

ACL 2019

**The 1st International Workshop on Computational  
Approaches to Historical Language Change**

**Proceedings of the Workshop**

August 2, 2019  
Florence, Italy

©2019 The 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](mailto:acl@aclweb.org)

ISBN 978-1-950737-31-4

## Introduction

Welcome to the 1st International Workshop on Computational Approaches to Historical Language Change (LChange'19) that was co-located with ACL 2019 in Florence, on August 2, 2019.

Human language changes over time, driven by the dual needs of adapting to ongoing sociocultural and technological development in the world and facilitating efficient communication. In particular, novel words are coined or borrowed from other languages, while obsolete words slide into obscurity. Similarly, words may acquire novel meanings or lose existing meanings. This workshop explores these phenomena by bringing to bear state-of-the-art computational methodologies, theories and digital text resources on exploring the time-varying nature of human language.

Although there exists rich empirical work on language change from historical linguistics, sociolinguistics and cognitive linguistics, computational approaches to the problem of language change – particularly how word forms and meanings evolve – have only begun to take shape over the past decade or so, with exemplary work on semantic change and lexical replacement. The motivation has long been related to search, and understanding in diachronic archives. The emergence of long-term and large-scale digital corpora was the prerequisite and has resulted in a slightly different set of problems for this strand of study than have traditionally been studied in historical linguistics. As an example, studies of lexical replacement have largely focused on named entity change (names of e.g., countries and people that change over time) because of the large effect these name changes have for temporal information retrieval.

The aim of this workshop is three-fold. First, we want to provide pioneering researchers who work on computational methods, evaluation, and large-scale modelling of language change an outlet for disseminating cutting-edge research on topics concerning language change. Currently, researchers in this area have published in a wide range of different venues, from computational linguistics, to cognitive science and digital archiving venues. We intended this workshop as a platform for sharing state-of-the-art research progress in this fundamental domain of natural language research.

Second, in doing so we want to bring together domain experts across disciplines. We want to connect those that have long worked on language change within historical linguistics and bring with them a large understanding for general linguistic theories of language change; those that have studied change across languages and language families; those that develop and test computational methods for detecting semantic change and laws of semantic change; and those that need knowledge (of the occurrence and shape) of language change, for example, in digital humanities and computational social sciences where text mining is applied to diachronic corpora subject to lexical semantic change.

Third, the detection and modelling of language change using diachronic text and text mining raise fundamental theoretical and methodological challenges for future research in this area. The representativeness of text is a first critical issue; works using large diachronic corpora and computational methods for detecting change often claim to find changes that are universally true for a language as a whole. But the jury is out on how results derived from digital literature or newspapers accurately represent changes in language as a whole. We hope to engage corpus linguists, big-data scientists, and computational linguists to address these open issues. Besides these goals, this workshop can also support discussion on the evaluation of computational methodologies for uncovering language change. Verifying change only using positive examples of change often confirms a corpus bias rather than reflecting genuine language change. Larger quantities and higher qualities of text over time result in the detection of more semantic change. In fact, multiple semantic laws have been linked to frequency rather than underlying semantic change. The methodological issue of evaluation, together with good evaluation testsets and standards are of high importance to the research community. We aim to shed some light on these issues and encourage the community to collaborate to find solutions.

The work in semantic change detection has, to a large extent, moved to (neural) embedding techniques in recent years. These methods have several drawbacks: the need for very large datasets to produce stable embeddings, and the fact that all semantic information of a word is encoded in a single vector thus limiting the possibility to study word senses separately. A move towards multi-sense embeddings will most likely require even more texts per time unit, which will limit the applicability of these methods to other languages than English and a few others. We want to bring about a discussion on the need for methods that can discriminate and disambiguate among a word's senses (meanings) and that can be used for resource-poor languages with little hope of acquiring the order of magnitude of words needed for creating stable embeddings, possibly using dynamic embeddings that seem to require less text. Finally, knowledge of language change is useful not only on its own, but as a basis for other diachronic textual investigations and in search.

A digital humanities investigation into the living conditions of young women through history cannot rely on the word *girl* in English, as in the past the reference of *girl* also included young men. Automatic detecting of language change is useful for many researchers outside of the communities that study the changes themselves and develop methods for their detection. By reaching out to these other communities, we can better understand how to utilize the results for further research and for presenting them to the interested public. In addition, we need good user interfaces and systems for exploring language changes in corpora, for example, to allow for serendipitous discovery of interesting phenomena. In addition to facilitate research on texts, information about language changes is used for measuring document across-time similarity, information retrieval from long-term document archives, the design of OCR algorithms and so on.

