# Ekaterina Kochmar

1 WEST 3.55
University of Bath
Claverton Down
Bath BA2 7PB

Email: ek762bath.ac.uk
Homepage: researchportal.bath.ac.uk/en/persons/ekaterina-kochmar
github.com/ekochmar

## Research Interests

- Areas: natural language processing, artificial intelligence, machine learning, data science, computational linguistics
- Keywords: machine learning applications, intelligent tutoring systems, computational semantics, statistical natural language processing, models of natural language readability, text simplification, summarisation, language testing and assessment, error detection and correction, author profiling

## Education

### ***University of Cambridge***, *Cambridge, UK*

PhD in Computer Science, 2011 - 2016

- PhD Dissertation: "Error Detection in Content Word Combinations"
- Supervisor: Prof. Ted Briscoe

### ***University of Cambridge***, *Cambridge, UK*

MPhil in Advanced Computer Science, 2010 - 2011 (84.1, distinction)

- MPhil Dissertation: "Identification of a Writer's Native Language by Error Analysis"
- Supervisor: Prof. Ted Briscoe

### ***University of Tuebingen***, *Tuebingen, Germany*

M.A. in Computational Linguistics, 2008 - 2010 (GPA: 1.0, excellent)

- M.A. Thesis: "Ensemble-Based Learning for Morphological Analysis of German"
- Advisors: Prof. E. Hinrichs, Dr. D. Gerdemann

### ***St. Petersburg State University***, *St. Petersburg, Russia*

Diploma in Mathematical and Applied Linguistics, 2003 - 2008 (GPA: 5.0, excellent)

- Thesis: "Functional syntactic analysis of prepositional and causal constructions for a grammatical parser of Russian"
- Advisor: Associate professor I.V. Azarova

# Employment

***Department of Computer Science**, University of Bath, UK*

- Lecturer (Assistant Professor), February 2021 - present

***Korbit AI**, Cambridge, UK / Montreal, Canada*

- Co-founder and CSO, January 2017 - present

***Cambridge Spark**, Cambridge, UK*

- Senior Teaching Fellow, January 2015 - present

***Department of Computer Science and Technology**, University of Cambridge, UK*

- Affiliated Lecturer, October 2018 - January 2021
- Senior Research Associate, February 2018 - January 2021
- Research Associate, October 2014 - January 2018

***Trinity College**, University of Cambridge, UK*

- Postdoctoral Teaching Assistant, January 2019 - January 2021

***Daimler AG**, Research Center, Ulm, Germany*

- Software Developer in a project on Natural Language Processing, Data Mining and Text Mining for social media, May - August 2010
- Internship, March - April 2010

***ABBYY**, St. Petersburg, Russia*

- Linguist in a project on Machine Translation in the economic domain, October 2007 - June 2008

***Softissimo**, outsource project, St. Petersburg, Russia*

- Linguist in a project on Machine Translation, June 2006 - September 2007

# Publications

- Ekaterina Kochmar (2021). *Getting Started with Natural Language Processing*. In preparation; to be published by Manning Publications, ISBN 9781617296765
- Ekaterina Kochmar, Dung Do Vu, Robert Belfer, Varun Gupta, Iulian Vlad Serban, and Joelle Pineau (2021). *Automated Generation of Personalized Pedagogical Interventions in Intelligent Tutoring Systems*. International Journal of Artificial Intelligence in Education (IJAIED); in preparation

- Francois St-Hilaire, Nathan Burns, Robert Belfer, Muhammad Shayan, Ariella Smofsky, Dung Do Vu, Antoine Frau, Joseph Potochny, Farid Faraji, Vincent Pavero, Neroli Ko, Ansona Onyi Ching, Sabina Elkins, Anush Stepanyan, Adela Matajova, Laurent Charlin, Yoshua Bengio, Iulian Vlad Serban, and Ekaterina Kochmar. *A Comparative Study of Learning Outcomes for Online Learning Platforms*. Accepted at the 22nd International Conference on Artificial Intelligence in Education (AIED 2021)

- Sian Gooding, Ekaterina Kochmar, Seid Muhie Yimam, and Chris Biemann (2021). *Word Complexity is in the Eye of the Beholder*. Accepted at the 2021 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT 2021)

- Matt Grenander, Robert Belfer, Ekaterina Kochmar, Iulian Serban, François St-Hilaire, and Jackie Cheung (2021). *Deep Discourse Analysis for Generating Personalized Feedback in Intelligent Tutor Systems*. In Proceedings of the 11th Symposium on Educational Advances in Artificial Intelligence (EAAI-21)

