

LChange 2022

**3rd International Workshop on Computational Approaches
to Historical Language Change 2022**

Proceedings of the Workshop

May 26-27, 2022

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

Gold



©2022 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 978-1-955917-42-1

Preface by the General Chair

Welcome to the 3rd International Workshop on Computational Approaches to Historical Language Change (LChange'22) co-located with ACL 2022. This year, LChange is held over two days, May 26–27 2022, as a hybrid event with participation possible both virtually and on-site in Dublin, Ireland. To support efforts in evaluation of computational methodologies for uncovering language change, LChange'22 features a shared task on semantic change detection for Spanish as one track of the workshop.

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 focus on digital text corpora**. In doing so, this workshop carries out the triple goals of disseminating the state-of-the-art research on diachronic modelling 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 21 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 15 full and short papers as oral or poster presentations. We had two distinguished keynote presentations: the first by Dirk Geeraerts (KU Leuven/University of Gothenburg) who presented a talk entitled “Can historical semantics save lives? (And other questions for computational diachronic semantics)”, and the second by Dominik Schlechtweg (University of Stuttgart) with the talk “Human and Computational Measurement of Lexical Semantic Change”. Finally, we invited two additional papers to be presented as posters, one published at the ACL 2022 conference and one in *Findings of NAACL*, which are not included in the workshop proceedings.

The shared task on semantic change discovery and detection in Spanish was divided in two phases: (1) graded change discovery; and (2) binary change detection. The main novelty with respect to the previous tasks consisted in predicting and evaluating changes for all vocabulary words in the corpus. Six teams participated in phase 1 and seven teams in phase 2.

To further support the community, we offered two student scholarships for the main conference in addition to the workshop, as well as mentoring for young researchers. Five researchers were offered mentoring on a topic of their choice, 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 ACL 2022 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, INRIA Paris (France)

Andrey Kutuzov, University of Oslo (Norway)

Simon Hengchen, University of Gothenburg (Sweden)

Haim Dubossarsky, University of Cambridge (United Kingdom)

Lars Borin, University of Gothenburg (Sweden)

LChange'22 Workshop Chairs

Organizing Committee

General Chair

Nina Tahmasebi, University of Gothenburg, Sweden

Program Chairs

Lars Borin, University of Gothenburg, Sweden

Simon Hengchen, University of Gothenburg, Sweden

Syrielle Montariol, INRIA Paris, France

Haim Dubossarsky, University of Cambridge, United Kingdom

Andrey Kutuzov, University of Oslo, Norway

Program Committee

Program Committee

Aleksandrs Berdicevskis, University of Gothenburg
Animesh Mukherjee, Indian Institute of Technology Kharagpur
Annalina Caputo, Dublin City University
Barbara McGillivray, King's College London, University of London
Barend Beekhuizen, University of Toronto
Clémentine Fourier, INRIA Paris
Ekaterina Vylomova, University of Melbourne
Ella Rabinovich, University of Toronto
Enrique Manjavacas, Leiden University
Eyal Sagi, University of St. Francis
Filip Miletic, University of Toulouse
Ian Stewart, University of Michigan
Karlien Franco, KU Leuven
Lidia Pivovarova, University of Helsinki
Ludovic Tanguy, University of Toulouse
Maike Park, Leibniz-Institute for the German Language
Mario Giulianelli, University of Amsterdam
Martin Pömsl, McGill University
Matej Martinc, Jozef Stefan Institute
Mauricio Gruppi, Rensselaer Polytechnic Institute
Michael Färber, Karlsruhe Institute of Technology
Paul Cook, University of New Brunswick
Paul Nulty, University College Dublin
Pierluigi Cassotti, University of Bari
Pierpaolo Basile, University of Bari
Péter Jeszenszky, Universität Bern
Samia Touileb, University of Bergen
Stefano De Pascale, Vrije Universiteit Brussel
Taraka Rama, University of North Texas
Vaibhav Jain, Delhi Technological University, Dhirubhai Ambani Institute Of Information and
Communication Technology
Valentin Hofmann, University of Oxford
Yijun Duan, AIST
Ying Li, Max Planck Institut

