ARTUR KULMIZEV

Celestijnenlaan 200a ⋄ Leuven, Belgium 3000

artur.kulmizev@kuleuven.be

CURRENT POSITION

KU Leuven, Belgium   *2023-present*
Postdoctoral Researcher   *Leuven, BE*

· Postdoctoral Researcher in the LAGoM-NLP group within the Department of Computer Science at KU Leuven, led by Miryam de Lhoneux.
· Research into multilngual NLP methods (training dynamics, data/language sampling, tokenization, evaluation), with an emphasis on typological diversity and fairness.

EDUCATION

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
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:
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.

Advanced Programming, MA in Language Technology, Uppsala University   *2019-21*

- 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 include web scraping, data visualization, and modular programming.

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 Buzaité: *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 Sventickaité: *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 Masters 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.

- Interpretability methods for neural NLP models.

- Language language models and evaluation.

- Computational and cognitive models of natural language processing.

- Dependency grammar and Universal Dependencies.

- Information theoretic and efficiency-based models of syntax.

PUBLICATIONS

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 Sgaard. *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. *Kpsala: 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 Gins i Ametll. *AffecThor 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.

ACADEMIC ACTIVITIES

Defended PhD Dissertation.	*Uppsala, SE — 2023*
Attended first UniDive general meeting.	*Paris, FR — 2023*
Joined the UniDive (Universality, diversity, and idiosyncracy in language technology) COST action as a member.	*2022*
Served as collector, Dagstuhl Workshop.	*2021*
Participated in the Dagstuhl Workshop on Universals of Linguistic Idiosyncrasy in Multilingual Computational Linguistics.	*2021*
Oral Presentation, EACL 2021.	*2021*
Appointed Equal Opportunity Rep., Lingfil PhD Council, Uppsala University.	*Uppsala, SE — 2020*
Re-appointed Chair of Lingfil PhD Council, Uppsala University.	*Uppsala, SE — 2020*
Oral Presentation, ACL 2020.	*2020*
Participated in *IWPT 2020 Shared Task.*	*2020*
Invited talk at NLPL Winter School.	*Skeikampen, NO — 2020*
Oral Presentation.	*Hong Kong— 2019*
Poster Presentation.	*Hong Kong— 2019*
Attended 1st EurNLP summit.	*London, UK — 2019*
Appointed Chair PhD Council, Uppsala University.	*Uppsala, SE — 2019*
Attended LxMLS Summer School.	*Lisbon, PT — 2019*
Attended 3rd Google NLP summit.	*Zurich, CH — 2019*
Attended NLPL Winter School.	*Skeikampen, NO — 2019*
Poster Presentations, EMNLP 2018.	*Brussels, BE — 2018*
Poster Presentation, TABUDAG 2018.	*Groningen, NL — 2018*
Poster Presentation, LREC 2018.	*Miyazaki, JP — 2018*
Participated in *SemEval-2018 Task 1: Affect in Tweets.* Winning Group.	*2018*
Participated in *SemEval-2018 Task 10: Capturing Discriminative Attributes.*	*2018*
Participated in *2017 Shared Task on Native Language Identification.* Winning Group.	*2017*
Attended ESSLLI 2017.	*Toulouse, FR — 2017*
Attended LOT Summer School.	*Leiden, NL — 2017*

WORK EXPERIENCE

University of Massachusetts, Amherst   *2012-2016*
Learning Management System Administrator   *Amherst, Massachusetts, USA*

· 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.

RELEVANT SKILLS

Programming Languages	Python, R, Java
Machine Learning Toolkits	Tensorflow, PyTorch, Keras, allennlp, Huggingface, scikit-learn
Other	numpy; scipy; pandas; NLTK; gensim
	technical writing and documentation; git; LINUX/UNIX;
	web design (Bootstrap, Jekyll); data visualization
	(ggplot2, matplotlib, seaborn); LaTeX

LANGUAGES

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