- Shiva Taslimipoor, Sara Bahaadini, and Ekaterina Kochmar (2020). *MTLB-STRUCT @PARSEME 2020: Capturing Unseen Multiword Expressions Using Multi-task Learning and Pre-trained Masked Language Models*. In Proceedings of the Joint Workshop on Multiword Expressions and Electronic Lexicons (MWE-LEX 2020, COLING 2020)

- Iulian Vlad Serban, Varun Gupta, Ekaterina Kochmar, Dung D. Vu, Robert Belfer, Joelle Pineau, Aaron Courville, Laurent Charlin, and Yoshua Bengio (2020). *A Large-Scale, Open-Domain, Mixed-Interface Dialogue-Based ITS for STEM*. In Proceedings of the 21st International Conference on Artificial Intelligence in Education (AIED 2020)

- Ekaterina Kochmar, Dung Do Vu, Robert Belfer, Varun Gupta, Iulian Vlad Serban, and Joelle Pineau (2020). *Automated Personalized Feedback Improves Learning Gains in an Intelligent Tutoring System*. In Proceedings of the the 21st International Conference on Artificial Intelligence in Education (AIED 2020). A system based on this research is **deployed** in Korbi (`https://korbit.ai`)

- Sian Gooding, Shiva Taslimipoor, and Ekaterina Kochmar (2020). *Incorporating Multiword Expressions in Phrase Complexity Estimation*. In Proceedings of the 1st Workshop on Tools and Resources to Empower People with REAding DIfficulties (READI 2020)

- Ekaterina Kochmar, Sian Gooding, and Matthew Shardlow (2020). *Detecting Multiword Expression Type Helps Lexical Complexity Assessment*. In Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)

- David Strohmaier, Sian Gooding, Shiva Taslimipoor, and Ekaterina Kochmar (2020). *SeCoDa: Sense Complexity Dataset*. In Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)

- Sian Gooding and Ekaterina Kochmar (2019). *Recursive Context-Aware Lexical Simplification*. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP 2019)

- Sian Gooding and Ekaterina Kochmar (2019). *Complex Word Identification as a Sequence Labelling Task*. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019)

- Sian Gooding, Ekaterina Kochmar, Advait Sarkar, and Alan Blackwell (2019). *Comparative judgments are more consistent than binary classification for labelling word complexity*. In Proceedings of the 13th Linguistic Annotation Workshop (LAW XIII, ACL 2019)

- Menglin Xia, Ekaterina Kochmar, and Ted Briscoe (2019). *Automatic learner summary assessment for reading comprehension*. In Proceedings of the 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019)

- Sian Gooding and Ekaterina Kochmar (2018). *CAMB at CWI Shared Task 2018: Complex Word Identification with Ensemble-Based Voting*. In Proceedings of the 13th Workshop on Innovative Use of NLP for Building Educational Applications (BEA13, NAACL HLT 2018). This paper presents the **winning submission** to the shared task on Complex Word Identification.

- Ekaterina Kochmar and Ekaterina Shutova (2017). *Modelling semantic acquisition in second language learning*. In Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications (BEA12, EMNLP 2017)

- Zafar Gilani, Ekaterina Kochmar, and Jon Crowcroft (2017). *Classification of Twitter Accounts into Automated Agents and Human Users*. In Proceedings of the 9th IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2017). This research is **reviewed** in Scientific American and International Business Times.

- Aurelie Herbelot and Ekaterina Kochmar (2016). *'Calling on the classical phone': a distributional model of adjective-noun errors in learners' English*. In Proceedings of the 26th International Conference on Computational Linguistics (COLING 2016)

- Ekaterina Kochmar and Ekaterina Shutova (2016). *Cross-Lingual Lexico-Semantic Transfer in Language Learning*. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016)

- Menglin Xia, Ekaterina Kochmar, and Ted Briscoe (2016). *Text Readability Assessment for Second Language Learners*. In Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications (BEA11, NAACL 2016). A system based on this research is **deployed** in Read & Improve (`https://readandimprove.englishlanguageitutoring.com`)

- Ekaterina Kochmar (2015). *Error Detection in Content Word Combinations*. Technical report UCAM-CL-TR-886, Computer Laboratory, University of Cambridge, ISSN 1476-2986

- Ekaterina Kochmar and Ted Briscoe (2015). *Using Learner Data to Improve Error Correction in Adjective–Noun Combinations*. In Proceedings of the 10th Workshop on Innovative Use of NLP for Building Educational Applications (BEA10, NAACL 2015)

- Ekaterina Kochmar and Ted Briscoe (2014). *Detecting Learner Errors in the Choice of Content Words Using Compositional Distributional Semantics*. In Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers (COLING 2014)

- Mariano Felice, Zheng Yuan, Øistein Andersen, Helen Yannakoudakis, and Ekaterina Kochmar (2014). *Grammatical error correction using hybrid systems and type filtering*. In Proceedings of the 17th Conference on Computational Natural Language Learning (CoNLL 2014): Shared Task. This paper presents the **winning submission** to the shared task on Grammatical Error Correction.

