Ten years of linguistic diversity in language processing conferences

Aymeric Collart
Academia Sinica

Language processing research needs insights from typologically diverse languages. Anand et al. (2011) showed after exploring 4,000 psycholinguistic studies that only 57 languages were represented, and that 85% of the studies involved 10 languages, primarily Indo-European. In this study, we ask whether linguistic diversity in language processing research increased in the last ten years (2012 to 2022) by exploring the languages investigated in two main conferences: CUNY-HSP and AMLaP. Overall, the 5,549 abstracts cover 105 languages, with the number of languages increasing over the years (28 in 2012 to 51 in 2022). 84.36% of the studies focused on Indo-European languages, and 54.78% on English only. However, from 2012 to 2022, the proportion of English studies decreased (72% to 51%) while it increased for other Indo-European (12% to 29%) and non-Indo-European languages (15% to 18%). Nevertheless, the proportion of WEIRD (around 96%) and non-WEIRD (around 4%) languages did not significantly change. These data indicate that language processing studies got more diverse in the past ten years, even if these only cover a small part of the world's languages, and non-WEIRD languages are still underrepresented. Future investigations looking at language diversity according to different topics (e.g., argument structure processing) are planned.

Selected references

Anand, Pranav, Sandra Chung and Matthew Wagers (2011). *Widening the net: challenges for gathering linguistic data in the digital age*. White paper published in NSF project SBE 2020: Future research in the Social, Behavioral and Economics Sciences.

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., Roehm, D., & Schlesewsky, M. (2011). Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain & Language*, *117*(3), 133–152. https://doi.org/10.1016/j.bandl.2010.09.010

Collart, A. (2023). Experimental linguistics embracing linguistic diversity: On the contributions of Formosan languages to models of sentence processing. In P. Li, E. Zeitoun, & R. De Busser (Eds.), *Handbook of Formosan languages: The indigenous languages of Taiwan* (pp. 21–61), Leiden: Brill.

Dahl, Östen (2015, May 1–3). *How WEIRD are WALS languages?* [Conference presentation]. The Diversity Linguistics: Retrospect and Prospect conference. May 1-3, Leipzig, Germany.

Hammarström, H., Forkel, R., Haspelmath, M., & Bank, S. (2022). Glottolog 4.7. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://doi.org/10.5281/zenodo.7398962 (Available online at http://glottolog.org, Accessed on 2023-05-26.)

Haspelmath, M. (2001). The European linguistic area: Standard Average European. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), *Language typology and language universals: An international handbook* (pp. 1492–1510). Mouton de Gruyter. https://doi.org/10.1515/9783110194265-044

Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition? *First Language (Special issue: How diverse is child language acquisition research?*), *42*(6), 703–735. https://doi.org/10.1177/01427237211066405

Majid, A., & Levinson, S. C. (2010). WEIRD languages have misled us, too. *Behavioral and Brain Sciences*, *33*(2/3), 103. https://doi.org/10.1017/S0140525X1000018X