

Call for papers for a special issue of the journal *Computational linguistics* on: “Computational approaches in historical linguistics after the quantitative turn”

Deadline: July 15, 2018

Guest editors

Taraka Rama – University of Oslo <tarakark@ifi.uio.no>
Simon J. Greenhill – Max Planck Institute for the Science of Human History <greenhill@shh.mpg.de>
Harald Hammarström – Uppsala University <harald.hammarstrom@lingfil.uu.se>
Gerhard Jäger – Eberhard Karls University Tübingen <gerhard.jaeger@uni-tuebingen.de>
Johann-Mattis List – Max Planck Institute for the Science of Human History <mattis.list@shh.mpg.de>

Call for papers

Computational approaches play an increasingly important role in historical linguistics and typology. During the last two decades, scholars have made significant advances in automatizing and formalizing specific aspects of the workflow of the classic comparative method. Among these are novel techniques for phonetic alignment (Prokić et al. 2009, List 2014), the detection of cognate words (Kondrak 2009, Rama 2016, List et al. 2017, Jäger et al. 2017), and improved methods for phylogenetic reconstruction (Chang et al. 2015, Bouckaert et al. 2012), complemented by numerous pioneering approaches investigating specific aspects of language change, such as geographic diffusion (Prokić and Cysouw 2013), semantic shift (Dellert 2016) and regular sound change (Hruschka et al. 2015, Bouchard-Côté et al. 2013), and even fully automated work flows for large-scale language comparison have been proposed (Brown et al. 2008, Jäger 2015, Rama and Borin 2015). On the other hand, availability of typological databases such as WALS has allowed scholars to investigate specific hypotheses regarding evolution of grammar and lexicon (Greenhill et al. 2017), the interaction of typological features (Hammarström and O’Connor 2013), and the areal distribution of linguistic structures (Daume III 2009, Chang and Michael 2014).

Despite the growing interest in automated approaches to historical linguistics, it is obvious that automatic approaches are still far away from being able to replace human experts, and the majority of approaches still makes broad use of manually annotated datasets. More than two decades after the quantitative turn in historical linguistics, we think it is time to reconsider how computational approaches in historical linguistics can be further improved, and where their current limits can be found.

Topics of interest

We invite submissions of papers which address one or several of the following questions:

1. How can we improve the automatic identification of cognates?
2. How can computational methods help to infer deeper genetic relations between the world’s language families?
3. How can big data approaches from computational linguistics help to improve classical approaches to historical linguistics?
4. What are the strengths and shortcomings of phylogenetic methods?
5. How does demography and geography influence the spread of languages through time and space?
6. Are there universal tendencies in the evolution of the world’s languages?
7. How to integrate typological features with lexical features for inferring language phylogenies and predicting typological features for ancestral languages?

We strongly emphasize the role of sustainable research and therefore ask all submissions which involve data and/or code that are not already publically available to submit these along with the

paper anonymously. Papers which do not conform to this requirement will not be considered for publication.

Paper format and reviewing policy

Papers should be formatted according to the Computational Linguistics style (<http://cljournal.org/>) and submitted using the online submission system (<http://cljournal.org/submissions.html>). In Step 1 of the submission process, please select “Special Issue: Computational historical linguistics” under the “Journal Section” heading. Please note that papers submitted to a special issue undergo the same reviewing process as regular papers. Special issues are the same length as regular issues (at most 5-6 papers, see <http://cljournal.org/specialissues.html> for more information).

Deadline

Paper submission deadline: July 15, 2018 (23:59 PST)

Contact

calc@digling.org

References

- Bouchard-Côté, A., D. Hall, T. Griffiths, and D. Klein (2013): **Automated reconstruction of ancient languages using probabilistic models of sound change**. *Proceedings of the National Academy of Sciences* 110.11. 4224–4229.
- Bouckaert, R., P. Lemey, M. Dunn, S. Greenhill, A. Alekseyenko, A. Drummond, R. Gray, M. Suchard, and Q. Atkinson (2012): **Mapping the origins and expansion of the Indo-European language family**. *Science* 337.6097. 957-960.
- Brown, C., E. Holman, S. Wichmann, V. Velupillai, and M. Cysouw (2008): **Automated classification of the world’s languages. A description of the method and preliminary results**. *Sprachtypologie und Universalienforschung* 61.4. 285-308.
- Chang, W., C. Cathcart, D. Hall, and A. Garrett (2015): **Ancestry-constrained phylogenetic analysis support the Indo-European steppe hypothesis**. *Language* 91.1. 194-244.
- Chang, W. and L. Michael. (2014): **A relaxed admixture model of contact**. *Language Dynamics and Change* 4(1). 1-26.
- Daumé III, Hal. (2009): **Non-parametric Bayesian areal linguistics**. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL '09)*, 593--601. Morristown, NJ, USA: Association for Computational Linguistics.
- Dellert, J. and A. Buch (2016): **Using computational criteria to extract large Swadesh Lists for lexicostatistics**. In: *Proceedings of the Leiden Workshop on Capturing Phylogenetic Algorithms for Linguistics*.
- Greenhill, S. J., C. H. Wu, X. Hua, M. Dunn, S. C. Levinson, & R. D. Gray, (2017): **Evolutionary dynamics of language systems**. *Proceedings of the National Academy of Sciences* 114, no. 42 (2017): E8822-E8829.
- Hammarström, H. and L. O’Connor. (2013): **Dependency Sensitive Typological Distance**. In Lars Borin & Anju Saxena (eds.), *Approaches to measuring linguistic differences*, 337-360. Berlin: Mouton.
- Hruschka, D., S. Branford, E. Smith, J. Wilkins, A. Meade, M. Pagel, and T. Bhattacharya (2015): **Detecting regular sound changes in linguistics as events of concerted evolution**. *Curr. Biol.* 25.1. 1-9.
- Jäger, G. (2015): **Support for linguistic macrofamilies from weighted alignment**. *Proceedings of the National Academy of Sciences* 112.41. 12752–12757.
- Jäger, G., J.-M. List, and P. Sofroniev (2017): **Using support vector machines and state-of-the-art algorithms for phonetic alignment to identify cognates in multi-lingual wordlists**. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. Long Papers*. 1204-1215.
- Kondrak, G. (2009): **Identification of Cognates and Recurrent Sound Correspondences in Word Lists**. *Traitement Automatique des Langues* 50(2). 201-235.
- List, J.-M. (2014): **Sequence comparison in historical linguistics**. Düsseldorf University Press: Düsseldorf.
- List, J. M., S. J. Greenhill, & R. D. Gray, (2017): **The potential of automatic word comparison for historical linguistics**. *PLOS ONE* 12.1. 1-18.
- Prokić, J. and M. Cysouw. (2013): **Combining Regular Sound Correspondences and Geographic Spread**. *Language Dynamics and Change* 3. 147-168.
- Prokić, J., M. Wieling, & J. Nerbonne (2009): **Multiple sequence alignments in linguistics**. In: *Proceedings of the EACL 2009 Workshop on Language Technology and Resources for Cultural Heritage, Social Sciences, Humanities, and Education*.
- Rama, T. (2016): **Siamese Convolutional Networks for Cognate Identification**. In *COLING*, pp. 1018-1027.
- Rama, T. and L. Borin (2015): **Comparative evaluation of string similarity measures for automatic language classification**. In Ján Mačutek and George K. Mikros, editors, *Sequences in Language and Text*, pages 203–231. Berlin: Walter de Gruyter.