. Grishman, *Computational linguistics: an introduction*. Cambridge University Press, 1986.

[29] R. Hausser, *Foundations of Computational Linguistics: Human-Computer Communication in Natural Language*. Springer-Verlag, 3rd edition, 2001.

[30] N. Dershowitz and E. Nissan (eds.), *Language, Culture, Computation: Computational Linguistics*. Springer, 2014.

[31] S. Pogodalla, *Logical Aspects of Computational Linguistics*. Springer, 2011.

ABOUT THE AUTHORS

Matthew N.O. Sadiku is a professor emeritus in the Department of Electrical and Computer Engineering at Prairie View A&M University, Prairie View, Texas. He is the author of several books and papers. His areas of research interest include computational electromagnetics, computer networks, and marriage counseling. He is a life fellow of IEEE.

Chandra M. M. Kotteti is an assistant professor in the School of Computer Science and Information Systems at Northwest Missouri State University, Maryville, MO. His current research interests include machine learning, deep learning, data science, and computer science.

C. Author: Email: chandra@nwmissouri.edu

Janet O. Sadiku holds bachelor degree in Nursing Science in 1980 at the University of Ife, now known as Obafemi Awolowo University, Nigeria and Master's degree from Juliana King University, Houston, TX in December 2022. She has worked as a nurse, educator, and church minister in Nigeria, United Kingdom, Canada, and United States. She is a co-author of some papers and books.

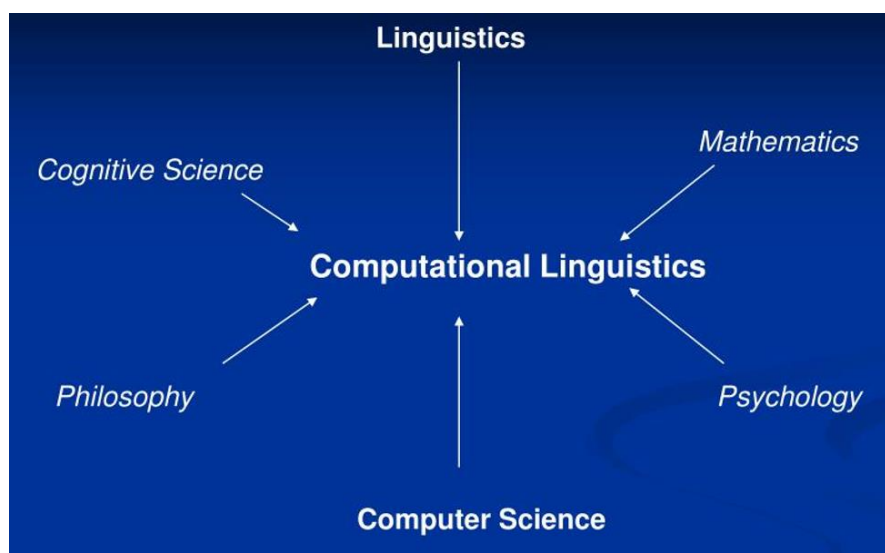


Figure 1 Computational linguistics as an interdisciplinary field [1]