ЦЕНТЪР ЗА ОБУЧЕНИЕ – БАН

1000 София
ул. „Сердика“ № 4
http://edu.bas.bg

email: tdc-phd@cu.bas.bg
тел.: 02 987 31 67
02 979 52 60

Basic Information:

Course Title: .Introduction to Computational Linguistics
Lecturer: Prof. Svetla Koeva
Phone: +35928722302
Email: svetla@dcl.bas.bg
Total Teaching Hours: 30

Annotation (up to 150 words)

This course introduces doctoral students to the foundational algorithms of modern computational linguistics and their most significant applications. The material is presented accessibly, requiring no prior background in mathematics or computer science.

Students will gain a solid understanding of computational linguistics as both a theoretical and applied discipline, developing familiarity with its core tasks, methods, and results – preparation that will equip them to engage with advanced developments throughout their careers.

Course content (brief description by topics or modules)

#	Topic	Hours
1	Computational linguistics: scope, tasks, and applications	2
2	Formal language theory: the Chomsky hierarchy	2
3	Regular expressions and finite-state methods	2
4	Context-free and context-sensitive grammars	2
5	Levels of text processing: tokenisation, tagging, parsing, and word sense disambiguation	2
6	Vector semantics and word embeddings	2
7	Neural network architectures for NLP	2
8	Transformer models and large language models	2
9	Dialogue systems and conversational agents	2
10	Prompt engineering and model interaction	2
11	Machine translation and computer-aided translation	2
12	Automatic speech recognition and text-to-speech synthesis	2
13	Linguistic annotation: principles and practice	2

14	Coreference resolution and entity linking	2
15	Semantic modelling and knowledge representation	2
Total		**30**

Teaching and assessment methods

The course is delivered through lectures and hands-on practical sessions. Lectures introduce the theoretical foundations and key concepts of each topic, while practical sessions allow students to apply computational tools and methods to real linguistic data, consolidating their understanding through direct engagement with the material.

Assessment is by coursework. Students must complete an individual project in which they apply one or more of the methods covered in the course to a problem of their choice, ideally related to their doctoral research. The project is accompanied by a written report demonstrating the student's understanding of the relevant methods, their implementation, and a critical discussion of the results.

Competencies acquired as a result of training (3–5 points)

Upon successful completion of the course, students will be able to:

Demonstrate a systematic understanding of the core tasks, methods, and theoretical foundations of computational linguistics, including formal language theory, text processing pipelines, and modern neural approaches.

Design and conduct a small-scale computational linguistics project, from problem formulation and method selection through to implementation, analysis, and written presentation of results.

Engage productively with the scholarly literature in computational linguistics and related fields, situating their doctoral research within the broader landscape of language technology and linguistic theory.

Reflect critically on the capabilities and limitations of current language technologies, including issues of data quality, model bias, and the relationship between computational models and linguistic theory.

Literature:

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., & Amodei, D. Language models are few-shot learners. In *Advances in Neural Information Processing Systems, 33.* Curran Associates, Red Hook, NY, 2020, pp. 1877–1901.

Carta, S., Giuliani, A., Piano, L., Podda, A. S., Pompianu, L., & Tiddia, S. G. Knowledge graph construction: Extraction, learning, and evaluation. *Applied Sciences, 15*(7), Article 3727. MDPI, Basel, 2025.

de Marneffe, M.-C., Manning, C. D., Nivre, J., & Zeman, D. Universal Dependencies. *Computational Linguistics*, 47(2) MIT Press, Cambridge, MA, 2021, pp. 255–308.

Jurafsky, Daniel and James H. Martin. 2025. *Speech and Language Processing: An Introduction to Natural Language Processing*, 3rd edition. Online manuscript released January 12, 2025.

Lane, Hobson and Maria Dyshel. 2025. Natural Language Processing in Action. 2nd edition. MEAP.

Manning, C. D., and Schutze, H. 1999. *Foundations of statistical natural language processing*. MIT Press.

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. Robust speech recognition via large-scale weak supervision. In *Proceedings of the 40th International Conference on Machine Learning*. PMLR, Honolulu, 2023, pp. 28492–28518.

Rothwell, A., Way, A., & Youdale, R. *Computer-Assisted Literary Translation*. Routledge, London, 2023.

Russell, Stuart J. and Peter Norvig. 2022. *Artificial Intelligence: A Modern Approach,* 4th edition. Pearson.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems, 30*. Curran Associates, Red Hook, NY, 2017, pp. 5998–6008.

Коева, Св., Д. Благоева (ред.). *Езикови ресурси и технологии за български език*. София: Академично издателство „Проф. “Марин Дринов, 2014, 310 с. ISBN: 978-954-322-797-6.

Additional information (optional) (e.g., special requirements, laboratory equipment, prior knowledge)

No special requirements.

The course is primarily designed for linguists, but is equally open to researchers from other humanities disciplines, as well as to mathematicians and computer scientists seeking to broaden their perspective.