How diverse is child language research?

Language exhibits a large degree of diversity for which any theory of language-related phenomena must account. However, the treatment of diversity across subdisciplines of Linguistics has been historically uneven. Writing about the large strides in the field of Language Documentation, Seifert et al. (2018) noted that we have some description of approximately 60% of the world's languages, but that continued effort to improve that number is needed because documentation continues to expand the hypothesis space of what is possible in natural language. Psycholinguistics has a comparatively poor record in studying diverse languages: Jaeger and Norcliffe (2009) reported that language production studies have covered only 0.6% of the world's languages. In a wider analysis, Anand et al. (2011) reported that 85% of adult psycholinguistic studies were based on only 10 languages (30% of which were on English). In the current paper we investigated the degree of linguistic diversity in the field of child language acquisition.

We coded every paper published in *Journal of Child Language* (1974 – 2020), *First Language* (1980 – 2020), *Language Acquisition* (1990 – 2020), and *Language Learning and Development* (2005 – 2020) on the following dimensions: (i) languages studied, (ii) topic studied, (iii) country of the authors, (iv) speaker group type (mono- versus multilingual), and (v) socio-economic status of the participant group. Here we focus on (i) – (iii). Overall, 2,834 papers were included in the dataset; an additional 417 were excluded because they did not report substantially new data analyses (e.g., editorials, methodological papers, replies). There were 104 unique languages reported in the data: if we take 6,000 languages as an approximate estimate of the number of language spoken on the planet today, this corresponds to just 1.7% coverage in the field's major journals. The number of languages is severely skewed towards English and other Indo-European languages, as shown in Figure 1, which shows the cumulative number of papers published on English, other Indo-European languages, and all 'Other' languages. The data show an almost exclusive initial focus on English, which weakens though remains strong across time. There were 63 'Other' languages, which came from 27 different language families. However, the distribution of languages in this category was skewed: some language had many papers (e.g., Mandarin, Hebrew, Japanese), while the modal number was 1 (Figure 2).

The distribution of different topics was broadly similar across English, other Indo-European languages, and 'Other' languages (Figure 3), showing that the field studies a broad range of topics independent of language. However, consistent with the bias towards European languages, the locus of research production is overwhelmingly located in the Global North, with the vast majority of papers coming from North America and Europe.

Overall, these data show that, despite a proud history of crosslinguistic research, the field has a lot more work to do to adequately widen its evidential base and widen participation to researchers who speak the languages we hope to include in our journals. We will discuss possible ways of increasing data coverage.

References

Anand et al. (2011). *Widening the net: Challenges for gathering linguistic data in the digital age.* NSF SBE 2020. Rebuilding the mosaic: Future research in the social, behavioral and economic sciences at the National Science Foundation in the next decade.

Jaeger, T. F. & Norcliffe, E. J. (2009). The Cross-linguistic Study of Sentence Production. *Language and Linguistics Compass* 3(4). 866–887.

Seifert, F. *et al.* (2018). Language documentation 25 years on. *Language, 94,* e324 – e345.

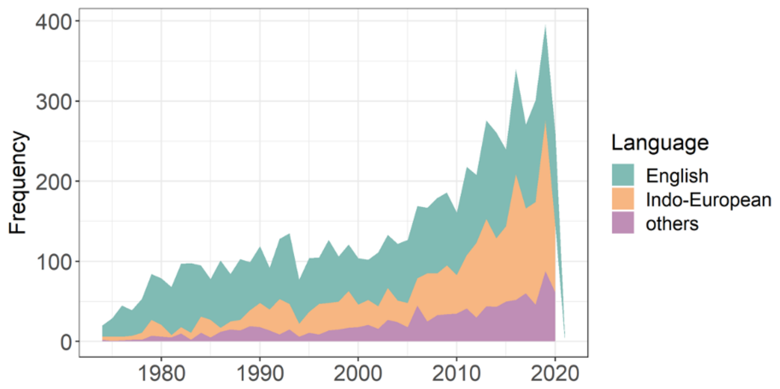

Figure 1. Papers published between 1974-2020 in child language journals in English, other Indo-European languages, and all other languages.