In response to the call we received 53 submissions, each of which were carefully evaluated by at least two members of the Program Committee. Based on the reviewer's feedback we accepted 34 full and short papers, which were then presented orally or as poster papers. We were also delighted to have two keynote presentations by Claire Bowers (Yale University) and Haim Dubossarsky (University of Cambridge). We hope that you will find the included papers as insightful and inspiring as we have.

We would like to thank the keynote speakers for their stimulating talks, the authors of papers for their interesting contributions and the members of the Program Committee for their insightful reviews. We also express our gratitude to the ACL 2019 workshop chairs for their kind assistance.

Nina Tahmasebi (University of Gothenburg)  
Lars Borin (University of Gothenburg)  
Adam Jatowt (Kyoto University)  
Yang Xu (University of Toronto)  
LChange'19 Workshop Chairs

**Organizers:**

Nina Tahmasebi, University of Gothenburg (Sweden)  
Lars Borin, University of Gothenburg (Sweden)  
Adam Jatowt, Kyoto University (Japan)  
Yang Xu, University of Toronto (Canada)

**Program Committee:**

Yvonne Adesam, University of Gothenburg (Sweden)  
Rami Aly, Universität Hamburg (Germany)  
Avishek Anand, L3S Research Center (Germany)  
Timothy Baldwin, University of Melbourne (Australia)  
Pierpaolo Basile, University of Bari (Italy)  
Barend Beekhuizen, University of Toronto Mississauga (Canada)  
Meriem Beloucif, Universität Hamburg (Germany)  
Klaus Berberich, MPI-INF (Germany)  
Aleksandrs Berdicevskis, University of Gothenburg (Sweden)  
Chris Biemann, Universität Hamburg (Germany)  
Damian Blasi, University of Zürich (Switzerland)  
Ricardo Campos, Polytechnic Institute of Tomar / INESC TEC, (Portugal)  
Annalina Caputo, Trinity College Dublin (Ireland)  
Brady Clark, Northwestern University (USA)  
Paul Cook, University of New Brunswick (Canada)  
Dana Dannells, University of Gothenburg (Sweden)  
Pavel Denisov, University of Stuttgart (Germany)  
Yijun Duan, Kyoto University (Japan)  
Haim Dubossarsky, University of Cambridge (UK)  
Stian Rødven Eide, University of Gothenburg (Sweden)  
Michael Färber, KIT (Germany)  
Antske Fokkens, Free University of Amsterdam (The Netherlands)  
Mats Fridlund, University of Gothenburg (Sweden)  
Mika Hämmäläinen, University of Helsinki (Finland)  
Johannes Hellrich, Friedrich Schiller University Jena (Germany)  
Simon Hengchen, University of Helsinki (Finland)  
Louise Holmer, University of Gothenburg (Sweden)  
Abhik Jana, IIT Kharagpur (India)  
Péter Jeszenszky, Ritsumeikan University (Japan)  
Dirk Johannßen, Universität Hamburg (Germany)  
Richard Johansson, University of Gothenburg (Sweden)  
Antti Kanner, University of Helsinki (Finland)  
Tom Kenter, Google (UK)  
Jey Han Lau, University of Melbourne (Australia)  
Nicholas A. Lester, University of Zürich (Switzerland)  
Liina Lindström, University of Tartu (Estonia)  
Behrooz Mansouri, Rochester Institute of Technology (USA)  
Animesh Mukherjee, IIT Kharagpur (India)  
Luis Nieto Piña, University of Gothenburg (Sweden)

