

ARTUR KULMIZEV

Antwerp, Belgium
akulmizev@gmail.com

EDUCATION

- UC Louvain; Belgium** *2025-current*
Postdoctoral Researcher, CENTAL (Center for Natural Language Processing)
Supervisor: Marie-Catherine de Marneffe
- KU Leuven; Belgium** *2023-24*
Postdoctoral Researcher, department of Computer Science
Supervisor: Miryam de Lhoneux
- Uppsala University; Sweden** *2019-23*
PhD in Computational Linguistics
Dissertation: *The Search for Syntax: Investigating the Syntactic Knowledge of Language Models Through the Lens of Dependency Parsing*
Supervisors: Joakim Nivre and Anders Søgaard (University of Copenhagen)
Opponent: Roger Levy (Massachusetts Institute of Technology)
Committee: Anna Korhonen (University of Cambridge), Jörg Tiedemann (University of Helsinki), Johan Boye (KTH)
- University of the Basque Country; Spain** *2017-18*
MSc in Language and Communication Technology
Year two of EM-LCT program.
- University of Groningen; Netherlands** *2016-17*
MA in Language and Communication Technology
Year one of EM-LCT program.
- University of Massachusetts-Amherst; USA** *2008-12*
BA in English Language and Literature
Cum Laude
Certificate in Technical Writing and Professional Communication
Minor in Spanish
Minor in Information Technology

TEACHING

- Courses:**
- Information Retrieval and Search Engines*, MSc in AI/Computer Science, KU Leuven *2023*
- Course for engineering students in information retrieval, search, and related algorithms.
 - Led development of final course project on retrieval-augmented generation (RAG).
 - Developed exercises and led practical sessions for latent semantic indexing and word embeddings.
- Natural Language Processing*, MA in Language Technology, Uppsala University *2019-21*
- Course for first-year MA students in fundamental NLP concepts: language modeling, part-of-speech tagging, dependency/constituency parsing, semantic role labeling. Co-taught with 1-2 instructors.
 - Administered lab sessions and gave lectures in flipped classroom format.
 - Designed and graded assignments and final projects.

- Course for first-year MA students in data structures and algorithms.
- Designed and administered *Practical Programming for Language Technologists Module*, consisting of 3 lectures and 3 labs. Contents included web scraping, data visualization, and modular programming.
- Re-designed entirety of course lecture and examination materials (year 2, single instructor)

MASTER'S THESIS SUPERVISION

Ivan Kiseliou: Better cooperation through communication in multi-agent reinforcement learning

Robin Hruska: Emergent Communication in Guessing Game Environment

Ziyang Luo: Positional Embeddings and Artefacts in Masked Language Models

Johnny Chang: Improving Transformer-Based Molecular Optimization Using Reinforcement Learning

Xi Yan: Pre-training a knowledge enhanced model in biomedical domain for information extraction

Viktorija Buzaitė: An End-to-End Native Language Identification Model without the Need for Manual Annotation

Jae Eun Hong: Regularized Fine-tuning Strategies for Neural Language Models: Application of entropy regularization on GPT-2

Eva Elzbieta Sventickaitė: Distillation or loss of information?: The effects of distillation on model redundancy

Jiayi Zhang: Analysis of Syntactic Behaviour of Neural Network Models by Using Gradient-Based Saliency Method: Comparative Study of Chinese and English BERT, Multilingual BERT and RoBERTa

Flor Ortiz: In-Domain and Cross-Domain Classification of Patronizing and Condescending Language in Social Media and News Texts

HONORS AND SCHOLARSHIPS

GUF GRANT

2017

Awarded to outstanding Master's students at the University of Groningen.

EM-LCT Consortium Scholarship

2016

Awarded to motivated students enrolled in the EM-LCT program.

Dean's List

2008-12

Awarded to Bachelor's students who maintain at least a 3.5 GPA per semester.

John and Abigail Adams Scholarship

2008-12

Awarded to in-state Massachusetts students who maintain at least a 3.5 per year.