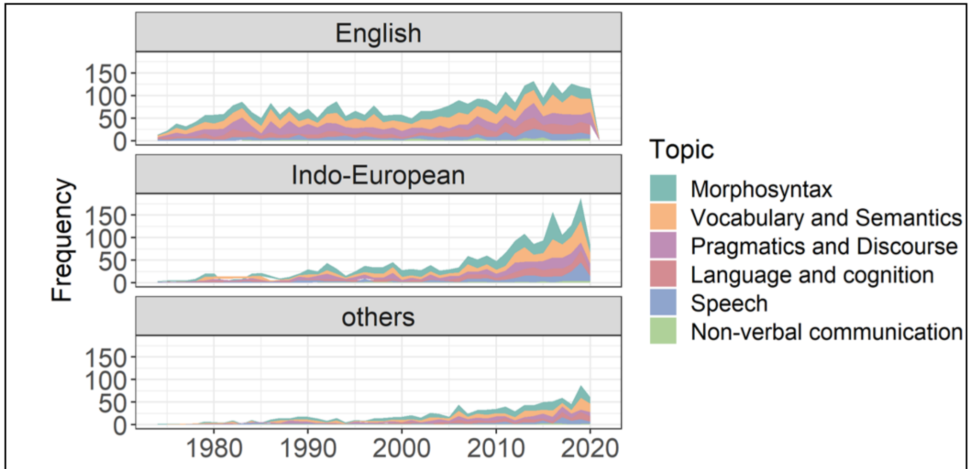

Figure 3. Papers published between 1974-2020 in child language journals in English, other Indo-European languages, and all other languages by topic.

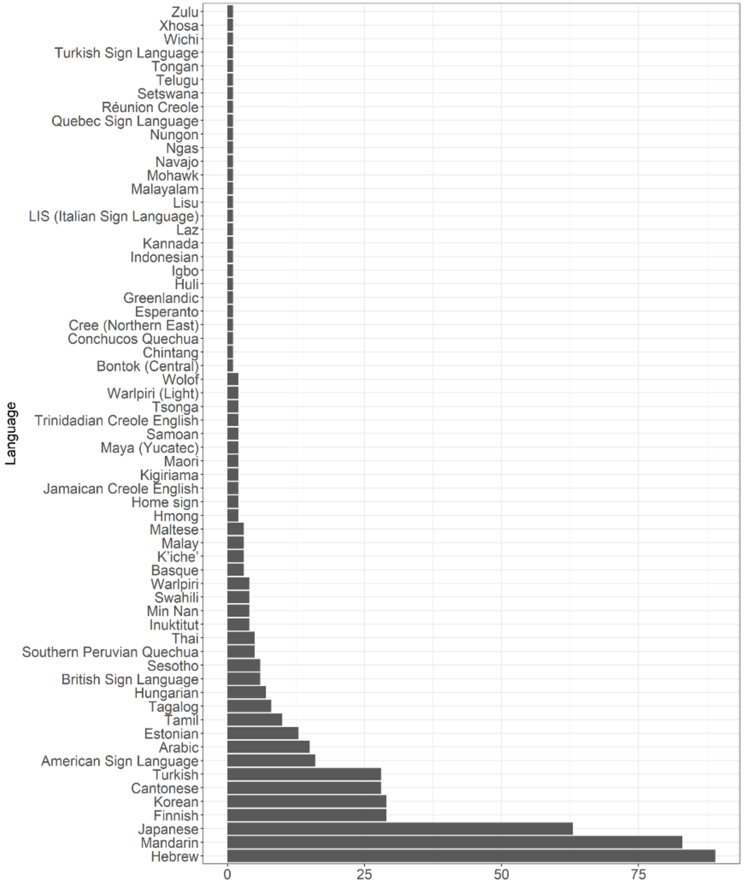

Figure 2. Frequency distribution of papers on non Indo-European papers published between 1974 – 2020.