Bill Noble, University of Gothenburg (Sweden)  
Kjetil Norvag, NTNU (Norway)  
Ella Rabinovich, University of Toronto (Canada)  
Taraka Rama, University of Oslo (Norway)  
Jacobo Rouces, University of Gothenburg (Sweden)  
Sylvie Saget, University of Gothenburg (Sweden)  
Eyal Sagi, Northwestern University (USA)  
Asad Sayeed, University of Gothenburg (Sweden)  
Dominik Schlechtweg, University of Stuttgart (Germany)  
Vidya Somashekarappa, University of Gothenburg (Sweden)  
Andreas Spitz, EPFL (Switzerland)  
Ian Stewart, Georgia Institute of Technology (USA)  
Suzanne Stevenson, University of Toronto (Canada)  
Susanne Vejdemo, Stockholm University (Sweden)  
Mikael Vejdemo Johansson, CUNY CSI (USA)  
Barbro Wallgren Hemlin, University of Gothenburg (Sweden)  
Melvin Wevers, KNAW Humanities Cluster (The Netherlands)  
Guanghao You, University of Zürich (Switzerland)  
Yihong Zhang, Osaka University (Japan)

**Invited Speakers:**

Claire Bower, Yale University (USA)  
Haim Dubossarsky, University of Cambridge (UK)

## Table of Contents

<i>From Insanely Jealous to Insanely Delicious: Computational Models for the Semantic Bleaching of English Intensifiers</i>	
Yiwei Luo, Dan Jurafsky and Beth Levin .....	1
<i>Computational Analysis of the Historical Changes in Poetry and Prose</i>	
Amitha Gopidi and Aniket Alam .....	14
<i>Studying Semantic Chain Shifts with Word2Vec: FOOD&gt;MEAT&gt;FLESH</i>	
Richard Zimmermann .....	23
<i>Evaluation of Semantic Change of Harm-Related Concepts in Psychology</i>	
Ekaterina Vylomova, Sean Murphy and Nicholas Haslam .....	29
<i>Contextualized Diachronic Word Representations</i>	
Ganesh Jawahar and Djamé Seddah .....	35
<i>Semantic Change and Semantic Stability: Variation is Key</i>	
Claire Bower .....	48
<i>GASC: Genre-Aware Semantic Change for Ancient Greek</i>	
Valerio Perrone, Marco Palma, Simon Hengchen, Alessandro Vatri, Jim Q. Smith and Barbara McGillivray .....	56
<i>Modeling Markedness with a Split-and-Merger Model of Sound Change</i>	
Andrea Ceolin and Ollie Sayeed .....	67
<i>A Method to Automatically Identify Diachronic Variation in Collocations.</i>	
Marcos Garcia and Marcos García Salido .....	71
<i>Written on Leaves or in Stones?: Computational Evidence for the Era of Authorship of Old Thai Prose</i>	
Attapol Rutherford and Santhawat Thanyawong .....	81
<i>Identifying Temporal Trends Based on Perplexity and Clustering: Are We Looking at Language Change?</i>	
Sidsel Boldsen, Manex Agirrezabal and Patrizia Paggio .....	86
<i>Using Word Embeddings to Examine Gender Bias in Dutch Newspapers, 1950-1990</i>	
Melvin Wevers .....	92
<i>Predicting Historical Phonetic Features using Deep Neural Networks: A Case Study of the Phonetic System of Proto-Indo-European</i>	
Frederik Hartmann .....	98
<i>ParHistVis: Visualization of Parallel Multilingual Historical Data</i>	
Aikaterini-Lida Kalouli, Rebecca Kehlbeck, Rita Sevastjanova, Katharina Kaiser, Georg A. Kaiser and Miriam Butt .....	109
<i>Tracing Antisemitic Language Through Diachronic Embedding Projections: France 1789-1914</i>	
Rocco Tripodi, Massimo Warglien, Simon Levis Sullam and Deborah Paci .....	115
<i>DiaHClust: an Iterative Hierarchical Clustering Approach for Identifying Stages in Language Change</i>	
Christin Schätzle and Hannah Booth .....	126