RESEARCH INTERESTS

- Multilingual and cross-lingual NLP models and methods.
- Explainability, faithfulness, and LLM reasoning.
- Data quality and evaluation.
- Interpretability methods for neural NLP models.
- Computational and cognitive models of natural language processing.
- Dependency grammar and Universal Dependencies.
- Information theoretic and efficiency-based models of syntax.

Journal Articles:

Artur Kulmizev, Joakim Nivre. *Schrödinger's Tree — On the Role of Syntax in Contemporary NLP*. Frontiers in Artificial Intelligence, Volume 5. 2022.

Main Conference Proceedings:

Artur Kulmizev, Joakim Nivre. *Investigating UD Treebanks via Dataset Difficulty Measures*. EACL 2023.

Mostafa Abdou, Vinit Ravishankar, Artur Kulmizev, Anders Søgaard. *Word Order Does Matter and Shuffled Language Models Know It*. ACL 2022.

Mostafa Abdou, Artur Kulmizev, Daniel Hershcovich, Stella Frank, Ellie Pavlick, Anders Søgaard. *Can Language Models Encode Perceptual Structure Without Grounding? A Case Study in Color*. CoNLL 2021.

Ziyang Luo, Artur Kulmizev, Xiaoxi Mao. *Positional Artefacts Propagate Through Masked Language Model Embeddings*. ACL 2021.

Artur Kulmizev, Vinit Ravishankar, Mostafa Abdou, Anders Søgaard, Joakim Nivre. *Attention Can Reflect Syntactic Structure — If You Let It*. EACL 2021.

Artur Kulmizev, Vinit Ravishankar, Mostafa Abdou, Joakim Nivre. *Do Neural Language Models Show Preferences for Syntactic Formalisms?*. ACL 2020.

Artur Kulmizev, Miryam de Lhoneux, Johannes Gontrum, Elena Fano, Joakim Nivre. *Deep Contextualized Word Embeddings in Transition-Based and Graph-Based Dependency Parsing—A Tale of Two Parsers Revisited*. EMNLP-IJCNLP 2019.

Mostafa Abdou, Artur Kulmizev, Felix Hill, Daniel M. Low, Anders Søgaard. *Higher-order comparisons of sentence encoder representations*. EMNLP-IJCNLP 2019.

Mostafa Abdou, Artur Kulmizev, Vinit Ravishankar, Lasha Abzianidze, Johan Bos. EMNLP 2018.

Mostafa Abdou, Artur Kulmizev, Vinit Ravishankar. *MGAD: Multilingual Generation of Analogy Datasets*. LREC 2018.

Shared Task and Workshop Submissions:

Daniel Hershcovich, Miryam de Lhoneux, Artur Kulmizev, Elham Pejhan, Joakim Nivre. *Køpsala: Transition-Based Graph Parsing via Efficient Training and Effective Encoding*. IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies, 2020.

Mostafa Abdou, Artur Kulmizev, Joan Ginés i Ametllé. *AffectThor at SemEval-2018 Task 1: A cross-linguistic approach to sentiment intensity quantification in tweets*. SemEval 2018.

Artur Kulmizev, Mostafa Abdou, Vinit Ravishankar, Malvina Nissim. *Discriminator at SemEval-2018 Task 10: Minimally Supervised Discrimination*. SemEval 2018.

Artur Kulmizev, Bo Blankers, Johannes Bjerva, Malvina Nissim, Gertjan van Noord, Barbara Plank, Martijn Wieling. *The Power of Character N-grams in Native Language Identification*. Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications (BEA), 2018.

Pre-prints:

Kushal Tatariya, Artur Kulmizev, Wessel Poelman, Esther Ploeger, Marcel Bollmann, Johannes Bjerva, Jiaming Luo, Heather Lent, and Miryam de Lhoneux. *How Good is Your Wikipedia?*. arXiv preprint. Under review at TACL.

