

Computational Linguistics in Bulgaria

Svetla Koeva

Department of Computational
Linguistics,
Institute for Bulgarian Language,
Bulgarian Academy of Sciences
52 Shipchenski prohod, Bldg. 17,
1113 Sofia, Bulgaria
svetlal@dcl.bas.bg

*The article introduces the journal **Computational Linguistics in Bulgaria**, an annual open access peer-reviewed journal published by the Department of Computational Linguistics at the Institute for Bulgarian Language of the Bulgarian Academy of Sciences. The relationship between the terms computational linguistics, natural language processing and artificial intelligence is briefly commented on in order to clarify the concept behind the journal's name. The focus is then placed on the Bulgarian language and the Bulgarian research community, emphasising the importance of international contributions for the development of scientific cooperation and progress.*

*The scope of the journal **Computational Linguistics in Bulgaria** is presented: It publishes articles on all areas of theoretical computational linguistics as well as on existing language resources, datasets and technologies for natural language processing and artificial intelligence. The journal promotes new approaches and methods, especially those aimed at applying language technologies to small and still resource-poor languages such as Bulgarian.*

Keywords: computational linguistics, natural language processing, artificial intelligence, *Computational Linguistics in Bulgaria*

1. Introduction

The first issue of the **Computational Linguistics in Bulgaria** journal (JCLIB),¹ an annual open-access peer-reviewed journal, is published by the Department of Computational Linguistics at the Institute for Bulgarian Language of the Bulgarian Academy of Sciences. The editorial policy of the *Computational Linguistics in Bulgaria* journal includes the publication of articles from all areas of **theoretical computational linguistics** in combination with **available language resources, datasets and technologies for natural language processing and artificial intelligence**. The focus is on new approaches and methods, especially with

¹ <https://jclib.dcl.bas.bg/>

regard to their application to **small and resource-poor languages such as Bulgarian**, in order to bridge the gap between large and small languages in terms of language technologies.

The idea for the journal is not new, it was born together with the idea for the biennial conference *Computational Linguistics in Bulgaria*, organised by the Department of Computational Linguistics at the Institute for Bulgarian Language of the Bulgarian Academy of Sciences, which started in 2014. However, there were many objective and subjective reasons that prevented the publication of the journal. The Scientific Council of the Institute for Bulgarian Language of the Bulgarian Academy of Sciences has decided to launch the journal on **7 November 2024**.

The aim of the publisher and the editorial board is to make the journal a recognised forum for the publication of scientific research in the field of computational linguistics, natural language processing and artificial intelligence, with a focus on the Bulgarian language, which is either the direct subject of the studies and applications or whose research could be significantly influenced in the future by a variety of innovative developments. It goes without saying that both young and established researchers from Bulgaria and abroad, as well as outstanding researchers who can contribute to significant advances in the field of computational linguistics, natural language processing and artificial intelligence, are welcome as authors.

Since the terms computational linguistics, natural language processing and artificial intelligence appear several times in this text, it is worth clarifying what we mean by each of them, outlining their specific area of application and explaining how they relate to each other and overlap.

2. Computational Linguistics, Natural Language Processing and Artificial Intelligence

Several terms (and corresponding concepts) are used to describe related areas of research and development: computational linguistics (CL), natural language processing (NLP), language engineering (LE), human language technology (HLT), language technology (LT), artificial intelligence (AI), etc. However, we will only focus on some of them.

If we compare the terms computational linguistics and natural language processing, the word *linguistics* in the first term refers to the scientific discipline and, with its modifier, forms a term for a new specific field of research, namely computational linguistics. In contrast, the word *processing* from the term *natural language processing* refers to the processing of certain data, in this case natural language.

There are many definitions for both terms, computational linguistics and natural language processing, which show a different understanding of their content: In different interpretations, the terms can overlap, subsume or refer to related but nevertheless different concepts.

For example, there is a narrow understanding of computational linguistics that implies that computational linguistics provides sophisticated methods for linguistic research. In the modern development of technologies, this is a more appropriate understanding of theoretical linguistics itself, which uses various language data analyses to prove or reject theoretical linguistic hypotheses.

Another, more widespread view of computational linguistics is that computational linguistics seeks to define how humans compute and produce language by formulating formal grammars and probabilistic models and developing efficient algorithms for machine learning, language generation and understanding that appear suitable for capturing the range of phenomena in human languages.

Almost 16 years ago, the following definition was made (Calzolari 2009):

The term CL includes the disciplines dealing with models, methods, technologies, systems and applications concerning the automatic processing of a language, both spoken and written. CL therefore includes both Speech Processing (or processing of the spoken word, and Natural Language Processing (NLP) or text processing. SP and NLP have closely linked objectives such as human-machine vocal interaction and human language understanding, to be used in many applications, such as machine translation, speech-to-speech translation, information retrieval, and so on.

