

LChange 2023

**4th International Workshop on Computational Approaches to
Historical Language Change 2023**

Proceedings of the Workshop

December 6, 2023

The LChange organizers gratefully acknowledge the support from the following sponsors.

Gold



©2023 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 979-8-89176-043-1

Preface by the General Chair

Welcome to the 4th International Workshop on Computational Approaches to Historical Language Change (LChange'23) co-located with EMNLP 2023. LChange is held on December 6th, 2023, as a hybrid event with participation possible both virtually and on-site in Singapore.

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 28 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 Gemma Boleda (Research Professor in the Department of Translation and Language Sciences of the Universitat Pompeu Fabra, Spain) who presented a talk entitled "What does semantic change have to do with Hello Kitty? Referring as the source of change", and the second by Mario Giulianelli (a postdoctoral fellow at ETH Zurich) with the talk "Neural language models for word usage representation and analysis". Finally, we invited five EMNLP'23 Findings papers to be presented as posters, which are not included in the workshop proceedings.

To further support the community, we offered five student scholarships to cover registration fees. We also offered mentoring for four young researchers on their research topic in the field of language change, either during the workshop or virtually.

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 EMNLP 2023 workshop chairs for their kind assistance during the organization process. Finally, our thanks go to our gold sponsor iguanodon.ai, as well as the research project "Towards Computational Lexical Semantic Change Detection" (Swedish Research Council, contract 2018-01184) and the research program "Change is Key!" (Riksbankens Jubileumsfond, contract M21-0021).

Nina Tahmasebi, workshop chair, University of Gothenburg (Sweden)

Syrielle Montariol, EPFL (Switzerland)

Haim Dubossarsky, Queen Mary University of London and University of Cambridge (United Kingdom)

Andrey Kutuzov, University of Oslo (Norway)

Simon Hengchen, iguanodon.ai and University of Geneva (Switzerland)

David Alfter, University of Gothenburg (Sweden)

Francesco Periti, University of Milan (Italy)

Pierluigi Cassotti, University of Gothenburg (Sweden)

LChange'23 Workshop Chairs

Organizing Committee

General Chair

Nina Tahmasebi, University of Gothenburg, Sweden

Program Chairs

Syrielle Montariol, École polytechnique fédérale de Lausanne, Switzerland

Haim Dubossarsky, Queen Mary University of London and University of Cambridge, United Kingdom

Andrey Kutuzov, University of Oslo, Norway

Simon Hengchen, iguanodon.ai and University of Geneva, Switzerland

David Alfter, University of Gothenburg, Sweden

Francesco Periti, University of Milan, Italy

Pierluigi Cassotti, University of Gothenburg, Sweden

Program Committee

Program Chairs

David Alfter, University of Gothenburg
Pierluigi Cassotti, University of Gothenburg
Haim Dubossarsky, Queen Mary University of London and University of Cambridge
Simon Hengchen, iguanodon.ai and University of Geneva
Andrey Kutuzov, University of Oslo
Syrielle Montariol, Ecole Polytechnique Fédérale de Lausanne
Francesco Periti, University of Milan
Nina Tahmasebi, University of Gothenburg

Reviewers

Nikolay Arefyev, University of Oslo
Ehsaneddin Asgari, University of California, Berkeley, Data Lab, Volkswagen Group and Helmholtz Center for Infection Research
Pierpaolo Basile, University of Bari
Christin Beck, Universität Konstanz
Aleksandrs Berdicevskis, Gothenburg University
Pierluigi Cassotti, University of Gothenburg
Paul Cook, University of New Brunswick
Stefano De Pascale, Vrije Universiteit Brussel and KU Leuven
Chiara Di Bonaventura, King's College London, University of London
Clémentine Fourier, HuggingFace
Karlien Franco, QLVL | KU Leuven & FWO Vlaanderen
Mario Giulianelli, University of Amsterdam
Mauricio Gruppi, Villanova University
Simon Hengchen, iguanodon.ai and University of Geneva
Valentin Hofmann, University of Oxford
Abhik Jana, IIT Bhubaneswar
Adam Jatowt, The University of Tokyo, Tokyo Institute of Technology
Andres Karjus, Tallinn University
Andrey Kutuzov, University of Oslo
Barbara McGillivray, King's College London, University of London
Syrielle Montariol, Ecole Polytechnique Fédérale de Lausanne
Paul Nulty, Birkbeck College, University of London and University College Dublin
Lidia Pivovarova, University of Helsinki
Martin Pömsl, School of Computer Science, McGill University
Taraka Rama, University of Gothenburg
Angelika Romanou, Ecole Polytechnique Fédérale de Lausanne
Eyal Sagi, University of St. Francis
Asad B. Sayeed, University of Gothenburg
Dominik Schlechtweg, Institute for Natural Language Processing, University of Stuttgart
Stephen Eugene Taylor, University of West Bohemia
Samia Touileb, University of Bergen, Norway
Ekaterina Vylomova, The University of Melbourne
Melvin Wevers, University of Amsterdam

