Andrey Kutuzov

PERSONAL DETAILS

E-Mail	andreku@ifi.uio.no
Address	Postboks 1080, Blindern
	University of Oslo
	Norway, Oslo, 0316
Mobile	(+47) 4064-8218
Homepage	https://www.mn.uio.no/ifi/english/people/aca/andreku/
Github	akutuzov
ORCID	0000-0003-2540-5912
GScholar	Jq4Wq7AAAAAJ

PROFESSIONAL EXPERIENCE

Associate professor   University of Oslo, Norway
Research, teaching
   2022-now

Postdoctoral fellow   University of Oslo, Norway
Research, teaching
   2021-2022

Researcher   Nordic Language Processing Laboratory (NLPL)
Maintaining parts of NLPL infrastructure, including Word Embeddings Repository (`http: // vectors. nlpl. eu/ repository/ `)
   2019-2020

Doctoral research fellow   University of Oslo, Norway
Research, teaching, writing a PhD thesis
   2015-2020

Computational linguist and analytic   Mail.ru Search, Russia
linguistic support for a general-purpose web search engine
   2012-2015

Computational linguist   Lionbridge Technologies, Inc., Finland
automatic grammar checker for Russian
   2011-2014

EDUCATION

PhD in Computational Linguistics   2020
University of Oslo, Norway
   Supervisors: Associate professor Erik Velldal, Professor Lilja Øvrelid
   Thesis title: Distributional word embeddings in modeling diachronic semantic change
   Thesis text: https://www.duo.uio.no/handle/10852/81045

Machine Learning course   2016
Coursera
   Stanford University, 96.5% grade
   Certificate: https://www.coursera.org/account/accomplishments/certificate/9QBANEUKQBY9

M.Sc. Computational Linguistics   2014
National Research University Higher School of Economics, Russia
   Supervisors: Professor Svetlana Toldova
   Thesis Title: Semantic clustering of Russian web search results: possibilities and problems

Specialist in Linguistics and Language Teaching   2003
Tyumen State University, Russia

Roughly equivalent to Bachelor degree

TEACHING

Assistant lecturer   University of Oslo, Norway
Developed and taught both lectures and seminars, designed and graded intermediate assignments and exams at Master level courses
   2016-2022

1. Fall 2016: '**Distributional Semantics: Extracting Meaning from Data**' (half of the INF5820 course, developed entirely from scratch).

2. Fall 2017: '**Dependency parsing**', '**Word sense disambiguation**' and '**Semantic role labelling**' parts of the INF5830 course.

3. Fall 2018: '**Supervised machine learning**', '**Deep learning**', '**Language modelling**' and '**Word embeddings**' parts of the INF5820 course (developed from scratch).

4. Spring 2019: '**Supervised machine learning**', '**Deep learning**', '**Language modelling**' and '**Word embeddings**' parts of the IN5550 course.

5. Spring 2020, 2021, 2022: '**Deep learning**', '**Language modelling**', '**Word embeddings**' and '**Contextualised word embeddings**' parts of the IN5550 course.

PhD thesis advisor   University of Oslo, Norway
NLP
   2021-

1. '**Inductive bias for more efficient language modelling**' by David Samuel (started in 2021, ongoing).

Master theses advisor   University of Oslo, Norway
NLP
   2016-2022

1. '**Word embedding models as graphs**' by Erik Winge (graduated in 2018).

2. '**Adapting named entity recognition for the conflict research domain**' by Petter Kolbræk Eikeland (graduated in 2018).

3. '**Semantic change detection for Norwegian**' by Sigrid Riiser Kristiansen (started in 2021, ongoing).

External Master theses advisor National Research University Higher School of Economics, Russia
NLP
   2016-2020

1. '**Pronominal anaphora resolution as linguistic preprocessing for distributional semantic models**' by Olga Koslowa (graduated in 2016).

2. '**Linguistic representativeness of distributional semantic models on the lexical level: Can it be general or is it always task-specific?**' by Amir Bakarov (graduated in 2019).

3. '**Diachronic word representation shifts in contextualized embedding models**' by Julia Rodina (graduated in 2020).

Guest lecturer   National Research University Higher School of Economics, Russia
natural language processing, information retrieval
   2014-2015

Assistant lecturer   Tyumen State University, Russia
corpus linguistics, translation studies
   2006-2011

FUNDING ACQUISITION

High performance language technologies (HPLT) 2022
Coordinated by Kenneth Heafield, University of Edinburgh
    Horizon Europe Framework Programme; I took part in drafting the proposal and will be the project onsite manager in Oslo.

