Delineating the field of language evolution research:
A quantitative study of submission types and peer-review patterns at the Joint Conference on Language Evolution (JCoLE)

Stefan Hartmann[*1], Yannick Jadoul[2], Rie Asano[3], Evelina Daniela Rodrigues[4], Daria Valente[5], Andrea Ravignani[2], and Slawomir Wacewicz[6]

[*]Corresponding Author: hartmast@hhu.de
[1]German Department, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany
[2]Department of Human Neurosciences, Sapienza University of Rome, Rome, Italy
[3]Advanced Comprehensive Research Organization, Teikyo University, Tokyo, Japan
[4]William James Center for Research, ISPA–Instituto Universitário, Lisbon, Portugal
[5]Department of Life Sciences and Systems Biology, University of Turin
[6]Center for Language Evolution Studies (CLES), Nicolaus Copernicus University, Torun, Poland

Like any field of research (Kuhn, 1970), language evolution research is in constant flux. As such, it seems promising to find bottom-up answers on what is seen as central, and what as more peripheral, in current language evolution research. In this paper, we draw on submission data from the Joint Conference on Language Evolution (JCoLE 2022) to address this question in more detail. Held in September 2022, JCoLE was a joint enterprise of the field's three main conferences (Evolang, Protolang, and Evolinguistics). Following up on previous scientometric studies on Evolang (Bergmann & Dale, 2016, Wacewicz et al., 2023), we use aggregated, anonymized data of the submissions and the peer-review results.

We grouped all submissions by the main research field, as indicated by the submitting author. The core research areas mentioned on the Evolang website are not represented equally. Perhaps unsurprisingly, more than a third of all submissions (37.5%) fall into the linguistics category. Cognitive science is the second most chosen topic (16.9%). Modelling, psychology, primatology, neuroscience, biology follow in this order. When it comes to the relevance scores given by reviewers, the data does not show strong differences between fields. Nonetheless some trends emerge: there is higher consensus in classifying research from biology as relevant; other fields, by contrast, show a broad range of relevance ratings. As could be expected, this higher dispersion is especially true for the many submissions classified under linguistics and cognitive science, but also for e.g. primatology. Papers from psychology and neuroscience show a comparatively lower mean relevance score.

Working with 150-word, Porter-stemmed summaries of submissions (Porter, 1980; implemented in the NTLK Python library, Bird, Klein, & Loper, 2009), we

calculated their *tf-idf* score vectors (term frequency-inverse document frequency). We calculated the pairwise cosine similarity between documents, and used these similarity measures as the edge weights in an undirected graph. On this graph, we ran an off-the-shelf community detection algorithm (Clauset-Newman-Moore greedy modularity maximisation, implemented in the NetworkX Python library; Hagberg, Schult, & Swart, 2008) to find 15 disjunct groups of submissions, clustered by the terms used in these author-provided summaries (Figure 1). To get an idea of the topics of submissions within each group, we removed stopwords from the summaries and again used the tf-idf to determine the most relevant words in each group of submissions. The main finding is that differences between groups in the assessed relevance of the submissions are small, testifying to the inherent multidisciplinary nature of language evolution as a research field. The fact that no research topic is entirely dominated by submissions from a single main research field further highlights this point.

Finally, our investigation of the evaluation of paper types (i.e., empirical, modelling, and theoretical work) shows that the community strongly values empirical work: the rejection rate was about 10% for empirical and modelling papers and about 24% for theoretical papers. This indicates that language evolution is considered an empirically tractable phenomenon by the community members. There is also a substantial and growing tendency to re-use existing data, with roughly 25% of submissions reporting studies working with data already available from databases and corpora. As for experimental work, empirical research using communication game setups is considered highly relevant despite the theoretical concerns with using modern humans as participants. Overall, then, our results offer a simple snapshot of language evolution as a dynamically developing field. As new questions, theories and methods keep emerging, the scope of language evolution research and the concept of language evolution will keep adapting.
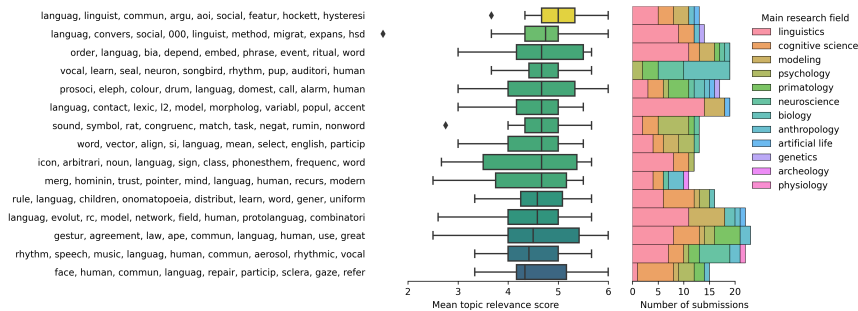

Figure 1. Groups of submissions detected based on the 10 most relevant terms. The boxplots (left) show the distribution of mean relevance scores within each group, and the barplots (right) show the composition of the "main research field" of the group's submissions

References

Bergmann, T., & Dale, R. (2016). A sociometric analysis of Evolang: Intersections and authorships. In S. G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Fehér, & T. Verhoef (Eds.), *The Evolution of Language. Proceedings of the 11th International Conference* (pp. 79–86).

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit.* O'Reilly Media, Inc.

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. In G. Varoquaux, T. Vaught, & J. Millman (Eds.), *Proceedings of the 7th Python in Science conference* (p. 11-15). Pasadena, CA USA.

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions* (2 ed.). Chicago: The University of Chicago Press.

Porter, M. F. (1980). An algorithm for suffix stripping. *Program*, *14*(3), 130–137.

Wacewicz, S., Sibierska, M., Placiński, M., Szczepańska, A., Poniewierska, A., Ng, Y., & Żywiczyński, P. (2023). The scientometric landscape of Evolang: A comprehensive database of the Evolang conference. *Journal of Language Evolution*, lzad003.