Another definition of computational linguistics, published in 2020, states that (Schubert 2020):

Computational linguistics is the scientific and engineering discipline concerned with understanding written and spoken language from a computational perspective, and building artifacts that usefully process and produce language, either in bulk or in a dialogue setting. To the extent that language is a mirror of mind, a computational understanding of language also provides insight into thinking.

Some authors emphasise that it is difficult to distinguish between computational linguistics and natural language processing (Hirschberg and Manning 2015, 261):

Computational linguistics, also known as natural language processing (NLP), *is the subfield of computer science concerned with using computational techniques to learn, understand, and produce human language content. Computational linguistic systems can have multiple purposes: The goal can be aiding human-human communication, such as in machine translation (MT); aiding human-machine communication, such as with conversational agents; or benefiting both humans and machines by analyzing and learning from the enormous quantity of human language content that is now available online.*

Many more definitions could be given, but we can summarise that computational linguistics is concerned with the theoretical modelling and formal description of language, while natural language processing applies theoretical investigations to solve real language problems in the context of interaction with computers. The terms are often used interchangeably because theory and application are inherently interdependent – neither can exist without the other.

To complicate things further, let us briefly examine how computational linguistics and natural language processing relate to artificial intelligence. In many views, the first two (or at least natural language processing) are areas of artificial intelligence (Navigli 2018, 5697):

Natural Language Processing (NLP) is a challenging field of Artificial Intelligence which is aimed at addressing the issue of automatically processing human language, called natural language, in written form. This is to be achieved by way of the automatic analysis, understanding and generation of language.

Cole Stryker and Jim Holdsworth have posted similar thoughts:²

Natural language processing (NLP) is a subfield of computer science and artificial intelligence (AI) that uses machine learning to enable computers to understand and communicate with human language. NLP enables computers and digital devices to recognize, understand and generate text and speech by combining computational linguistics, the rule-based modeling of human language together with statistical modeling, machine learning and deep learning. NLP research has helped enable the era of generative AI, from the communication skills of large language models (LLMs) to the ability of image generation models to understand requests.

One of the most frequently cited definitions of artificial intelligence is that of John McCarthy, which was given several decades ago and revised in 2007 (McCarthy 2007, 2):

Artificial intelligence is the science and engineering of making intelligent machines, especially intelligent computer programs.

(Human) intelligence is not just about a person's ability to learn and use language. There is ample evidence from various fields where machines successfully (in some cases even better than humans) perform tasks that require human intelligence. In road transport, for example, self-driving cars are no longer experiments: Waymo, one of the largest providers in the US, offers more than 150,000 autonomous rides every week, while Baidu's affordable Apollo Go robotaxi fleet now serves numerous cities in China (Maslej et al. 2025, 156).

When we turn to the question of the relationship between computational linguistics, natural language processing and artificial intelligence, it may be helpful to use the categorisation of artificial intelligence systems into four basic design principles: acting humanly, thinking humanly, thinking rationally and acting rationally (Russell and Norvig 2022, 19-20):

Historically, researchers have pursued several different versions of AI. Some have defined intelligence in terms of fidelity to human performance, while others prefer an abstract, formal definition of intelligence called rationality – loosely speaking, doing the “right thing”. The subject matter itself also varies: some consider intelligence to be a property of internal thought processes and reasoning, while others focus on intelligent behavior, an external characterization. From these two dimensions – human vs. rational and thought vs. behavior – there are four possible combinations, and there have been adherents and research programs for all four.

² <https://www.ibm.com/think/topics/natural-language-processing>

It is also pointed out that the rationalist approach to artificial intelligence involves a combination of mathematics and engineering and is associated with statistics and control theory (Russell and Norvig 2022, 20). However, when it comes to human language, it has not been possible to avoid (computational) linguistics.

Even if we restrict ourselves to the fact that artificial intelligence is currently mainly concerned with acting rationally, the activities related to human language are a subset of the field artificial intelligence, so we should agree that computational linguistics (natural language processing) is part of artificial intelligence understood in this way.

3. Computational Linguistics in Bulgaria

The connection between the journal *Computational Linguistics in Bulgaria* and the conference *Computational Linguistics in Bulgaria* is close, although the conference publishes its own proceedings. The *Computational Linguistics in Bulgaria* (CLIB)³ conference is an international event with the aim of exploring new approaches and methods in computational linguistics and natural language processing, especially with regard to their application to small and less well-resourced languages such as Bulgarian, and bridging the gap between “large” and “small” languages in terms of language technologies.