Kristine Bonnevie scholarship for a research stay 2022
Institute for Logic, Language and Computation, University of Amsterdam, Netherlands
    Hosts: professor Raquel Fernández, PhD fellow Mario Giulianelli

Teksthub support grant 2021
Annotation of historical semantic change test set for Norwegian
    `https://www.uio.no/tjenester/it/forskning/kompetansehuber/teksthub/utviklingsprosjekter/`
    I coordinated the work of 6-person team; the dataset is now created and available online
    (`https://github.com/ltgoslo/nor_dia_change`)

Young scientist mobility grant for a research internship 2018
Language Technology group, University of Hamburg, Germany
    Hosts: professor Chris Biemann, research assistant Alexander Panchenko

PROJECTS COORDINATION

NorLM (large-scale language models for Norwegian)) `http://norlm.nlpl.eu`

3rd International Workshop on Computational Approaches to Historical Language Change
(co-organizer) `https://languagechange.org/events/2022-acl-lchange/`

SemEval 2022 shared task on structured sentiment analysis
(co-organizer) `https://competitions.codalab.org/competitions/33556`

Maintaining software infrastructure on multiple HPC clusters
for the UiO Language Technology Group 2020-2022

RusVectōrēs (word embeddings for Russian online) `https://rusvectores.org/en`

Nordic Language Processing Laboratory
embeddings repository `http://vectors.nlpl.eu/repository/`

ShiftRy (diachronic semantic changes in Russian news) `https://shiftry.rusvectores.org/en/`

Simple-elmo library (maintainer) `https://github.com/ltgoslo/simple_elmo`

Gensim library (contributor) `https://github.com/RaRe-Technologies/gensim`

AWARDS

NORA Award for Distinguished Early Career Investigator NORA
Norwegian Artificial Intelligence Research Consortium (`https://www.nora.ai/`)
    2022

Area Chair Favourite Paper COLING-2018
Kutuzov et al. (2018)
    2018

2nd place in Russian Word Sense Induction and Disambiguation shared task Dialog-2018
Kutuzov (2018) `https://russe.nlpub.org/2018/wsi/`
    2018

Best NLP talk award AIST-2017
Kutuzov and Kunilovskaya (2017)
    2017

2nd place in Russian Semantic Similarity shared task Dialog-2015
Kutuzov and Andreev (2015) `https://russe.nlpub.org/`
    2015

CODING AND COMPUTING SKILLS

Python	Expert
Bash	Advanced
***NIX systems**	Advanced
Deep learning frameworks	
(*TensorFlow*, *PyTorch*, *Keras*, *HuggingFace Transformers*)	Expert

PEER REVIEWED FOR

- ACL rolling review (ARR, 2021, 2022)

- ACL main conference (2018, 2019, 2020, 2021)

- EACL main conference (2021)

- COLING main conference (2020)

- EMNLP main conference (2020, 2021)

- Journal of Natural Language Engineering (JNLE) (2019, 2020)

- Language Resources and Evaluation (LRE) (2017, 2018, 2019, 2020)

- LREC main conference (2016, 2018, 2020)

- NODALIDA main conference (2017, 2019, 2021)

- RANLP main conference (2019)

- Dialog main conference (2017, 2018, 2019, 2020, 2021)

- AINL main conference (2015, 2016, 2017, 2018, 2019, 2020)

- AIST main conference (2016, 2017, 2018, 2019, 2020)

INVITED TALKS AND TUTORIALS

Grammatical Profiling for Semantic Change Detection Research seminar
talk
 Montclair, 2021
New Language models NorwAI Innovate
talk
 Trondheim, 2021
NLP and Language Modelling for Norwegian Swedish NLP webinars
talk
 online, 2021
Contextualized embeddings for semantic change detection: lessons learned Department colloquium
talk
 Stuttgart, 2020
Diachronic word embeddings for semantic shift modelling: how to trace changes of meaning in time AINL
tutorial
 Tartu, Estonia, 2019
'It depends on the context': why word2vec is out of fashion and what is contextualised embeddings 4th Moscow-Tartu Digital Humanities School
invited talk
 Moscow, Russia, 2019
Deep neural networks in NLP: hype or not? 4th Kolmogorov seminar on computational linguistics

invited talk

Moscow, Russia, 2019

An Introduction to Deep Learning for NLP NeIC 2019 - Nordic Infrastructure for Open Science
tutorial

