LChange 2024

5th International Workshop on Computational Approaches to Historical Language Change 2024

Proceedings of the Workshop

August 15, 2024

©2024 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 979-8-89176-138-4

Preface by the General Chair

Welcome to the 5th International Workshop on Computational Approaches to Historical Language Change (LChange'24) co-located with ACL 2024. LChange is held on August 15th, 2024, as a hybrid event with participation possible both virtually and on-site in Thailand.

Characterizing the time-varying nature of language will have broad implications and applications in multiple fields including linguistics, artificial intelligence, digital humanities, computational cognitive and social sciences. In this workshop, we bring together the world's pioneers and experts in **computational approaches to historical language change with a focus on digital text corpora**. In doing so, this workshop carries out the triple goals of disseminating state-of-the-art research on diachronic modeling of language change, fostering cross-disciplinary collaborations, and exploring the fundamental theoretical and methodological challenges in this growing niche of computational linguistic research.

In response to the call, we received 24 submissions. Each of them was carefully evaluated by at least two members of the Program Committee, whom we believed to be most appropriate for each paper. Based on the reviewers' feedback we accepted 17 full and short papers as oral or poster presentations. We had two distinguished keynote presentations: the first by Antske Fokkens (Professor at the Computational Linguistics and Text Mining Lab at the Vrije Universiteit Amsterdam, Netherlands) who presented a talk entitled "What Changes in Language Modeling mean for Modeling Language Change", and the second by Johann-Mattis List (Professor and Chair of Multilingual Computational Linguistics at the University of Passau, Germany) with the talk "New Approaches in Computer-Assisted Language Comparison". Finally, we invited two ACL'24 Findings papers to be presented at the workshop, which are not included in the workshop proceedings.

We hope that you will find the workshop papers insightful and inspiring. We would like to thank the keynote speakers for their stimulating talks, the authors of all papers for their interesting contributions, and the members of the Program Committee for their insightful reviews. Our special thanks go to the emergency reviewers who stepped in to provide their expertise. We also express our gratitude to the ACL 2024 workshop chairs for their kind assistance during the organization process. Finally, our thanks go to our sponsors, the research program "Change is Key!" (Riksbankens Jubileumsfond, contract M21-0021).

Nina Tahmasebi, chair, University of Gothenburg (Sweden)
Syrielle Montariol, EPFL (Switzerland)
Andrey Kutuzov, University of Oslo (Norway)
David Alfter, University of Gothenburg (Sweden)
Francesco Periti, University of Milan (Italy)
Pierluigi Cassotti, University of Gothenburg (Sweden)
Netta Huebscher, University of Gothenburg (Sweden)
LChange'24 Workshop Chairs

Organizing Committee

General Chair

Nina Tahmasebi, University of Gothenburg, Sweden

Program Chairs

Syrielle Montariol, École polytechnique fédérale de Lausanne, Switzerland
Andrey Kutuzov, University of Oslo, Norway
David Alfter, University of Gothenburg, Sweden
Francesco Periti, University of Milan, Italy
Pierluigi Cassotti, University of Gothenburg, Sweden
Netta Huebscher, University of Gothenburg, Sweden

Program Committee

Program Chairs

David Alfter, University of Gothenburg
Pierluigi Cassotti, University of Gothenburg
Netta Huebscher, University of Gothenburg
Andrey Kutuzov, University of Oslo
Syrielle Montariol, Ecole Polytechnique Fédérale de Lausanne
Francesco Periti, University of Milan
Nina Tahmasebi, University of Gothenburg

Reviewers

Pierpaolo Basile, University of Bari
Pierluigi Cassotti, University of Gothenburg
Jing Chen, Hong Kong Polytechnic University
Stefano De Pascale, Vrije Universiteit Brussel and KU Leuven
Haim Dubossarsky, Queen Mary University of London and University of Cambridge
Mariia Fedorova, University of Oslo
Karlien Franco, QLVL | KU Leuven & FWO Vlaanderen
Mario Giulianelli, University of Amsterdam
Mauricio Gruppi, Villanova University
Vaibhav Jain, Delhi Technological University
Andres Karjus, Tallinn University
Andrey Kutuzov, University of Oslo
Barbara McGillivray, King's College London, University of London
Timothee Mickus, University of Helsinki
Filip Miletić, University of Stuttgart
Syrielle Montariol, Ecole Polytechnique Fédérale de Lausanne
Pablo Mosteiro, Utrecht University
Krzysztof Nowak, Polish Academic of Science
Lidia Pivovarova, University of Helsinki
Martin Pömsl, School of Computer Science, McGill University
Ella Rabinovich, International Business Machines
Martin Ruskov, University of Milan
Dominik Schlechtweg, Institute for Natural Language Processing, University of Stuttgart
Taichi Aida, Tokyo Metropolitan University
Ludovic Tanguy, University of Toulouse
Stephen Eugene Taylor, University of West Bohemia

Keynote Talk: What Changes in Language Modeling mean for Modeling Language Change