Keynote Talk: Can historical semantics save lives? (And other questions for computational diachronic semantics)

Dirk Geeraerts

KU Leuven/University of Gothenburg

Abstract: Drawing on a number (methodologically non-computational) diachronic semantic studies that I have carried out at various points over the past decades, I would like to draw the attention to three issues that have so far played only a secondary role in the booming field of computational diachronic semantics but that might provide some inspiration for a further expansion: first, the double-sided status of textual interpretation, which can feature both as a descriptive target and as a methodological source in historical semantics; second, the relevance of incorporating an onomasiological dimension in the definition of semantic change; and third, the distinction between generalizations about semantic change that are formulated in terms of structural and functional features (like isomorphism or frequency) and generalizations that correlate semantic changes with external phenomena (like societal changes).

Bio: Dirk Geeraerts is professor emeritus of linguistics at the University of Leuven. His main research focus involves the fields of lexical semantics and lexicology, with specific attention for social variation and diachronic change of meaning and vocabulary. He is the founder of the journal *Cognitive Linguistics*, and editor of *The Oxford Handbook of Cognitive Linguistics* (2007). Publications include *The Structure of Lexical Variation* (Mouton De Gruyter 1994), *Diachronic Prototype Semantics* (OUP 1997), *Words and Other Wonders* (Mouton De Gruyter 2006), *Theories of Lexical Semantics* (OUP 2010), and *Ten Lectures on Cognitive Sociolinguistics* (Brill 2018).

Keynote Talk: Human and Computational Measurement of Lexical Semantic Change

Dominik Schlechtweg

University of Stuttgart/University of Texas, Austin

Abstract: Human language changes over time. This change occurs on several linguistic levels such as grammar, sound or meaning. The study of meaning changes on the word level is often called Lexical Semantic Change (LSC) and is traditionally either approached from an onomasiological perspective asking by which words a meaning can be expressed, or a semasiological perspective asking which meanings a word can express over time. In recent years, the task of automatic detection of semasiological LSC from textual data has been established as a proper field of computational linguistics under the name of Lexical Semantic Change Detection (LSCD). Two main factors have contributed to this development: (i) the *digital turn* in the humanities has made large amounts of historical texts available in digital form. (ii) New *computational models* have been introduced efficiently learning semantic aspects of words solely from text. One of the main motivations behind the work on LSCD are their applications in historical semantics and historical lexicography where researchers are concerned with the classification of words into categories of semantic change. Automatic methods have the advantage to produce semantic change predictions for large amounts of data in small amounts of time and could thus considerably decrease human efforts in the mentioned fields, while being able to scan more data and thus to uncover more semantic changes which are at the same time less biased towards ad hoc sampling criteria used by researchers. On the other hand, automatic methods may also be hurtful when their predictions are biased, i.e., they may miss numerous semantic changes or label words as changing which are not. Results produced in this way may then lead researchers to make empirically inadequate generalizations on semantic change. Hence, automatic change detection methods should not be trusted until they have been evaluated thoroughly and their predictions have been shown to reach an acceptable level of correctness.

Despite the rapid growth of LSCD as a field a solid evaluation of the wealth of proposed models was still missing in 2017. The reasons were multiple, but most importantly there was no annotated benchmark test set available. In this talk I will describe the work done for my PhD from the last five years aimed at standardizing the evaluation of LSCD models.

Bio: Dominik Schlechtweg did his PhD at the IMS (University of Stuttgart) working together with Sabine Schulte im Walde on automatic detection of lexical semantic change. He held a PhD scholarship from Konrad Adenauer Foundation. Since February 2022 he is a post-doctoral researcher at the IMS (University of Stuttgart), working in the 6-year research program *Change is Key!* and in the research project *Towards computational lexical semantic change detection*. Currently, he is doing a research internship with Katrin Erk at the University of Texas, Austin.