Frank D. Zamora-Reina, University of Chile

Keynote Talk: What does semantic change have to do with Hello Kitty? Referring as the source of change

Gemma Boleda

University Pompeu Fabra

Abstract: It has long been noted that lexical semantic change is rooted in specific utterances, specific reference acts: for instance, in Old English *deor* ("deer") meant wild animal, and it acquired its current meaning probably via hunting, deer being the favorite animal of the chase".[1] However, traditional historical linguistics lacked the tools to explore the process from reference to semantic change on a large scale. Current methods in computational linguistics, as well as the increasing availability of large-scale linguistic resources, afford precisely that. In this talk, I will present work that links reference to change by examining different phenomena (production of referring expressions, regular polysemy) at different timescales (language development, synchronic use, language evolution), using quantitative and computational methods.

[1] <https://www.etymonline.com/search?q=deer>

Bio: Gemma Boleda is a Research Professor in the Department of Translation and Language Sciences of the Universitat Pompeu Fabra (Barcelona, Spain) and co-director of the Computational Linguistics and Linguistic Theory (COLT) research group. She is a linguist who uses quantitative and computational methods to investigate how language works. She is in particular interested in how people convey meaning through language.

Keynote Talk: Word usage representations from neural language models

Mario Giulianelli
ETH Zurich

Abstract: Neural language models are powerful tools for language scientists interested in studying word usage and its evolution over time. Drawing from a series of recent findings, I will argue that contemporary neural language models can infer contextually appropriate word interpretations which are predictive of human comprehension behaviour, and that they allow for quantitative yet interpretable comparisons between word usages. I will discuss methods to engage with language models for obtaining word representations, including the collection of neural representations generated during the processing of word usage examples, and the direct input of natural language instructions to induce human-readable word definitions. These approaches hold significant relevance for examining shifts and variations in word usage across the temporal and spatial dimensions.

Bio: Mario is a postdoctoral researcher at ETH Zurich, where he works with the Rycolab in the Institute for Machine Learning, Department of Computer Science; and a member of the ELLIS Society. Previously, he was a PhD student at the University of Amsterdam in the Institute for Logic, Language and Computation. He studies language use and evolution using tools from computer science, linguistics, and cognitive science.

Table of Contents

<i>Literary Intertextual Semantic Change Detection: Application and Motivation for Evaluating Models on Small Corpora</i>	
Jackson Ehrenworth and Katherine A. Keith	1
<i>Domain-Adapting BERT for Attributing Manuscript, Century and Region in Pre-Modern Slavic Texts</i>	
Piroska Lendvai, Uwe Reichel, Anna Jouravel, Achim Rabus and Elena Renje	15
<i>Representing and Computing Uncertainty in Phonological Reconstruction</i>	
Johann-Mattis List, Nathan Hill, Robert Forkel and Frederic Blum	22
<i>GHisBERT – Training BERT from scratch for lexical semantic investigations across historical German language stages</i>	
Christin Beck and Marisa Köllner	33
<i>A longitudinal study about gradual changes in the Iranian Online Public Sphere pre and post of ‘Mahsa Moment’: Focusing on Twitter</i>	
Sadegh Jafari, Amin Fathi, Abolfazl Hajizadegan, Amirmohammad Kazemeini and Sauleh Eetemadi	46
<i>Political dogwhistles and community divergence in semantic change</i>	
Max Boholm and Asad B. Sayeed	53
<i>EvoSem: A database of polysemous cognate sets</i>	
Mathieu Dehouck, Alex François, Siva Kalyan, Martial Pastor and David Kletz	66
<i>Multi-lect automatic detection of Swadesh list items from raw corpus data in East Slavic languages</i>	
Iliia Afanasev	76
<i>Anchors in Embedding Space: A Simple Concept Tracking Approach to Support Conceptual History Research</i>	
Jetske Adams, Martha Larson, Jaap Verheul and Michael Boyden	87
<i>ChiWUG: A Graph-based Evaluation Dataset for Chinese Lexical Semantic Change Detection</i>	
Jing Chen, Emmanuele Chersoni, Dominik Schlechtweg, Jelena Prokic and Chu-Ren Huang ..	93
<i>Towards Detecting Lexical Change of Hate Speech in Historical Data</i>	
Sanne Hoeken, Sophie Jasmin Spliethoff, Silke Schwandt, Sina Zarriëß and Özge Alacam ...	100
<i>Changing usage of Low Saxon auxiliary and modal verbs</i>	
Janine Siewert, Martijn Wieling and Yves Scherrer	112
<i>Semantic Shifts in Mental Health-Related Concepts</i>	
Naomi Baes, Nick Haslam and Ekaterina Vylomova	119
<i>Automating Sound Change Prediction for Phylogenetic Inference: A Tukanoan Case Study</i>	
Kalvin Chang, Nathaniel Romney Robinson, Anna Cai, Ting Chen, Annie Zhang and David R Mortensen	129
<i>Scent and Sensibility: Perception Shifts in the Olfactory Domain</i>	
Teresa Paccosi, Stefano Menini, Elisa Leonardelli, Iliaria Barzon and Sara Tonelli	143
<i>From Diachronic to Contextual Lexical Semantic Change: Introducing Semantic Difference Keywords (SDKs) for Discourse Studies</i>	
Isabelle Gribomont	153