Antske Fokkens
Vrije Universiteit Amsterdam

Abstract: Language change detection has emerged as a subdomain that has caught the interest (computational) linguistics, historians, social scientists and computer scientists. Despite this enthusiasm and stable attention from the NLP community over multiple years, our methods keep on having difficulties in distinguishing valid signals of change from noise. This holds both for methods using static word embeddings as well as for more recent explorations with methods that make use of contextual embeddings. The question of how to distinguish true signal from noise has received substantial attention from the field, with the design of benchmarks, control tests and artificially created samples and data. An aspect that has, to my knowledge, received less attention is the fundamental differences between most methods using static on the one hand, and most methods using contextualized embeddings on the other hand. Mainly, methods that make use of static embeddings involve creating new embeddings for the full vocabular creating general shifts in space.
Methods using contextualized embeddings on the other hand mostly make use of pretrained language models, either as is or with some continual training on the target corpus. Change is then studied by comparing instances including target terms from different corpora. In this talk, I will explore what these fundamental differences mean when carrying out methodological checks and balances for studying language change with the aim of answering the question: how can we find meaningful change and know that is meaningful.

Bio: Antske Fokkens is a researcher at the Computational Lexicology and Terminology Lab and a visiting researcher at the Web and Media group at VU University Amsterdam, where she is also part of the Network Institute. Her main interest lies in the methodological aspects of Computational Linguistics, particularly how computational models of language work and which methods are suitable for modeling or analyzing linguistic phenomena. Her recent work focuses on applying NLP to digital humanities, enhancing historical research through the BiographyNet project. Additionally, she addresses methodological issues in system architecture and large-scale news processing through projects like NewsReader and Can we Handle the News. Her PhD thesis proposed a methodology for developing linguistic precision grammars, applicable across various theories.

Keynote Talk: New Approaches in Computer-Assisted Language Comparison

Johan-Mattis List
University of Passau

Abstract: The field of computer-assisted language comparison seeks to develop interactive computational workflows that facilitate those tasks that linguists working in the field of historical or typological language comparison usually carry out manually. While the field has substantially grown over the past decade, with new tools and new workflows that support computer-assisted analyses, there remain many challenges that have so far not yet been addressed in computer-assisted approaches. In this study, three new approaches that facilitate detailed comparative analysis will be presented. The first approach allows for an efficient manual labeling of correspondence patterns in comparative wordlists, the second approach allows to group sounds in phonetically transcribed wordlists and to segment words into morphemes. The third approach allows to correct individual word forms in comparative wordlists, by contrasting the reflexes of a proto-form that one would expect under the assumption of regular sound change with the reflexes that are attested in the data. All approaches are implemented in an interactive web-based tool that is freely available and integrated with previous computer-assisted tools and workflows.

Bio: Johan-Mattis List is a comparative linguist and Chair for Multilingual Computational Linguistics since January 2023, leading the ERC-funded ProduSemyresearch group. Previously, he was a stand-in professor at Bielefeld University and a senior researcher at the Max Planck Institutes in Leipzig and Jena. He earned his doctorate at Heinrich Heine University in Düsseldorf and completed his habilitation at Friedrich Schiller University in Jena. His research focuses on the evolution of human language lexicons and language change, with particular interest in Southeast Asian and South American languages. He advocates for open research and draws inspiration from bioinformatics to improve language comparison methods.

Table of Contents

***Invited paper**: Computer-Assisted Language Comparison with EDICTOR 3*
Johann-Mattis List and Kellen Parker van Dam 1

Exploring Diachronic and Diatopic Changes in Dialect Continua: Tasks, Datasets and Challenges
Melis Çelikkol, Lydia Körber and Wei Zhao 12

Similarity-Based Cluster Merging for Semantic Change Modeling
Christopher Brückner, Leixin Zhang and Pavel Pecina 23

Historical Ink: Semantic Shift Detection for 19th Century Spanish
Tony Montes, Laura Manrique-Gómez and Rubén Manrique 29

Presence or Absence: Are Unknown Word Usages in Dictionaries?
Xianghe Ma, Dominik Schlechtweg and Wei Zhao 42

Towards a GoldenHymns Dataset for Studying Diachronic Trends in 19th Century Danish Religious Hymns
Ea Lindhardt Overgaard, Pascale Feldkamp and Yuri Bizzoni 55

A Feature-Based Approach to Annotate the Syntax of Ancient Chinese
Chenrong Zhao 62

AXOLOTL'24 Shared Task on Multilingual Explainable Semantic Change Modeling
Mariia Fedorova, Timothee Mickus, Niko Partanen, Janine Siewert, Elena Spaziani and Andrey Kutuzov 72

Improving Word Usage Graphs with Edge Induction
Bill Noble, Francesco Periti and Nina Tahmasebi 92

Towards a Complete Solution to Lexical Semantic Change: an Extension to Multiple Time Periods and Diachronic Word Sense Induction
Francesco Periti and Nina Tahmasebi 108