Table of Contents

| | |
|--|-----|
| <i>A Multilingual Benchmark to Capture Olfactory Situations over Time</i> Stefano Menini, Teresa Paccosi, Sara Tonelli, Marieke Van Erp, Inger Leemans, Pasquale Lisenà, Raphael Troncy, William Tullett, Ali Hürriyetoğlu, Ger Dijkstra, Femke Gordijn, Elias Jürgens, Josephine Koopman, Aron Ouwerkerk, Sanne Steen, Inna Novalija, Janez Brank, Dunja Mladenic and Anja Zidar | 1 |
| <i>Language Acquisition, Neutral Change, and Diachronic Trends in Noun Classifiers</i> Aniket Kali and Jordan Kodner | 11 |
| <i>Deconstructing destruction: A Cognitive Linguistics perspective on a computational analysis of diachronic change</i> Karlien Franco, Mariana Montes and Kris Heylen | 23 |
| <i>What is Done is Done: an Incremental Approach to Semantic Shift Detection</i> Francesco Periti, Alfio Ferrara, Stefano Montanelli and Martin Ruskov | 33 |
| <i>From qualifiers to quantifiers: semantic shift at the paradigm level</i> Quentin Feltgen | 44 |
| <i>Do Not Fire the Linguist: Grammatical Profiles Help Language Models Detect Semantic Change</i> Mario Giulianelli, Andrey Kutuzov and Lidia Pivovarova | 54 |
| <i>Explainable Publication Year Prediction of Eighteenth Century Texts with the BERT Model</i> Iiro Rastas, Yann Ciarán Ryan, Iiro Tiihonen, Mohammadreza Qaraei, Liina Repo, Rohit Babbar, Eetu Mäkelä, Mikko Tolonen and Filip Ginter | 68 |
| <i>Using Cross-Lingual Part of Speech Tagging for Partially Reconstructing the Classic Language Family Tree Model</i> Anat Samohi, Daniel Weisberg Mitelman and Kfir Bar | 78 |
| <i>A New Framework for Fast Automated Phonological Reconstruction Using Trimmed Alignments and Sound Correspondence Patterns</i> Johann-Mattis List, Robert Forkel and Nathan Hill | 89 |
| <i>Caveats of Measuring Semantic Change of Cognates and Borrowings using Multilingual Word Embeddings</i> Clémentine Fourier and Syrielle Montariol | 97 |
| <i>Lexicon of Changes: Towards the Evaluation of Diachronic Semantic Shift in Chinese</i> Jing Chen, Emmanuele Chersoni and Chu-ren Huang | 113 |
| <i>Low Saxon dialect distances at the orthographic and syntactic level</i> Janine Siewert, Yves Scherrer and Martijn Wieling | 119 |
| <i>Vaderland, Volk and Natie: Semantic Change Related to Nationalism in Dutch Literature Between 1700 and 1880 Captured with Dynamic Bernoulli Word Embeddings</i> Marije Timmermans, Eva Vanmassenhove and Dimitar Shterionov | 125 |
| <i>Using neural topic models to track context shifts of words: a case study of COVID-related terms before and after the lockdown in April 2020</i> Olga Kellert and Md Mahmud Uz Zaman | 131 |
| <i>Roadblocks in Gender Bias Measurement for Diachronic Corpora</i> Saied Alshahrani, Esma Wali, Abdullah R Alshamsan, Yan Chen and Jeanna Matthews | 140 |

