A Dynamic Network Approach to bilingual child data

Antje Quick[1], Stefan Hartmann[2] & Paul Ibbotson3[3]
[1]University of Leipzig, antje.quick@uni-leipzig.de [2] University of Düsseldorf, [3] The Open University

Keywords: Bilingual Language Acquisition, Patterns, Dynamic Network

Languages allow us to construct an astronomic variety of utterances by relying on a rather limited number of words. This compositional nature is one of the most remarkable features of language and has led some researchers to conclude that it would be impossible for children to learn a language by relying solely on basic cognitive mechanisms. However, usage-based paradigms have shown that children's early language is remarkably restricted and constructed around lexically specific frames; productivity gradually emerges in a piecemeal way showing an intimate relationship between linguistic knowledge and input (e.g. Tomasello 2003).

Most work on language acquisition has concentrated mainly on monolingual settings. However, most children actually grow up in a multilingual environment and often produce code-mixed utterances, cf. e.g. German-English utterances like *was ist dein picture* 'What is your picture' (Fion, 03;02.12). To systematically detect patterns in bilingual children and to relate them to the input they receive, we need to have a reliable data-driven approach. A number of such pattern detection approaches have already been proposed in the literature, e.g. traceback (Dąbrowska & Lieven 2005, Hartmann et al. 2021). A fairly recent addition to the repertoire of exploratory data-driven methods is Ibbotson et al.'s (2019) dynamic network model (DNM), which uses simple frequency measures to detect networks of co-occurring words (and part-of-speech tags) from the data. In particular, the DNM combines two measures: word frequencies and transition probabilities, gradually building up a network of patterns of use based on distributional information. This allows for identifying what Ibbotson et al. (2019) call "communities" in the data. In Ibbotson et al.'s (2019) original application of the method, it was applied to child-directed speech, showing that the emerging patterns cluster together to form clear "hubs", which in turn can be assumed to facilitate language acquisition. We argue that the same approach can also be used to investigate children's data, especially when dealing with code-mixed utterances.

As such, the aim of this paper is to model the "building blocks" of Fion' early language, a 2-to-3-year-old German-English bilingual child (n = 47,812 utterances). We want to detect patterns in his language use based on transitional probabilities as applied in a DNM (Ibbotson et al. 2019). Additionally, we want to relate those patterns to the child-directed speech he hears from their caregivers (n = 228,221 utterances). Finally, we want to investigate whether frequently co-occurring words are clustered together and are predictive of the child's code-mixed utterances.

References

Dąbrowska, Ewa, and Elena Lieven. 2005. Towards a lexically specific grammar of children's question constructions. Cognitive Linguistics 16(3): 437–474.

Hartmann, Stefan, Nikolas Koch, and Antje Endesfelder Quick. 2021. The traceback method in child language acquisition research: Identifying patterns in early speech. Language and Cognition 13: 227–53.

Ibbotson, Paul, Vsevolod Salnikov, and Richard Walker. 2019. A dynamic network analysis of emergent grammar. First Language, 39(6), 652–680.

Tomasello, Michael. 2003. Constructing a Language: A Usage-Based Theory of Language Acquisition. Cambridge and London: Harvard University Press.