ACADEMIC ACTIVITIES

Invited talk, UC Louvain.	<i>Louvain-la-Neuve, BE — 2025</i>
Invited talk, University of Antwerp.	<i>Antwerp, BE — 2024</i>
Poster presentation, CLIN 34.	<i>Leiden, NL — 2024</i>
Area Chair, ARR.	<i>online — 2024</i>
Area Chair, ARR.	<i>online — 2024</i>
Presentation, Leuven.AI Junior Researcher day.	<i>Leuven, BE — 2023</i>
Started postdoc.	<i>Leuven, BE — 2023</i>
Attended first UniDive general meeting.	<i>Paris, FR — 2023</i>
Defended PhD Dissertation.	<i>Uppsala, SE — 2023</i>
Attended first UniDive general meeting.	<i>Paris, FR — 2023</i>
Invited talk, University of Helsinki.	<i>Helsinki, FI — 2022</i>
Joined the UniDive (Universality, diversity, and idiosyncrasy in language technology) COST action as a member.	<i>2022</i>
Served as collector, Dagstuhl Workshop.	<i>2021</i>
Participated in the Dagstuhl Workshop on Universals of Linguistic Idiosyncrasy in Multilingual Computational Linguistics.	<i>2021</i>
Oral Presentation, EACL 2021.	<i>2021</i>
Appointed Equal Opportunity Rep., Lingfil PhD Council, Uppsala University.	<i>Uppsala, SE — 2020</i>
Re-appointed Chair of Lingfil PhD Council, Uppsala University.	<i>Uppsala, SE — 2020</i>
Oral Presentation, ACL 2020.	<i>2020</i>
Participated in <i>IWPT 2020 Shared Task</i> .	<i>2020</i>
Invited talk at NLPL Winter School.	<i>Skeikampen, NO — 2020</i>
Oral Presentation.	<i>Hong Kong — 2019</i>
Poster Presentation.	<i>Hong Kong — 2019</i>
Attended 1st EurNLP summit.	<i>London, UK — 2019</i>
Appointed Chair PhD Council, Uppsala University.	<i>Uppsala, SE — 2019</i>
Attended LxMLS Summer School.	<i>Lisbon, PT — 2019</i>
Attended 3rd Google NLP summit.	<i>Zurich, CH — 2019</i>
Attended NLPL Winter School.	<i>Skeikampen, NO — 2019</i>
Poster Presentations, EMNLP 2018.	<i>Brussels, BE — 2018</i>
Poster Presentation, TABUDAG 2018.	<i>Groningen, NL — 2018</i>
Poster Presentation, LREC 2018.	<i>Miyazaki, JP — 2018</i>
Participated in <i>SemEval-2018 Task 1: Affect in Tweets</i> . Winning Group.	<i>2018</i>
Participated in <i>SemEval-2018 Task 10: Capturing Discriminative Attributes</i> .	<i>2018</i>
Participated in <i>2017 Shared Task on Native Language Identification</i> . Winning Group.	<i>2017</i>
Attended ESSLI 2017.	<i>Toulouse, FR — 2017</i>
Attended LOT Summer School.	<i>Leiden, NL — 2017</i>

WORK EXPERIENCE

University of Massachusetts, Amherst	<i>2012-2016</i>
<i>Learning Management System Administrator</i>	<i>Amherst, Massachusetts, USA</i>
<ul style="list-style-type: none">· Front-end administration for Blackboard Learn and various building blocks.· Management of institution-wide transition from self-hosted environment to managed hosting by Blackboard.· Technical project management and logistical planning for online teaching and learning environments.· Testing and implementation of service packs, cumulative patches, building blocks and integrations of other applications and services.	

SKILLS

Programming	Python (PyTorch, TensorFlow, JAX), R, Java, SQL
ML/NLP Frameworks	transformers, datasets, PEFT, accelerate, bitsandbytes, langchain fairseq, allennlp, scikit-learn, spaCy
LLM Frameworks	OpenAI, vllm, ollama, litellm, langchain
Infrastructure	SLURM, Docker, Google Cloud, Azure, AWS

LANGUAGES

English	Native-level
Russian	First language
Dutch	Written & Oral Proficiency (B1)
Spanish	Written & Oral Proficiency
Swedish	Basic (A1)
Basque	Basic (A1)

LEAVE PERIODS

<i>Birth of second child</i> — six month leave after Postdoc contract	<i>Antwerp, BE — 2024</i>
<i>Birth of first child</i>	<i>Antwerp, BE — 2022</i>