| | |
|---|-----|
| [LSCDISCOVERY SHARED TASK] <i>LSCDiscovery: A shared task on semantic change discovery and detection in Spanish</i> | |
| Frank D. Zamora-Reina, Felipe Bravo-Marquez and Dominik Schlechtweg | 149 |
| [LSCDISCOVERY SHARED TASK] <i>BOS at LSCDiscovery: Lexical Substitution for Interpretable Lexical Semantic Change Detection</i> | |
| Artem Kudisov and Nikolay Arefyev | 165 |
| [LSCDISCOVERY SHARED TASK] <i>DeepMistake at LSCDiscovery: Can a Multilingual Word-in-Context Model Replace Human Annotators?</i> | |
| Daniil Homskiy and Nikolay Arefyev | 173 |
| [LSCDISCOVERY SHARED TASK] <i>UAlberta at LSCDiscovery: Lexical Semantic Change Detection via Word Sense Disambiguation</i> | |
| Daniela Teodorescu, Spencer von der Ohe and Grzegorz Kondrak | 180 |
| [LSCDISCOVERY SHARED TASK] <i>CoToHiLi at LSCDiscovery: the Role of Linguistic Features in Predicting Semantic Change</i> | |
| Ana Sabina Uban, Alina Maria Cristea, Anca Daniela Dinu, Liviu P Dinu, Simona Georgescu and Laurentiu Zoicas | 187 |
| [LSCDISCOVERY SHARED TASK] <i>HSE at LSCDiscovery in Spanish: Clustering and Profiling for Lexical Semantic Change Discovery</i> | |
| Kseniia Kashleva, Alexander Shein, Elizaveta Tukhtina and Svetlana Vydrina | 193 |
| [LSCDISCOVERY SHARED TASK] <i>GlossReader at LSCDiscovery: Train to Select a Proper Gloss in English – Discover Lexical Semantic Change in Spanish</i> | |
| Maxim Rachinskiy and Nikolay Arefyev | 198 |

Program

Thursday, May 26, 2022

09:00 - 09:15 *Introduction*

09:15 - 10:35 *Session 1*

Low Saxon dialect distances at the orthographic and syntactic level

Janine Siewert, Yves Scherrer and Martijn Wieling

A New Framework for Fast Automated Phonological Reconstruction Using Trimmed Alignments and Sound Correspondence Patterns

Johann-Mattis List, Robert Forkel and Nathan Hill

What is Done is Done: an Incremental Approach to Semantic Shift Detection

Francesco Periti, Alfio Ferrara, Stefano Montanelli and Martin Ruskov

10:35 - 11:05 *BREAK*

11:05 - 12:30 *Session 2*

Lexicon of Changes: Towards the Evaluation of Diachronic Semantic Shift in Chinese

Jing Chen, Emmanuele Chersoni and Chu-ren Huang

Deconstructing destruction: A Cognitive Linguistics perspective on a computational analysis of diachronic change

Karlien Franco, Mariana Montes and Kris Heylen

Using Cross-Lingual Part of Speech Tagging for Partially Reconstructing the Classic Language Family Tree Model

Anat Samohi, Daniel Weisberg Mitelman and Kfir Bar

12:30 - 14:00 *LUNCH / BREAK*

14:00 - 15:00 *Keynote 1 - Dirk Geeraerts*

15:00 - 15:30 *Session 3*

Do Not Fire the Linguist: Grammatical Profiles Help Language Models Detect Semantic Change

Mario Giulianelli, Andrey Kutuzov and Lidia Pivovarova

Thursday, May 26, 2022 (continued)

15:30 - 17:00 *Poster session + COFFEE*

From qualifiers to quantifiers: semantic shift at the paradigm level

Quentin Feltgen

Explainable Publication Year Prediction of Eighteenth Century Texts with the BERT Model

Iiro Rastas, Yann Ciarán Ryan, Iiro Tiihonen, Mohammadreza Qaraei, Liina Repo, Rohit Babbar, Eetu Mäkelä, Mikko Tolonen and Filip Ginter

Caveats of Measuring Semantic Change of Cognates and Borrowings using Multilingual Word Embeddings