<i>Treat the Word As a Whole or Look Inside? Subword Embeddings Model Language Change and Typology</i> Yang Xu, Jiasheng Zhang and David Reitter .....	136
<i>Times Are Changing: Investigating the Pace of Language Change in Diachronic Word Embeddings</i> Stephanie Brandl and David Lassner .....	146
<i>The Rationality of Semantic Change</i> Omer Korat .....	151
<i>Studying Laws of Semantic Divergence across Languages using Cognate Sets</i> Ana Uban, Alina Maria Ciobanu and Liviu P. Dinu .....	161
<i>Detecting Syntactic Change Using a Neural Part-of-Speech Tagger</i> William Merrill, Gigi Stark and Robert Frank .....	167
<i>Grammar and Meaning: Analysing the Topology of Diachronic Word Embeddings</i> Yuri Bizzoni, Stefania Degaetano-Ortlieb, Katrin Menzel, Pauline Krielke and Elke Teich .....	175
<i>Spatio-Temporal Prediction of Dialectal Variant Usage</i> Péter Jeszenszky, Panote Siriaraya, Philipp Stoeckle and Adam Jatowt .....	186
<i>One-to-X Analogical Reasoning on Word Embeddings: a Case for Diachronic Armed Conflict Prediction from News Texts</i> Andrey Kutuzov, Erik Velldal and Lilja Øvrelid .....	196
<i>Measuring Diachronic Evolution of Evaluative Adjectives with Word Embeddings: the Case for English, Norwegian, and Russian</i> Julia Rodina, Daria Bakshandaeva, Vadim Fomin, Andrey Kutuzov, Samia Touileb and Erik Velldal 202	
<i>Semantic Change in the Language of UK Parliamentary Debates</i> Gavin Abercrombie and Riza Batista-Navarro .....	210
<i>Semantic Change and Emerging Tropes In a Large Corpus of New High German Poetry</i> Thomas Haider and Steffen Eger .....	216
<i>Conceptual Change and Distributional Semantic Models: an Exploratory Study on Pitfalls and Possibilities</i> Pia Sommerauer and Antske Fokkens .....	223
<i>Measuring the Compositionality of Noun-Noun Compounds over Time</i> Prajit Dhar, Janis Pagel and Lonneke van der Plas .....	234
<i>Towards Automatic Variant Analysis of Ancient Devotional Texts</i> Amir Hazem, Béatrice Daille, Dominique Stutzmann, Jacob Currie and Christine Jacquin .....	240
<i>Understanding the Evolution of Circular Economy through Language Change</i> Sampriti Mahanty, Frank Boons, Julia Handl and Riza Theresa Batista-Navarro .....	250
<i>Gaussian Process Models of Sound Change in Indo-Aryan Dialectology</i> Chundra Cathcart .....	254
<i>Modeling a Historical Variety of a Low-Resource Language: Language Contact Effects in the Verbal Cluster of Early-Modern Frisian</i> Jelke Bloem, Arjen Versloot and Fred Weerman .....	265



*Visualizing Linguistic Change as Dimension Interactions*

Christin Schätzle, Frederik L. Dennig, Michael Blumenschein, Daniel A. Keim and Miriam Butt

272



# Conference Program

August 2, 2019

**9:00–9:15**     *Introduction*

**9:15–10:30**     **Session 1**

9:15–10:00     *Semantic Change in the Time of Machine Learning, Doing It Right!*  
Haim Dubossarsky

10:00–10:30     *From Insanely Jealous to Insanely Delicious: Computational Models for the Semantic Bleaching of English Intensifiers*  
Yiwei Luo, Dan Jurafsky and Beth Levin

**10:30–10:45**     *Coffee Break*

**10:45–12:30**     **Session 2**

10:45–11:15     *Computational Analysis of the Historical Changes in Poetry and Prose*  
Amitha Gopidi and Aniket Alam

11:15–11:35     *Studying Semantic Chain Shifts with Word2Vec: FOOD>MEAT>FLESH*  
Richard Zimmermann

11:35–11:55     *Evaluation of Semantic Change of Harm-Related Concepts in Psychology*  
Ekaterina Vylomova, Sean Murphy and Nicholas Haslam

11:55–12:25     *Contextualized Diachronic Word Representations*  
Ganesh Jawahar and Djamé Seddah

**12:30–13:30**     *Lunch Break*

**August 2, 2019 (continued)**

**13:30–14:30 Session 3**

13:30–14:30 *Semantic Change and Semantic Stability: Variation is Key*

Claire Bowern

**14:30–16:00 Session 4 (Poster Session)**

*GASC: Genre-Aware Semantic Change for Ancient Greek*

Valerio Perrone, Marco Palma, Simon Hengchen, Alessandro Vatri, Jim Q. Smith and Barbara McGillivray

*Modeling Markedness with a Split-and-Merger Model of Sound Change*

Andrea Ceolin and Ollie Sayeed

*A Method to Automatically Identify Diachronic Variation in Collocations.*

Marcos Garcia and Marcos García Salido

*Written on Leaves or in Stones?: Computational Evidence for the Era of Authorship of Old Thai Prose*