Copenhagen, Denmark, 2019 (together with Stephan Oepen)

Evolution of concepts in Russian news texts 3rd Moscow-Tartu Digital Humanities School
tutorial

Moscow, Russia, 2018 (together with Elizaveta Kuzmenko)

Historical geography in language statistics 2nd Moscow-Tartu Digital Humanities School
tutorial

Moscow, Russia, 2017 (together with Fedor Korandey and Elizaveta Kuzmenko)

Distributional semantic models and their applications Computational linguistics workshop
invited talk

Moscow, Russia, 2017 (Institute for System Analysis, Russian Academy of Science)

REFERENCES

Dr. Erik Velldal
University of Oslo, Norway
Department of Informatics
Postboks 1080
Blindern 0316 OSLO
erikve@ifi.uio.no
+47-22840119

Dr. Chris Biemann
University of Hamburg, Germany
Department of Informatics
22527
Vogt-Kölln-Str. 30, Hamburg
Chris.Biemann@informatik.uni-hamburg.de
+49(0)40-42883-2386

Prof. Svetlana Toldova
National Research University Higher School of Economics, Russia
Moscow
21/4 Staraya Basmannaya Ulitsa
Room 518
stoldova@hse.ru
+7(495)772-9590

Publications

Anna Aksenova, Ekaterina Gavrishina, Elisei Rykov, and Andrey Kutuzov. 2022. RuDSI: Graph-based word sense induction dataset for Russian. In *Proceedings of TextGraphs-16: Graph-based Methods for Natural Language Processing*, pages 77–88, Gyeongju, Republic of Korea. Association for Computational Linguistics.

Andrey Kutuzov, Samia Touileb, Petter Mæhlum, Tita Enstad, and Alexandra Wittemann. 2022. NorDiaChange: Diachronic semantic change dataset for Norwegian. In *Proceedings of the Language Resources and Evaluation Conference*, pages 2563–2572, Marseille, France. European Language Resources Association.

Mario Giulianelli, Andrey Kutuzov, and Lidia Pivovarova. 2022. Do not fire the linguist: Grammatical profiles help language models detect semantic change. In *Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change*, pages 54–67, Dublin, Ireland. Association for Computational Linguistics.

Andrey Kutuzov, Erik Velldal, and Lilja Øvrelid. 2022. Contextualized embeddings for semantic change detection: Lessons learned. *Northern European Journal of Language Technology*, 8(1).

Tatyana Iazykova, Denis Kapelyushnik, Olga Bystrova, and Andrey Kutuzov. 2021. Unreasonable effectiveness of rule-based heuristics in solving Russian SuperGLUE tasks. *Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii: Dialog conference*.

Andrey Kutuzov and Lidia Pivovarova. 2021. Three-part diachronic semantic change dataset for Russian. In *Proceedings of the 2nd International Workshop on Computational Approaches to Historical Language Change 2021*, pages 7–13, Online. Association for Computational Linguistics.

Andrey Kutuzov and Elizaveta Kuzmenko. 2021. Representing ELMo embeddings as two-dimensional text online. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 143–148, Online. Association for Computational Linguistics.

Vinit Ravishankar, Andrey Kutuzov, Lilja Øvrelid, and Erik Velldal. 2021. Multilingual ELMo and the effects of corpus sampling. In *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 378–384, Reykjavik, Iceland (Online). Linköping University Electronic Press, Sweden.

Andrey Kutuzov, Jeremy Barnes, Erik Velldal, Lilja Øvrelid, and Stephan Oepen. 2021. Large-scale contextualised language modelling for Norwegian. In *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 30–40, Reykjavik, Iceland (Online). Linköping University Electronic Press, Sweden.

Andrey Kutuzov, Lidia Pivovarova, and Mario Giulianelli. 2021. Grammatical profiling for semantic change detection. In *Proceedings of the 25th Conference on Computational Natural Language Learning*, pages 423–434, Online. Association for Computational Linguistics.

Julia Rodina, Yuliya Trofimova, Andrey Kutuzov, and Ekaterina Artemova. 2021. ELMo and BERT in semantic change detection for Russian. In *Analysis of Images, Social Networks and Texts*, pages 175–186, Cham. Springer International Publishing.