- Ekaterina Kochmar and Ted Briscoe (2013). *Capturing Anomalies in the Choice of Content Words in Compositional Distributional Semantic Space*. In Proceedings of the Recent Advances in Natural Language Processing (RANLP 2013)

- Ekaterina Kochmar, Øistein Andersen, and Ted Briscoe (2012). *HOO 2012 Error Recognition and Correction Shared Task: Cambridge University Submission Report*. In Proceedings of the 7th Workshop on Innovative Use of NLP for Building Educational Applications (BEA7, NAACL-HLT 2012)

- Ekaterina Kochmar (2011). *Identification of a Writer's Native Language by Error Analysis*. MPhil Dissertation, Computer Laboratory, University of Cambridge, UK

- Ekaterina Kochmar (2010). *Ensemble-Based Learning for Morphological Analysis of German*. MA Thesis, University of Tuebingen, Germany

# Teaching Experience

- **Lecturing (academia):**
  - **Developing course material and lecturing**, Data Science: Principles and Practice, Part II CST course, Department of Computer Science and Technology, University of Cambridge, 2018 - 2021. The course received **positive feedback** from the students and is one of the three **most popular** courses in Part II of the Computer Science Tripos.
  - **Lecturing**, Introduction to Machine Learning, Schmidt Data for Science Residency Programme (part of the Accelerate Programme for Scientific Discovery), Department of Computer Science and Technology, University of Cambridge, 2020
  - **Giving guest lectures**, Information Retrieval, Part II course, Department of Computer Science and Technology, University of Cambridge, 2017 - 2018
  - **Giving guest lectures**, Machine Learning for Language Processing, Advanced Computer Science MPhil course, Department of Computer Science and Technology, University of Cambridge, 2013 - 2018
  - **Part-time lecturing**, Computational Linguistics, Part II course, Department of Theoretical and Applied Linguistics, University of Cambridge, 2017
  - **Lecturing**, Advanced Topics in Natural Language Processing, Advanced Computer Science MPhil course, Department of Computer Science and Technology, University of Cambridge, 2015 - 2017
- **Project supervision (academia):**
  - **Supervising**, MSc projects on Repairing Misunderstandings between Humans and AI Conversational Agents, Department of Computer Science, University of Bath, 2021 - present.
  - **Supervising**, MSc projects on Automated Generation of Knowledge Assessment Items, Department of Computer Science, University of Bath, 2021 - present.
  - **Co-supervising**, PhD project on Text Simplification, Department of Computer Science and Technology, University of Cambridge, 2018 - 2021.
  - **Supervising**, Advanced Computer Science MPhil project on Graphical concept representations for prerequisite and explanation extraction from textbooks, Department of Computer Science and Technology, University of Cambridge, 2020 - 2021.
  - **Supervising**, Part II (undergraduate) project on Automated Alignment in Parallel Texts, Department of Computer Science and Technology, University of Cambridge, 2020 - 2021.
  - **Supervising**, Part II (undergraduate) project on Exploring Memory Organisation Structures used in Memory Networks, University of Cambridge, 2020 - 2021.
  - **Supervising**, cross-disciplinary projects on applications of data science for scientific discovery, Schmidt Data for Science Residency Programme (part of the Accelerate Programme for Scientific Discovery), Department of Computer Science and Technology, University of Cambridge, 2020
  - **Supervising**, Advanced Computer Science MPhil project on Complex Word Identification for Readability Assessment, Department of Computer Science and Technology, University of Cambridge, 2019 - 2020. The student got **distinction** on the project.
  - **Supervising**, Advanced Computer Science MPhil project on Lexical Text Adaptation for Different Readability Levels, Department of Computer Science and Technology, University of Cambridge, 2019 - 2020. The student got **distinction** on the project.
  - **Co-supervising**, PhD project on Text Readability Assessment for Second Language Learners, Department of Computer Science and Technology, University of Cambridge, 2015 - 2018. The student passed the viva with **minor corrections**.

