

All the publications are peer-reviewed in a single or double blind process.

Publications

Rama, Taraka and Johann-Mattis List. An automated framework for fast cognate detection and Bayesian phylogenetic inference in computational historical linguistics. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 6225–6235, Florence, Italy, July 2019. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/P19-1627>.

Stig Johan Berggren, **Rama, Taraka**, and Lilja Øvrelid. Regression or classification? automated essay scoring for Norwegian. In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 92–102, Florence, Italy, August 2019. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/W19-4409>.

Brad Aiken, Jared Kelly, Alexis Palmer, Suleyman Olcay Polat, **Rama, Taraka**, and Rodney Nielsen. Sigmorphon 2019 task 2 system description paper: Morphological analysis in context for many languages, with supervision from only a few. In *Proceedings of the 16th Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 87–94, Florence, Italy, August 2019. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/W19-4211>.

Çağrı Çöltekin, **Rama, Taraka**, and Verena Blaschke. Tübingen-Oslo team at the vardial 2018 evaluation campaign: An analysis of n-gram features in language variety identification. In *Proceedings of the Fifth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2018)*, pages 55–65. Association for Computational Linguistics, 2018. URL <http://aclweb.org/anthology/W18-3906>.

Rama, Taraka, Johannes Wahle, Pavel Sofroniev, and Gerhard Jäger. Fast and unsupervised methods for multilingual cognate clustering. *Language Dynamics and Change*, Forthcoming.

Rama, Taraka and Søren Wichmann. Towards identifying the optimal data-size for lexically-based bayesian inference of linguistic phylogenies. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1578–1590. Association for Computational Linguistics, 2018. URL <http://aclweb.org/anthology/C18-1134>.

Rama, Taraka. Three tree priors and five datasets: A study of the effect of tree priors in Indo-European phylogenetics. *Language Dynamics and Change*, Forthcoming.

- Rama, Taraka**, Johann-Mattis List, Johannes Wahle, and Gerhard Jäger. Are automatic methods for cognate detection good enough for phylogenetic reconstruction in historical linguistics? In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 393–400. Association for Computational Linguistics, 2018. URL <http://aclweb.org/anthology/N18-2063>.
- Sowmya Vajjala and **Rama, Taraka**. Experiments with Universal CEFR classification. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 147–153. Association for Computational Linguistics, 2018. URL <http://aclweb.org/anthology/W18-0515>.
- Çağrı Çöltekin and **Rama, Taraka**. Exploiting universal dependencies treebanks for measuring morphosyntactic complexity. In Christian Bentz and Aleksandrs Berdicevskis, editors, *Proceedings of First Workshop on Measuring Language Complexity*, pages 1–7, 2018.
- Çağrı Çöltekin and **Rama, Taraka**. Tübingen-Oslo at SemEval-2018 Task 2: SVMs perform better than RNNs in Emoji Prediction”. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 34–38. Association for Computational Linguistics, 2018. URL <http://aclweb.org/anthology/S18-1004>.
- Rama, Taraka** and Sowmya Vajjala. A Dependency Treebank for Telugu. In *Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories*, pages 119–128, Prague, Czech Republic, 2017. Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/W17-7616>.
- Rama, Taraka** and Çağrı Çöltekin. Fewer features perform well at native language identification task. In *Proceedings of The 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 255–260, Copenhagen, Denmark, 2017. Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/W17-5028>.
- Søren Wichmann and **Rama, Taraka**. Jackknifing the black sheep: ASJP classification performance and Austronesian. In Ritsuko Kikusawa and Lawrence A. Reid, editors, *Let’s Talk about Trees: Genetic Relationships of Languages and Their Phylogenetic Representation*. National Museum of Ethnology, Osaka, Japan, 2018.
- Roland Mühlenbernd and **Rama, Taraka**. What phoneme networks tell us about the age of language families. *Journal of Language Evolution*, 2(1):67–76, 2017. doi: 10.1093/jole/lzx007. URL <http://dx.doi.org/10.1093/jole/lzx007>.

