Bilingual language acquisition from a usage-based perspective:
A corpus study on the code-mixing of a German-English bilingual child

Stefan Hartmann[1], Antje Quick[2], Ad Backus[3], Elena Lieven[4]
[1]University of Bamberg, [2]University of Leipzig, [3]Tilburg University, [4]University of Manchester
[1]stefan1.hartmann@uni-bamberg.de, [2]antje.quick@uni-leipzig.de, [3]a.m.backus@uvt.nl,
[4]elena.lieven@manchester.ac.uk

Keywords: Language acquisition, multilingualism, traceback, acquisition of constructions

Lexically specific patterns such as *What's this?* and patterns consisting of a lexically specific "frame" and an open slot such as *look, an* x! play an important role in language acquisition scenarios (Cameron-Faulkner, Lieven, and Tomasello 2003). Recently, Quick et al. (2018) have shown that partially schematic patterns also seem to play a key role in the code-mixing of bilingual children, e.g. *I want* x, as in *I want die paint* 'I want the paint'. Following up on this observation, the present study investigates whether lexically specific as well as partially schematic patterns in the code-mixing of a German-English bilingual child (n=1,024 code-mixed utterances out of 47,812 utterances in total) aged 2;3 to 3;11 can be traced back to patterns found in the input (n=61,077). In line with usage-based accounts of language acquisition (e.g. Tomasello 2003), this would suggest that bilingual children construct their code-mixed utterances on the basis of concrete linguistic material they have heard before.

In our operationalization of utterance-initial chunks as well as partially schematic patterns, we follow e.g. Dąbrowska & Lieven (2005) or Cameron-Faulkner et al. (2003): Repeated strings are considered patterns if they occur at least four times in the corpus. In a first step, only utterance-initial *n*-grams were retrieved automatically to identify recurring lexical patterns. It could be shown that a large proportion of the code-mixed data contain utterance-initial lexical chunks (82.2%), a large proportion of which (72.7 %) can be traced back to the parental input.

A more in-depth study of a subset of the data using manual annotation, which allows for identifying partially schematic rather than just lexically specific chunks, reveals that code-mixing often involves the filling of an open slot in a partially schematic construction. In addition, much of the code-mixing turns out to be primed by the occurrence of the same forms in the immediately preceding discourse. In sum, these findings lend further support to usage-based theories of language acquisition, and they can help understand code-mixing in early bilingual language acquisition from a construction-based perspective.

References

Cameron-Faulkner, Thea, Elena Lieven & Michael Tomasello. 2003. A construction based analysis of child directed speech. *Cognitive Science* 27(6). 843–873.

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

Lieven, Elena, Dorothé Salomo & Michael Tomasello. 2009. Two-year-old Children's Production of Multiword Utterances: A Usage-Based Analysis. *Cognitive Linguistics* 20. 481–507.

Quick, Antje Endesfelder, Elena Lieven, Malinda Carpenter & Michael Tomasello. 2018. Identifying partially schematic units in the code-mixing of an English and German speaking child. *Linguistic Approaches to Bilingualism* 8(4). 477–501.

Stoll, Sabine, Kirsten Abbot-Smith & Elena Lieven. 2009. Lexically Restricted Utterances in Russian, German, and English Child-Directed Speech. *Cognitive Science* 33(1). 75–103.

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