Program

Wednesday, December 6, 2023

09:15 - 09:30 *Introduction*

09:30 - 10:30 *Keynote Mario Giulianelli*

10:30 - 11:00 *Coffee Break*

11:00 - 12:00 *Session 1*

EvoSem: A database of polysemous cognate sets

Mathieu Dehouck, Alex François, Siva Kalyan, Martial Pastor and David Kletz

Semantic Shifts in Mental Health-Related Concepts

Naomi Baes, Nick Haslam and Ekaterina Vylomova

Scent and Sensibility: Perception Shifts in the Olfactory Domain

Teresa Paccosi, Stefano Menini, Elisa Leonardelli, Ilaria Barzon and Sara Tonelli

12:00 - 13:30 *Lunch Break*

13:30 - 14:30 *Keynote Gemma Boleda*

14:30 - 15:30 *Session 2*

Political dogwhistles and community divergence in semantic change

Max Boholm and Asad B. Sayeed

Automating Sound Change Prediction for Phylogenetic Inference: A Tukanooan Case Study

Kalvin Chang, Nathaniel Romney Robinson, Anna Cai, Ting Chen, Annie Zhang and David R Mortensen

Domain-Adapting BERT for Attributing Manuscript, Century and Region in Pre-Modern Slavic Texts

Piroska Lendvai, Uwe Reichel, Anna Jouravel, Achim Rabus and Elena Renje

Wednesday, December 6, 2023 (continued)

15:30 - 16:30 *Poster Session*

ChiWUG: A Graph-based Evaluation Dataset for Chinese Lexical Semantic Change Detection

Jing Chen, Emmanuele Chersoni, Dominik Schlechtweg, Jelena Prokic and Chu-Ren Huang

Changing usage of Low Saxon auxiliary and modal verbs

Janine Siewert, Martijn Wieling and Yves Scherrer

Towards Detecting Lexical Change of Hate Speech in Historical Data

Sanne Hoeken, Sophie Jasmin Spliethoff, Silke Schwandt, Sina Zarriß and Özge Alacam

From Diachronic to Contextual Lexical Semantic Change: Introducing Semantic Difference Keywords (SDKs) for Discourse Studies

Isabelle Gribomont

Multi-lect automatic detection of Swadesh list items from raw corpus data in East Slavic languages

Ilia Afanasev

Literary Intertextual Semantic Change Detection: Application and Motivation for Evaluating Models on Small Corpora

Jackson Ehrenworth and Katherine A. Keith

A longitudinal study about gradual changes in the Iranian Online Public Sphere pre and post of 'Mahsa Moment': Focusing on Twitter

Sadegh Jafari, Amin Fathi, Abolfazl Hajizadegan, Amirmohammad Kazemeini and Sauleh Eetemadi

16:30 - 17:30 *Session 3*

Representing and Computing Uncertainty in Phonological Reconstruction

Johann-Mattis List, Nathan Hill, Robert Forkel and Frederic Blum

Anchors in Embedding Space: A Simple Concept Tracking Approach to Support Conceptual History Research

Jetske Adams, Martha Larson, Jaap Verheul and Michael Boyden

GHisBERT – Training BERT from scratch for lexical semantic investigations across historical German language stages

Christin Beck and Marisa Köllner

Wednesday, December 6, 2023 (continued)

17:30 - 17:45 *Closing Remarks*