- Rama, Taraka**, Çağrı Çöltekin, and Pavel Sofroniev. Computational analysis of Gondi dialects. In *Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial)*, pages 26–35, Valencia, Spain, April 2017. Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/W17-1203>.
- Çağrı Çöltekin and **Rama, Taraka**. Tübingen system in VarDial 2017 shared task: Experiments with language identification and cross-lingual parsing. In *Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial)*, pages 146–155, Valencia, Spain, April 2017. Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/W17-1218>.
- Rama, Taraka**. Siamese convolutional networks for cognate identification. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1018–1027. The COLING 2016 Organizing Committee, 2016a. URL <http://www.aclweb.org/anthology/C16-1097>.
- Çağrı Çöltekin and **Rama, Taraka**. Discriminating similar languages with linear svms and neural networks. In *Proceedings of the Third Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial3)*, pages 15–24, Osaka, Japan, December 2016. The COLING 2016 Organizing Committee. URL <http://aclweb.org/anthology/W16-4802>.
- Rama, Taraka** and Çağrı Çöltekin. LSTM autoencoders for dialect analysis. In *Proceedings of the Third Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial3)*, pages 25–32, Osaka, Japan, December 2016. The COLING 2016 Organizing Committee. URL <http://aclweb.org/anthology/W16-4803>.
- Rama, Taraka**. Ancestry sampling for Indo-European phylogeny. In *Proceedings of the Leiden Workshop on Capturing Phylogenetic Algorithms for Linguistics*. University of Tübingen, online publication system, 2016b.
- Rama, Taraka** and Lars Borin. 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. Walter de Gruyter, 2015.
- Rama, Taraka**. Automatic cognate identification with gap-weighted string subsequences. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, May 31 – June 5, 2015 Denver, Colorado, USA*, pages 1227–1231, 2015. ISBN 978-1-941643-49-5. URL <http://aclweb.org/anthology/N/N15/N15-1130.pdf>.

- Lars Borin, Anju Saxena, **Rama, Taraka**, and Bernard Comrie. Linguistic landscaping of South Asia using digital language resources: Genetic vs. areal linguistics. In *Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3137–3144, 2014.
- Rama, Taraka** and Lars Borin. N-gram approaches to the historical dynamics of basic vocabulary. *Journal of Quantitative Linguistics*, 21(1):50–64, 2013.
- Rama, Taraka**. Phonotactic diversity predicts the time depth of the world’s language families. *PLOS ONE*, 8(5):1–9, 2013. URL <https://doi.org/10.1371/journal.pone.0063238>.
- Rama, Taraka**, Prasant Kolachina, and Sudheer Kolachina. Two methods for automatic identification of cognates. In *Quantitative Investigations in Theoretical Linguistics*, volume 5, pages 76–80, 2013.
- Rama, Taraka** and Sudheer Kolachina. Distance-based phylogenetic inference algorithms in the subgrouping of Dravidian languages. In Lars Borin and Anju Saxena, editors, *Approaches to Measuring Linguistic Differences*, pages 141–174. De Gruyter Mouton, Berlin, 2013. ISBN 978-3-11-030525-8.
- Rama, Taraka** and Prasanth Kolachina. How good are typological distances for determining genealogical relationships among languages? In *Proceedings of the 24th International Conference on Computational Linguistics*, pages 975–984, 2012.
- Søren Wichmann, Eric Holman, **Rama, Taraka**, and Robert S. Walker. Correlates of reticulation in linguistic phylogenies. *Language Dynamics and Change*, 1(2): 205–240, 2011a.
- Søren Wichmann, **Rama, Taraka**, and Eric Holman. Phonological diversity, word length, and population sizes across languages: The ASJP evidence. *Linguistic Typology*, 15(2):177–197, 2011b. URL <http://dx.doi.org/10.1515/lity.2011.013>.
- Sudheer Kolachina, **Rama, Taraka**, and Lakshmi Bai. Maximum parsimony method in the subgrouping of Dravidian languages. In *Quantitative Investigations in Theoretical Linguistics*, volume 4, pages 52–56, 2011.
- Rama, Taraka** and Lars Borin. Estimating language relationships from a parallel corpus. A study of the Europarl corpus. In *NEALT Proceedings Series (NODAL-IDA 2011 Conference Proceedings)*, volume 11, pages 161–167, 2011. URL <http://hdl.handle.net/10062/17303>.
- Rama, Taraka**, Anil Kumar Singh, and Sudheer Kolachina. Modeling letter-to-phoneme conversion as a phrase based statistical machine translation problem

with minimum error rate training. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Student Research Workshop and Doctoral Consortium*, pages 90–95. Association for Computational Linguistics, 2009a.

Rama, Taraka and Anil Kumar Singh. From bag of languages to family trees from noisy corpus. In *Proceedings of the Conference on Recent Advances in Natural Language Processing*, pages 355–359, Borovets, Bulgaria, 2009.

Anil Kumar Singh, Sethu Subramaniam, and **Rama, Taraka**. Transliteration as alignment vs. transliteration as generation for crosslingual information retrieval. *Traitement Automatique des Langues*, 51(2), 2010.

Rama, Taraka, Sudheer Kolachina, and B. Lakshmi Bai. Quantitative methods for phylogenetic inference in historical linguistics: An experimental case study of South Central Dravidian. *Indian Linguistics*, 70:265–282, 2009b.