  - **Supervising**, Advanced Computer Science MPhil project on Complex Word Identification, Department of Computer Science and Technology, University of Cambridge, 2017 - 2018. The student got **distinction** on the project and received a **Google prize award** for the thesis.
  - **Supervising**, Advanced Computer Science MPhil project on Headline Generation, Department of Computer Science and Technology, University of Cambridge, 2017 - 2018
  - **Supervising**, Part II (undergraduate) project on Study Helper based on Natural Language Processing, Department of Computer Science and Technology, University of Cambridge, 2017 - 2018. The student got a **first-class mark** on the project.
  - **Supervising**, Advanced Computer Science MPhil project on Incorporating L1 Detection into Error Detection in English Text Produced by EFL Learners, Department of Computer Science and Technology, University of Cambridge, 2016 - 2017. The student got **distinction** on the project.
  - **Supervising**, Advanced Computer Science MPhil project on Improving and Evaluating Methods for Automatic Factual Question Generation, Department of Computer Science and Technology, University of Cambridge, 2016 - 2017. The student got **distinction** on the project.
  - **Supervising**, Part II (undergraduate) project on Directory Structure Generation Using Information Retrieval Techniques, Department of Computer Science and Technology, University of Cambridge, 2016 - 2017. The student got a **first-class mark** on the project.
  - **Co-supervising**, Advanced Computer Science MPhil project on Reverse Dictionary Search using Neural Network Embeddings for Vocabulary Acquisition, Department of Computer Science and Technology, University of Cambridge, 2015 - 2016
  - **Co-supervising**, Advanced Computer Science MPhil project on Error Detection for Prepositions and Articles in ESL Text as Discriminative Classification, Department of Computer Science and Technology, University of Cambridge, 2011 - 2012. The student got **distinction** on the project.

- **Undergraduate course supervision (academia):**

  - **Supervising**, Further Human–Computer Interaction, Part IB CST course, Department of Computer Science and Technology, University of Cambridge, 2020 - 2021
  - **Supervising**, Interaction Design, Part IA CST course, Department of Computer Science and Technology, University of Cambridge, 2019 - 2021
  - **Supervising**, Foundations of Data Science, Part IB CST course, Department of Computer Science and Technology, University of Cambridge, 2018 - 2021
  - **Supervising**, Formal Models of Language, Part IB CST course, Department of Computer Science and Technology, University of Cambridge, 2017 - 2021
  - **Supervising**, Machine Learning and Real-world Data, Part IA CST course, Department of Computer Science and Technology, University of Cambridge, 2016 - 2021. I received **positive feedback** from the students submitted through the student self-assessment forms.
  - **Supervising**, Natural Language Processing, Part II course, Department of Computer Science and Technology, University of Cambridge, 2011 - 2017
  - **Supervising**, Information Retrieval, Part II course, Department of Computer Science and Technology, University of Cambridge, 2012 - 2016

- **Lecturing (industry):**

  - **Developing and teaching**, Natural Language Processing with Python, Advanced Data Science course, Cambridge Spark, Cambridge / London, 2017 - present
  - **Developing and teaching**, Natural Language Processing with Python, Corporate Training for Morgan Stanley, Cambridge Spark, New York (remotely), 2020

  - **Teaching**, Introduction to Machine Learning, Corporate Training for Morgan Stanley, Cambridge Spark, London (remotely), 2020
  - **Teaching**, Introduction to Python, Advanced Data Science course, Cambridge Spark, Warwick Business School (remotely), 2020
  - **Developing and teaching**, Text Mining and Natural Language Processing with Python, 2-days course, Cambridge Spark, Cambridge, 2016 - 2017
  - **Teaching**, Machine Learning and Data Science Bootcamp, Cambridge Coding Academy, Cambridge, 2016
  - **Teaching**, Workshop on Machine Learning with Python, Cambridge Coding Academy, 2015 - 2016

- **Project supervision (industry):**
  - Via Cambridge Spark, I have **supervised** a variety of real-world machine learning projects with industry partners such as Cytora, Wavelength, Zoopla, Ernst & Young, and HBO, among others, and **advised** on an academic project at UCL, Developmental Biology and Cancer Department.