Original contributions on the following topics are expected at the CLIB conference (and also in the journal): computer-assisted language learning, training and education; information retrieval; information extraction; text mining and knowledge graph inference; linguistic foundations for computer vision and robotics; language modelling; language theories and cognitive modelling for NLP; large language models and NLP evaluation methods; language resources and benchmarking for large language models; language resource construction and annotation; machine learning for NLP; machine translation; multilingualism; translation aids; morphology and segmentation; natural language generation, understanding, summarisation and simplification; ontologies, terminology and knowledge representation; sentiment analysis; authorship analysis; opinion and argumentation analysis; speech recognition, synthesis and understanding of spoken language; tagging, chunking, syntax and parsing; and other related topics.

The part “in Bulgaria” in the names of the journal and the conference means several things:

- The journal is published in **Bulgaria**, the conference takes place in **Bulgaria**.
- The focus of the journal (as well as the conference) is on the **Bulgarian language** in the broadest sense: computational linguistic research in Bulgarian, but also datasets, models, technologies for other languages that can be newly implemented or adopted for Bulgarian. Of course, outstanding achievements in the field of computational linguistics, which can influence not only computational linguistic research in Bulgarian, but also the whole scientific field

³ <https://dcl.bas.bg/clib/>

of computational linguistics and natural language processing, are extremely important and of particular interest to the journal and the conference.

- The aim of the journal and the conference is to connect **Bulgarian researchers** at home and abroad and to promote the exchange of ideas, resources and successes. Both forums welcome research contributions from all over the world and promote scientific communication, collaboration and mutual support in the knowledge that research thrives on exchange and joint endeavour.

4. Journal *Computational Linguistics in Bulgaria*

It has already been mentioned that the journal *Computational Linguistics in Bulgaria* publishes research papers from all areas of theoretical computational linguistics as well as studies on existing resources, datasets and technologies for natural language processing and artificial intelligence. The journal emphasises innovative approaches and methods, especially those aimed at the application of language technologies to small and resource-poor languages such as Bulgarian, with the overarching goal of narrowing the gap between large and small languages in the development and accessibility of language technologies.

The journal accepts submissions of original research with a particular focus on Bulgarian and other related languages, but also welcomes contributions on new theories, datasets and technologies applicable to a wide range of languages.

Finally, we must mention some formal features of the journal *Computational Linguistics in Bulgaria*:

It is an **open access** journal, which means that the full text of all articles is freely available on the Internet so that users can read, download, copy, distribute, print, search, link or index the content and use it as data in software or for any other lawful purpose – without financial, legal or technical barriers beyond Internet access.⁴

The research results published in the journal are in the public domain and may be used under the terms of a Creative Commons Attribution 4.0 International Public Licence (CC-BY-4.0).⁵ Anyone is free to share – copy, distribute and transmit, remix – adapt the work, under the condition of attribution – the original authors must be credited.

The journal *Computational Linguistics in Bulgaria* adheres to the ethical guidelines for journal publications of the **Committee on Publication Ethics** (COPE).⁶

To summarise, by combining research on Bulgarian and related languages with global perspectives, we aim to create a space where ideas thrive, innovation flourishes and language technologies bridge the gap between resource-rich and resource-poor languages.

⁴ <https://www.budapestopenaccessinitiative.org/>

⁵ <https://creativecommons.org/licenses/by/4.0/>

⁶ <https://publicationethics.org/>

References

- Calzolari, Nicoletta. 2009. *Encyclopedia of Life Support Systems (EOLSS)*, volume 1, chapter Linguistics. Computational Linguistics. EOLSS Publishers, Paris, France. Developed under the Auspices of the UNESCO.
- Hirschberg, Julia and Christopher D. Manning. 2015. Advances in natural language processing. *Science*, 349(6245):261–266.
- Maslej, Nestor, Loredana Fattorini, Raymond Perrault, Yolanda Gil, Vanessa Parli, Njenga Kariuki, Emily Capstick, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Juan Carlos Nibbles, Yoav Shoham, Russell Wald, Toby Walsh, Armin Hamrah, Lapo Santarlasci, Julia Betts Lotufo, Alexandra Rome, Andrew Shi, and Sukrut Oak. 2025. *The AI Index 2025 Annual Report*. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA. https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf.
- McCarthy, John. 2007. *What is artificial intelligence?* Computer Science Department, Stanford University.
- Russell, Stuart J. and Peter Norvig. 2022. *Artificial Intelligence: A Modern Approach*, 4 edition. Pearson.
- Schubert, Lenhart. 2020. Computational Linguistics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, Spring 2020 edition. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/spr2020/entries/computational-linguistics/>.