Attapol Rutherford and Santhawat Thanyawong

*Identifying Temporal Trends Based on Perplexity and Clustering: Are We Looking at Language Change?*

Sidsel Boldsen, Manex Agirrezabal and Patrizia Paggio

*Using Word Embeddings to Examine Gender Bias in Dutch Newspapers, 1950-1990*

Melvin Wevers

*Ab Antiquo: Proto-language Reconstruction with RNNs*

Carlo Meloni, Shauli Ravfogel and Yoav Goldberg

*Predicting Historical Phonetic Features using Deep Neural Networks: A Case Study of the Phonetic System of Proto-Indo-European*

Frederik Hartmann

*ParHistVis: Visualization of Parallel Multilingual Historical Data*

Aikaterini-Lida Kalouli, Rebecca Kehlbeck, Rita Sevastjanova, Katharina Kaiser, Georg A. Kaiser and Miriam Butt

**August 2, 2019 (continued)**

*Tracing Antisemitic Language Through Diachronic Embedding Projections: France 1789-1914*

Rocco Tripodi, Massimo Warglien, Simon Levis Sullam and Deborah Paci

*DiaHClust: an Iterative Hierarchical Clustering Approach for Identifying Stages in Language Change*

Christin Schätzle and Hannah Booth

*Treat the Word As a Whole or Look Inside? Subword Embeddings Model Language Change and Typology*

Yang Xu, Jiasheng Zhang and David Reitter

*Times Are Changing: Investigating the Pace of Language Change in Diachronic Word Embeddings*

Stephanie Brandl and David Lassner

*The Rationality of Semantic Change*

Omer Korat

*Studying Laws of Semantic Divergence across Languages using Cognate Sets*

Ana Uban, Alina Maria Ciobanu and Liviu P. Dinu

*Detecting Syntactic Change Using a Neural Part-of-Speech Tagger*

William Merrill, Gigi Stark and Robert Frank

*Grammar and Meaning: Analysing the Topology of Diachronic Word Embeddings*

Yuri Bizzoni, Stefania Degaetano-Ortlieb, Katrin Menzel, Pauline Krielke and Elke Teich

*Spatio-Temporal Prediction of Dialectal Variant Usage*

Péter Jeszenszky, Panote Siriaraya, Philipp Stoeckle and Adam Jatowt

*One-to-X Analogical Reasoning on Word Embeddings: a Case for Diachronic Armed Conflict Prediction from News Texts*

Andrey Kutuzov, Erik Velldal and Lilja Øvreliid

*Measuring Diachronic Evolution of Evaluative Adjectives with Word Embeddings: the Case for English, Norwegian, and Russian*

Julia Rodina, Daria Bakshandaeva, Vadim Fomin, Andrey Kutuzov, Samia Touileb and Erik Velldal

*Semantic Change in the Language of UK Parliamentary Debates*

Gavin Abercrombie and Riza Batista-Navarro

**August 2, 2019 (continued)**

*Semantic Change and Emerging Tropes In a Large Corpus of New High German Poetry*

Thomas Haider and Steffen Eger

*Conceptual Change and Distributional Semantic Models: an Exploratory Study on Pitfalls and Possibilities*

Pia Sommerauer and Antske Fokkens

*Measuring the Compositionality of Noun-Noun Compounds over Time*

Prajit Dhar, Janis Pagel and Lonneke van der Plas

*Towards Automatic Variant Analysis of Ancient Devotional Texts*

Amir Hazem, Béatrice Daille, Dominique Stutzmann, Jacob Currie and Christine Jacquin

*Understanding the Evolution of Circular Economy through Language Change*

Sampriti Mahanty, Frank Boons, Julia Handl and Riza Theresa Batista-Navarro

*Gaussian Process Models of Sound Change in Indo-Aryan Dialectology*

Chundra Cathcart

**16:00–16:40 Session 5**

16:00–16:20 *Modeling a Historical Variety of a Low-Resource Language: Language Contact Effects in the Verbal Cluster of Early-Modern Frisian*

Jelke Bloem, Arjen Versloot and Fred Weerman

16:20–16:40 *Visualizing Linguistic Change as Dimension Interactions*

Christin Schätzle, Frederik L. Dennig, Michael Blumenschein, Daniel A. Keim and Miriam Butt

**16:40–17:15 Discussion and Closing**