Varvara Logacheva, Denis Teslenko, Artem Shelmanov, Steffen Remus, Dmitry Ustalov, Andrey Kutuzov, Ekaterina Artemova, Chris Biemann, Simone Paolo Ponzetto, and Alexander Panchenko. 2020. Word sense disambiguation for 158 languages using word embeddings only. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 5943–5952, Marseille, France. European Language Resources Association.

Nadezda Katricheva, Alyaxey Yaskevich, Anastasiya Lisitsina, Tamara Zhordaniya, Andrey Kutuzov, and Elizaveta Kuzmenko. 2020. Vec2graph: A Python library for visualizing word embeddings as graphs. In *Analysis of Images, Social Networks and Texts*, pages 190–198, Cham. Springer International Publishing.

Andrey Kutuzov and Mario Giulianelli. 2020. UiO-UvA at SemEval-2020 task 1: Contextualised embeddings for lexical semantic change detection. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 126–134, Barcelona (online). International Committee for Computational Linguistics.

Maria Kunilovskaya, Andrey Kutuzov, and Alistair Plum. 2020. Taxonomy enrichment for Russian: synset classification outperforms linear hyponym-hypernym projections. *Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii: Dialog conference*, pages 459–469.

Andrey Kutuzov, Vadim Fomin, Julia Rodina, and Vladislav Mikhailov. 2020. ShiftRy: web service for diachronic analysis of Russian news. *Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii: Dialog conference*, pages 485–501.

Julia Rodina and Andrey Kutuzov. 2020. RuSemShift: a dataset of historical lexical semantic change in Russian. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1037–1047, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Vadim Fomin, Daria Bakshandaeva, Julia Rodina, and Andrey Kutuzov. 2019. Tracing cultural diachronic semantic shifts in Russian using word embeddings: test sets and baselines. *Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii: Dialog conference*, pages 203–218.

Andrey Kutuzov and Elizaveta Kuzmenko. 2019. To lemmatize or not to lemmatize: How word normalisation affects ELMo performance in word sense disambiguation. In *Proceedings of the First NLPL Workshop on Deep Learning for Natural Language Processing*, pages 22–28, Turku, Finland. Linköping University Electronic Press.

Andrey Kutuzov, Erik Velldal, and Lilja Øvrelid. 2019. One-to-X analogical reasoning on word embeddings: a case for diachronic armed conflict prediction from news texts. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pages 196–201, Florence, Italy. Association for Computational Linguistics.

Julia Rodina, Daria Bakshandaeva, Vadim Fomin, Andrey Kutuzov, Samia Touileb, and Erik Velldal. 2019. Measuring diachronic evolution of evaluative adjectives with word embeddings: the case for English, Norwegian, and Russian. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pages 202–209, Florence, Italy. Association for Computational Linguistics.

Andrey Kutuzov, Mohammad Dorgham, Oleksiy Oliynyk, Chris Biemann, and Alexander Panchenko. 2019b. Making fast graph-based algorithms with graph metric embeddings. In *Proceedings of the 57th Conference of the Association for Computational Linguistics*, pages 3349–3355, Florence, Italy. Association for Computational Linguistics.

Andrey Kutuzov, Mohammad Dorgham, Oleksiy Oliynyk, Chris Biemann, and Alexander Panchenko. 2019a. Learning graph embeddings from WordNet-based similarity measures. In *Proceedings of the Eighth Joint Conference on Lexical and Computational Semantics (*SEM 2019)*, pages 125–135, Minneapolis, Minnesota. Association for Computational Linguistics.

Dmitry Ustalov, Alexander Panchenko, Andrey Kutuzov, Chris Biemann, and Simone Paolo Ponzetto. 2018. Unsupervised semantic frame induction using triclustering. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 55–62. Association for Computational Linguistics.

Andrey Kutuzov and Elizaveta Kuzmenko. 2018. Two centuries in two thousand words: Neural embedding models in detecting diachronic lexical changes. In *Quantitative Approaches to the Russian Language*, pages 95–112. Routledge.

Andrey Kutuzov. 2018. Russian word sense induction by clustering averaged word embeddings. *Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii: Dialog conference*, pages 391–403.

Andrey Kutuzov, Lilja Øvrelid, Terrence Szymanski, and Erik Velldal. 2018. Diachronic word embeddings and semantic shifts: a survey. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1384–1397. Association for Computational Linguistics.