- **Other:**
  - **Interviewing** for Computer Science Tripos Admissions, Trinity College, University of Cambridge, 2019 - 2021
  - **Developing and assessing** test questions for Part IA CST Machine Learning and Real-world Data section of the College test, St John's College, University of Cambridge, 2017 - 2018
  - **Teaching**, Raspberry Pi Practical Session, Women into Mathematics, Year 10 Subject Day, St John's College, University of Cambridge, 2015
  - **Demonstrating**, Syntax and Semantics of Natural Language, Advanced Computer Science MPhil course, Department of Computer Science and Technology, University of Cambridge, 2013 - 2015

## Presentations and Invited Talks

- Panelist on "Education and AI: solutions, challenges and critical perspectives" ("Éducation et IA: solutions, enjeux et perspectives critiques"), Centre of Pedagogy, University of Montreal , April 2021
- Invited speaker at the AAAI 2021 Spring Symposium on AI in K-12 Education, vitrual, March 2021
- Keynote speaker at live@Manning, a virtual conference by Manning Publications, August 2020
- Keynote speaker at Allianz SE NLP meetup, Munich, Germany, scheduled for 2020
- Keynote at PyData, Cambridge, UK, November 2019
- Invited talk at JPMorgan Chase AI Research Paper Club, London, UK, May 2019
- Invited talk at Psychometrics Seminar, Cambridge, UK, May 2019
- Invited talk at ARM, Cambridge, UK, January 2019
- Invited talk at the Sutton Trust Summer School, Cambridge, UK, August 2018
- Workshop organisation at Cambridge Artificial Intelligence Summit 2018, Cambridge, UK, June 2018
- Invited talk at the ALMAnaCH, Inria, Paris, France, May 2018
- Invited talk at the UCL Linguistics Society, London, UK, February 2018

- Invited talk at AstraZeneca, Cambridge, UK, November 2017
- Workshop organisation at London Data Science Summit, London, UK, October 2017
- Invited talk at the Sutton Trust Summer School, Cambridge, UK, August 2017
- Invited talk at the Cambridge Coding Academy Summer School, Cambridge, UK, July 2017
- Invited talk at the Summer School in Machine Learning for Digital English Language Teaching, Chania, Greece, July 2017
- Invited talk at the Technology Ventures Conference, Cambridge, UK, June 2017
- Invited talk at Cambridge Data Science Summit, Cambridge, UK, June 2017
- Invited talk at Microsoft Research, Cambridge, UK, April 2017
- Invited talk at the University of Copenhagen, Copenhagen, Denmark, October 2016
- Poster presentation at Deep Learning Summer School, Montreal, Canada, August 2016
- Invited talk at Cambridge Data Science Bootcamp, Cambridge, UK, March 2016
- Invited talk at the University of Warwick, Warwick, UK, October 2015

## Other Professional Activities

- **Conference organisation**:
  - Reviewer for the top-ranked international conferences in the field, including ACL, NAACL, EMNLP, AAAI, COLING, BEA, LREC, *SEM, AIST; ACL workshops. I am included in the list of ACL 2018, EMNLP 2018, NAACL 2019 **top reviewers**.
  - Reviewer for the following journals: ACM Computing Surveys; ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP); ACM Transactions on Information Systems (TOIS); Computational Linguistics (standing reviewer); Computer Speech and Language; IEEE Transactions on Knowledge and Data Engineering (TKDE); IEEE Transactions on Learning Technologies (TLT); International Journal of Artificial Intelligence in Education; International Journal of Speech, Language and the Law; Journal of the Association for Information Science and Technology; Journal of Cognitive Science; System
  - Area chair for ACL 2019
  - Secretary of the ACL Special Interest Group on Educational Applications (SIGEDU), 2017 - present
  - Co-organiser and committee member of the Workshop on Innovative Use of NLP for Building Educational Applications (BEA), 2015 - present
  - ACL Student Research Workshop mentor, 2018-2019
  - ACL-2014 Student Research Workshop co-chair, Baltimore, 2014
  - Co-organiser of the Oxbridge Women in Computer Science Conferences, Cambridge/Oxford, 2014-2016
- **Other**:
  - Postdoc representative to the Postdoc Forum, 2015 - 2019
  - Organising STEM Subject Day, St John's College, 2017-2018

- women@cl committee member and technical seminars organiser, 2014 - 2016
- Demonstrating at the Launch of the University Fund Raising Campaign, 2015
- Demonstrating at Raspberry Pi Practical Session, STEM Subject Day, 2015
- NLIP group research seminars organiser, 2011 - 2013
- MPhil student representative to the Graduate Students' Forum, 2010 - 2011

## Grant applications

- Ekaterina Kochmar, Sian Gooding, and Elaine Schmidt. *Eye-tracking methods for lexical complexity assessment*, **Language Sciences Research Incubator Fund**, 2020
- **NVidia** GPU Grant, 2019

## Awards and Scholarships

- **St John's College** Scholarship, 2011 - 2014
- **Cambridge Overseas Trust** Scholarship, 2010 - 2014
- German Academic Exchange Service (**DAAD**) Scholarship, 2008 - 2010
- Scholarship from the **Vladimir Potanin Foundation**, 2005 - 2007

Last updated: April 15, 2021