

Social structure and lexical variation: A cross-linguistic study of three sign languages

Lexical variation is a fundamental characteristic of all languages. It has been claimed that sociolinguistic factors, among them community structure, affect the rate of lexical convergence in early stages of language emergence. Yet results are contradictory as to the direction of the influence of social structure. Agent-based experiments show that smaller populations converge more quickly on a shared set of lexical items (Baronchelli et al., 2006) while studies on ‘emerging’ sign languages show the reverse effect suggesting that community size and social networks shape lexical variation with higher pressures in larger communities to converge (Meir et al., 2012; Meir & Sandler, 2019). To date, few quantitative comparisons with large datasets have tested these claims (Meir & Sandler 2019).

We investigate the claim that lexical variation is higher in smaller communities by comparing three unrelated sign languages: Kata Kolok (KK), British Sign Language (BSL), and Israeli Sign Language (ISL), which differ in terms of language age, community size, and social structure. BSL is at least 260 years old and used by a large social network of signers. ISL and KK are both younger than BSL but ISL is used over a wider geographical distribution (similar to BSL) and KK is used among hearing and deaf members of a Balinese village with more dense social network ties. The comparison of these three sign languages allows us to explore and isolate the social factors that influence lexical variation.

We re-analyse existing data from different lexical elicitation tasks (KK: 36 concepts, 46 signers, Mudd et al. 2020; BSL: 42 concepts, 249 signers, Stamp et al. 2014; ISL: 40 concepts, 62 signers, Stamp et al., 2021). We calculated the lexical distance between pairs of participants from the same language community by scoring whether two participants produced the same or different signs in response to each stimulus, resulting in a lexical distance matrix. We then compared the lexical distance matrix for each dataset to each other.

This cross-linguistic comparison does not support the claim that smaller and younger sign languages show higher degrees of lexical variation than larger and older sign languages (Figure 1A). In fact, BSL shows the highest average distance and highest standard deviation on the global level ($M = 0.57$; $SD = 0.14$). On the local level (e.g. comparing lexical distances of signers from the same region for BSL), the sociolinguistic variables governing variation are not always clear; variation is lowest for BSL, confirming previous work that found strong regional variation (Figure 1B; Stamp et al. 2014) and higher for KK and ISL, suggesting that no single sociolinguistic factor governs lexical variation.

Understanding what factors shape lexical variation within communities requires cross-linguistic comparisons and novel methodologies. Our results challenge the claim that smaller communities exhibit greater lexical variation and highlight the importance of quantitative comparisons to test assumptions and to disentangle the sociolinguistic variables driving variation. We show that variation measures need nuance: differentiating between global and local variation yields different results.

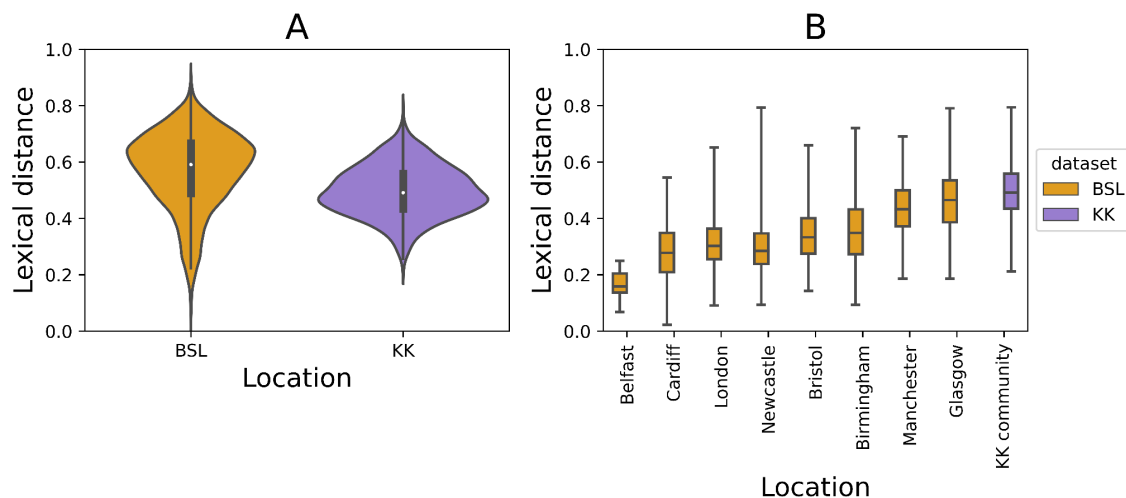


Figure 1. Lexical distances in BSL and KK. A) Violin-plot showing the median (white dot), the interquartile range (thick line) and 1.5x the interquartile range (line) of the lexical distances across the KK dataset (1045 distances) and the BSL dataset (25425 distances). B) boxplot showing lexical distance by region in the BSL dataset (only lexical distances of participants within the same region; 3358 distances) compared to the KK dataset (1045 distances); middle line representing the median, the box the interquartile range and the whiskers the range of the data points (0 to 100%).

References

- Baronchelli, A., Felici, M., Loreto, V., Caglioti, E., & Steels, L.** (2006). Sharp transition towards shared vocabularies in multi-agent systems. *Journal of Statistical Mechanics: Theory and Experiment*, (2006) P06014. <https://doi.org/10.1088/1742-5468/2006/06/P06014>
- Meir, I., Israel, A., Sandler, W., Padden, C., & Aronoff, M.** (2012). The influence of community on language structure: Evidence from two young sign languages. *Linguistic Variation*, 12(2), 247–291. <https://doi.org/10.1075/lv.12.2.04mei>
- Meir, I., & Sandler, W.** (2019). Variation and conventionalization in language emergence: The case of two young sign languages of Israel. In E. Doron, M. R. Hovav, Y. Reshef, & M. Taube (Eds.), *Linguistik Aktuell/Linguistics Today* (Vol. 256, pp. 337–363). John Benjamins Publishing Company. <https://doi.org/10.1075/la.256.13mei>
- Mudd, K., Lutzenberger, H., de Vos, C., Fikkert, P., Crasborn, O., & de Boer, B.** (2020). The effect of sociolinguistic factors on variation in the Kata Kolok lexicon. *Asia-Pacific Language Variation*, 6(1), 53–88. <https://doi.org/10.1075/aplv.19009.mud>
- Schembri, A., Fenlon, J., Rentelis, R. & Cormier, K.** (2017). *British Sign Language Corpus Project: A corpus of digital video data and annotations of British Sign Language 2008-2017* (Third Edition). London: University College London. (<https://www.bslcorpusproject.org>)
- Stamp, R., Schembri, A., Fenlon, J., Rentelis, R., Woll, B., & Cormier, K.** (2014). Lexical variation and change in British Sign Language. *PLOS ONE*, 9(4), e94053. <https://doi.org/10.1371/journal.pone.0094053>
- Stamp, R., Ohanin, O. & Lanesman, S.** (2022). *The Corpus of Israeli Sign Language*. Conference Proceedings (LREC): Language Resources (LRs) and Evaluation for Human Language Technologies (HLT).