Murhaf Fares, Andrey Kutuzov, Stephan Oepen, and Erik Velldal. 2017. Word vectors, reuse, and replicability: Towards a community repository of large-text resources. In *Proceedings of the 21st Nordic Conference on Computational Linguistics*, pages 271–276. Association for Computational Linguistics.

Andrey Kutuzov and Elizaveta Kuzmenko. 2017. Webvectors: A toolkit for building web interfaces for vector semantic models. In *Analysis of Images, Social Networks and Texts: 5th International Conference, AIST 2016, Yekaterinburg, Russia, April 7-9, 2016, Revised Selected Papers*, pages 155–161, Cham. Springer International Publishing.

Maria Kunilovskaya and Andrey Kutuzov. 2017. Universal dependencies-based syntactic features in detecting human translation varieties. In *Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories*, pages 27–36.

Andrey Kutuzov, Erik Velldal, and Lilja Øvrelid. 2017. Tracing armed conflicts with diachronic word embedding models. In *Proceedings of the Events and Stories in the News Workshop at ACL 2017*, pages 31–36, Vancouver, Canada.

Andrey Kutuzov, Erik Velldal, and Lilja Øvrelid. 2017. Temporal dynamics of semantic relations in word embeddings: an application to predicting armed conflict participants. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1824–1829, Copenhagen, Denmark.

Andrey Kutuzov and Maria Kunilovskaya. 2017. Size vs. structure in training corpora for word embedding models: Araneum Russicum Maximum and Russian National Corpus. In *Analysis of Images, Social Networks and Texts*, pages 47–58, Cham. Springer International Publishing.

Pierre Lison and Andrey Kutuzov. 2017. Redefining context windows for word embedding models: An experimental study. In *Proceedings of the 21st Nordic Conference on Computational Linguistics*, pages 284–288. Association for Computational Linguistics.

Andrey Kutuzov, Elizaveta Kuzmenko, and Lidia Pivovarova. 2017. Clustering of Russian adjective-noun constructions using word embeddings. In *Proceedings of the 6th Workshop on Balto-Slavic Natural Language Processing*, pages 3–13. Association for Computational Linguistics.

Andrey Kutuzov. 2017. Arbitrariness of linguistic sign questioned: correlation between word form and meaning in russian. *Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii: Dialog conference*, 1(16 (23)):109–120.

Andrey Kutuzov, Erik Velldal, and Lilja Øvrelid. 2016. Redefining part-of-speech classes with distributional semantic models. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, pages 115–125. Association for Computational Linguistics.

Andrey Kutuzov and Elizaveta Kuzmenko. 2016. Neural embedding language models in semantic clustering of web search results. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France. European Language Resources Association (ELRA).

Andrey Kutuzov and Olga Koslowa. 2016. Improving distributional semantic models using anaphora resolution during linguistic preprocessing. In *Computational Linguistics and Intellectual Technologies: papers from the Annual conference "Dialogue"*, pages 288–300. Izdatelskij centr RGGU.

Andrey Kutuzov, Elizaveta Kuzmenko, and Anna Marakasova. 2016. Exploration of register-dependent lexical semantics using word embeddings. In *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*, pages 26–34, Osaka, Japan.

Andrey Kutuzov and Elizaveta Kuzmenko. 2016. Cross-lingual trends detection for named entities in news texts with dynamic neural embedding models. In *First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval (ECIR 2016)*, pages 27–32. Technical University of Aachen.

Andrey Kutuzov, Mikhail Kopotev, Tatyana Sviridenko, and Lyubov Ivanova. 2016. Clustering comparable corpora of Russian and Ukrainian academic texts: Word embeddings and semantic fingerprints. In *Ninth Workshop on Building and Using Comparable Corpora*.

Andrey Kutuzov and Igor Andreev. 2015. Texts in, meaning out: neural language models in semantic similarity task for Russian. *Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii: Dialog conference*, 1(14 (21)).

Andrey Kutuzov and Elizaveta Kuzmenko. 2015. Comparing neural lexical models of a classic national corpus and a web corpus: The case for Russian. *Lecture Notes in Computer Science*, 9041:47–58.

Andrey Kutuzov and Maria Kunilovskaya. 2014. Russian learner translator corpus. In *Text, Speech and Dialogue*, pages 315–323, Cham. Springer International Publishing.

Andrey Kutuzov and Max Ionov. 2014. The impact of morphology processing quality on automated anaphora resolution for Russian. *Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii: Dialog conference*, 1(13 (20)).