Clémentine Fourrier and Syrielle Montariol

Vaderland, Volk and Natie: Semantic Change Related to Nationalism in Dutch Literature Between 1700 and 1880 Captured with Dynamic Bernoulli Word Embeddings

Marije Timmermans, Eva Vanmassenhove and Dimitar Shterionov

Language Acquisition, Neutral Change, and Diachronic Trends in Noun Classifiers

Aniket Kali and Jordan Kodner

A Multilingual Benchmark to Capture Olfactory Situations over Time

Stefano Menini, Teresa Paccosi, Sara Tonelli, Marieke Van Erp, Inger Leemans, Pasquale Lisena, Raphael Troncy, William Tullett, Ali Hürriyetoglu, Ger Dijkstra, Femke Gordijn, Elias Jürgens, Josephine Koopman, Aron Ouwerkerk, Sanne Steen, Inna Novalija, Janez Brank, Dunja Mladenec and Anja Zidar

Using neural topic models to track context shifts of words: a case study of COVID-related terms before and after the lockdown in April 2020

Olga Kellert and Md Mahmud Uz Zaman

Roadblocks in Gender Bias Measurement for Diachronic Corpora

Saied Alshahrani, Esma Wali, Abdullah R Alshamsan, Yan Chen and Jeanna Matthews

[LSCDISCOVERY SHARED TASK] *UAlberta at LSCDiscovery: Lexical Semantic Change Detection via Word Sense Disambiguation*

Daniela Teodorescu, Spencer von der Ohe and Grzegorz Kondrak

Friday, May 27, 2022

09:30 - 09:45 *Introduction day 2*

09:40 - 10:40 *Keynote 2 - Dominik Schlechtweg*

10:40 - 11:05 *BREAK*

11:05 - 11:25 *Task description paper*

[LSCDISCOVERY SHARED TASK] *LSCDiscovery: A shared task on semantic change discovery and detection in Spanish*

Frank D. Zamora-Reina, Felipe Bravo-Marquez and Dominik Schlechtweg

11:25 - 12:45 *Best task paper 1*

[LSCDISCOVERY SHARED TASK] *DeepMistake at LSCDiscovery: Can a Multilingual Word-in-Context Model Replace Human Annotators?*

Daniil Homskiy and Nikolay Arefyev

11:45 - 12:05 *Best task paper 2*

[LSCDISCOVERY SHARED TASK] *GlossReader at LSCDiscovery: Train to Select a Proper Gloss in English – Discover Lexical Semantic Change in Spanish*

Maxim Rachinskiy and Nikolay Arefyev

12:05 - 13:30 *LUNCH / BREAK*

13:30 - 15:00 *Virtual poster session + COFFEE*

[LSCDISCOVERY SHARED TASK] *BOS at LSCDiscovery: Lexical Substitution for Interpretable Lexical Semantic Change Detection*

Artem Kudisov and Nikolay Arefyev

[LSCDISCOVERY SHARED TASK] *UAlberta at LSCDiscovery: Lexical Semantic Change Detection via Word Sense Disambiguation*

Daniela Teodorescu, Spencer von der Ohe and Grzegorz Kondrak

[LSCDISCOVERY SHARED TASK] *CoToHiLi at LSCDiscovery: the Role of Linguistic Features in Predicting Semantic Change*

Ana Sabina Uban, Alina Maria Cristea, Anca Daniela Dinu, Liviu P Dinu, Simona Georgescu and Laurentiu Zoicas

[LSCDISCOVERY SHARED TASK] *HSE at LSCDiscovery in Spanish: Clustering and Profiling for Lexical Semantic Change Discovery*

Kseniia Kashleva, Alexander Shein, Elizaveta Tukhtina and Svetlana Vydrina

Friday, May 27, 2022 (continued)

15:00 - 16:00 *Mentoring*

16:00 - 16:30 *Closing*