TartuNLP @ AXOLOTL-24: Leveraging Classifier Output for New Sense Detection in Lexical Semantics
Aleksei Dorkin and Kairit Sirts 120

EtymoLink: A Structured English Etymology Dataset
Yuan Gao and Weiwei Sun 126

Complexity and Indecision: A Proof-of-Concept Exploration of Lexical Complexity and Lexical Semantic Change
David Alfter 137

Can political dogwhistles be predicted by distributional methods for analysis of lexical semantic change?
Max Boholm, Björn Rönnerstrand, Ellen Breitholtz, Robin Cooper, Elina Lindgren, Gregor Rettenegger and Asad Sayeed 144

Towards an Onomasiological Study of Lexical Semantic Change Through the Induction of Concepts
Bastien Liétard, Mikaela Keller and Pascal Denis 158

Deep-change at AXOLOTL-24: Orchestrating WSD and WSI Models for Semantic Change Modeling
Denis Kokosinskii, Mikhail Kuklin and Nikolay Arefyev 168

Exploring Sound Change Over Time: A Review of Computational and Human Perception
Siqi He and Wei Zhao 180

A Few-shot Learning Approach for Lexical Semantic Change Detection Using GPT-4
Zhengfei Ren, Annalina Caputo and Gareth J. F. Jones 187

Program

Wednesday, December 6, 2023

09:15 - 09:30 *Introduction*

09:30 - 10:30 *Keynote Antske Fokkens*

10:30 - 11:00 *Coffee Break*

11:00 - 12:00 *Session 1*

***Findings paper**: A Semantic Distance Metric Learning approach for Lexical Semantic Change Detection*
Taichi Aida and Danushka Bollegala

Towards a GoldenHymns Dataset for Studying Diachronic Trends in 19th Century Danish Religious Hymns
Ea Lindhardt Overgaard, Pascale Feldkamp and Yuri Bizzoni

***Findings paper**: Definition generation for lexical semantic change detection*
Mariia Fedorova, Andrey Kutuzov and Yves Scherrer

12:00 - 13:00 *Lunch Break*

13:00 - 13:45 *Keynote Johann-Mattis List*

13:45 - 14:45 *Session 2*

Towards an Onomasiological Study of Lexical Semantic Change Through the Induction of Concepts
Bastien Liétard, Mikaela Keller and Pascal Denis

Towards a Complete Solution to Lexical Semantic Change: an Extension to Multiple Time Periods and Diachronic Word Sense Induction
Francesco Periti and Nina Tahmasebi

AXOLOTL'24 Shared Task on Multilingual Explainable Semantic Change Modeling
Mariia Fedorova, Timothee Mickus, Niko Partanen, Janine Siewert, Elena Spaziani and Andrey Kutuzov

Wednesday, December 6, 2023 (continued)

14:45 - 15:30 *Session Poster Pitch*

15:30 - 16:30 *Poster Session*

TartuNLP @ AXOLOTL-24: Leveraging Classifier Output for New Sense Detection in Lexical Semantics
Aleksei Dorkin and Kairit Sirts

Deep-change at AXOLOTL-24: Orchestrating WSD and WSI Models for Semantic Change Modeling
Denis Kokosinskii, Mikhail Kuklin and Nikolay Arefyev

Can political dogwhistles be predicted by distributional methods for analysis of lexical semantic change?
Max Boholm, Björn Rönnerstrand, Ellen Breitholtz, Robin Cooper, Elina Lindgren, Gregor Rettenegger and Asad Sayeed

EtymoLink: A Structured English Etymology Dataset
Yuan Gao and Weiwei Sun

Similarity-Based Cluster Merging for Semantic Change Modeling
Christopher Brückner, Leixin Zhang and Pavel Pecina

Historical Ink: Semantic Shift Detection for 19th Century Spanish
Tony Montes, Laura Manrique-Gómez and Rubén Manrique

Complexity and Indecision: A Proof-of-Concept Exploration of Lexical Complexity and Lexical Semantic Change
David Alfter

Exploring Sound Change Over Time: A Review of Computational and Human Perception
Siqi He and Wei Zhao

A Few-shot Learning Approach for Lexical Semantic Change Detection Using GPT-4
Zhengfei Ren, Annalina Caputo and Gareth J. F. Jones

A Feature-Based Approach to Annotate the Syntax of Ancient Chinese
Chenrong Zhao

Wednesday, December 6, 2023 (continued)

Exploring Diachronic and Diatopic Changes in Dialect Continua: Tasks, Datasets and Challenges
Melis Çelikkol, Lydia Körber and Wei Zhao

Improving Word Usage Graphs with Edge Induction
Bill Noble, Francesco Periti and Nina Tahmasebi

Presence or Absence: Are Unknown Word Usages in Dictionaries?
Xianghe Ma, Dominik Schlechtweg and Wei Zhao

16:30 - 17:30 *Round Table*

17:30 - 17:45 *Closing Remarks*