Recycling constructional patterns: The role of chunks in child code-mixing.

Nikolas Koch
LMU Munich, koch@daf.lmu.de

Keywords: bilingual first language acquisition, code-mixing, computational methods

Language mixtures are often regarded as particularly creative utterances of children that cannot be attributed to the respective input (cf. De Houwer 2009: 44). In recent years these utterances have seen increased interest from a usage-based perspective. In usage-based approaches to monolingual acquisition, a number of methods have been developed that allow for detecting patterns from usage data. In this talk, we evaluate one of these methods with regard to its performance when applied to code-mixing (CM) data: the so-called Chunk-based-Learner (CBL) model (McCauley & Christiansen 2017, 2019). This method makes it possible to automatically detect patterns in speech data. In the process, the CBL focuses on chunking processes. However, unlike in most quantitative methods, chunks are not defined by fixed thresholds, but based on backward transitional probabilities (BTP). This method calculates how likely it is that a current word is preceded by the respective previous word. On this basis, chunks and chunk boundaries are determined in the utterances of a German-English bilingual boy. In addition, his input is also taken into account. A total of 228,221 utterances were included in the analysis, 3492 of which were CM utterances of the boy.

First results show that the patterns found in the CM utterances can largely be traced back to the input in the form of chunks. Here, about two-thirds (3,080 utterances) of the language boundaries within CM utterances coincide fully ([ein kleinen][shak]; 'a little shark') or partially ([I have][that werkzeug]; 'I have the tool') with chunk boundaries. In 426 CM utterances, there is no overlap between speech and chunk boundaries. These are often bilingual chunks such as [time out machen] ('take a timeout') or [ein anderer frog] ('another frog') which have no occurrences in the parental input but might be consolidated by the brother, who is a bilingual child himself. The question arises to what extent the extracted chunks in the boy's CM utterances can be attributed to input from his parents, his brother, or to his own utterances. Especially in the case of the bilingual chunks, we expect them to be consolidated not by his caregivers but by utterances of his brother or himself. Furthermore, we will test the hypothesis that the chunks that adhere to language boundaries can be attributed primarily to input from caregivers. In this context, the question arises whether these are also the chunks that occur frequently in the boy's monolingual utterances. The results and advantages and disadvantages of the method are discussed in the light of a usage-based approach (see also Koch et al. 2022).

References

De Houwer, Annick. 2009. *Bilingual First Language Acquisition*. Bristol: Multilingual Matters.

Koch, Nikolas, Stefan Hartmann & Antje Endesfelder Quick. 2022. Traceback and chunk-based learning: Comparing usage-based computational approaches to code-switching. In: *Languages* 7(4), 271; doi.org/10.3390/languages7040271.

McCauley, Stewart M. & Morten H. Christiansen. 2017. Computational Investigations of Multiword Chunks in Language Learning. *Topics in Cognitive Science* 9(3). 637-652.

McCauley, Stewart M. & Morten H. Christiansen. 2019. Language learning as language use: A cross-linguistic model of child language development. *Psychological Review* 126(1). 1-51.