

## **Bi-modal bi-Lingual bi-National corpora of child language**

Ronice QUADROS (UFSC)  
Diane LILLO-MARTIN (University of Connecticut)  
Deborah CHEN PICHLER (Gallaudet University)

In order to address numerous linguistic research questions, we have been building several corpora of sign language acquisition data. Until recently, our focus had been on sign language only with deaf children, from deaf parents, acquiring sign language as native language; now we are expanding our work to include bimodal bilingual children acquiring both a sign language and a spoken language, building comparable corpora across two sign/spoken language pairs. This adds considerable complexity to the already challenging prospect of corpus building. In this presentation, we explore some of the issues we have faced already and those we expect to face, in the context of our linguistic goals.

Recent research on childhood bilingualism has indicated that although children have two separate developing grammatical systems from very early on, there are instances of cross-linguistic influence, where grammatical structures from one language seem to exert atemporary influence on the child's grammar of the other language (e.g. Hulk & Müller 2000). An important question is to identify the loci of such influences based on linguistic criteria. In order for us to address such issues, we are developing corpora from individual children acquiring both a sign language and a spoken language. Many of the same data collection issues arise as those for projects investigating only sign language (see Baker & Woll 2005 for some best practices in this domain). However, in our BiBiBi project we are collecting samples of both sign language and spoken language – and, it turns out that this frequently means code-blended language as well (Bogaerde & Baker 2005, 2009; Emmorey et al. 2008).

We use different sets of researchers (deaf and hearing) to emphasize appropriate target language use, assuming the child's interlocutor sensitivity (Petitto et al. 2001), but we also recognize that code-blending is simply a part of the language system being acquired.

Following video collection, we invest considerable energy in the production of transcripts, to be used in conjunction with the videos for linguistic analyses. Following our earlier sign-only research, we use ELAN for time-locked videos with transcription (<http://www.latmpi.eu/tools/elan/>).

For bilingual research, we designed a different template so that both languages are parent tiers, to optimize the study of (sequential or simultaneous) bimodal productions. We have not yet resolved the following linguistic issues, but we hope that our project will contribute to their discussion in the field as a whole. Does bimodal bilingualism lead to cross-language influence different from that found in mono-modal bilingualism (e.g., due to code-blending, use of non-manuals)? When bimodal bilinguals code-blend, are they choosing grammatical structures which are permitted in both languages for maximum accommodation? What kinds of syntactic representations can account for code-blends?

Finally, we see a number of important implications and extensions of the system we are developing. For example, we are creating a specific identification for each sign to be used in our transcripts (in the same spirit of Johnson, in preparation, for Australian Sign Language), what we call "ID sign". Because there is no commonly accepted writing system for sign languages, sign researchers generally rely on a system of glossing; however, traditional transcription does not assign a consistent gloss for each sign, but different glosses depending on context and other aspects of the signed utterance. This means that it is very difficult for researchers to identify the locations of interest in a transcript using a search function to

discover all occurrences of a particular sign. Analysis must proceed at a much slower pace of handsearching transcripts one utterance at a time. In order to facilitate and expand the analysis of data collected in the parent project, the competitive revision application proposes the development of an ID sign lexicon containing the vocabulary items used most frequently by the children we are studying. ID signs are word labels chosen to represent each sign root systematically, so that every use of the sign has the same label, despite contextual or morphological differences which affect how the sign is interpreted. By using ID signs in our transcripts, we will be able to conduct our analyses more efficiently, using a wider range of data. The proposed ID sign lexicon will address the problem of transcript searchability and greatly facilitate the analysis of data to be collected for sign language corpora. This helps to standardize annotations and it can be freely accessed by other researchers. Another example, many similar issues arise in the study of co-speech gesture. Our template and corpus-building decisions can be applicable to the development of adult only bimodal bilingual corpora. And, we hope our collaboration across continents may contribute to and promote cross-linguistic research on sign languages as well.

## References

- Hulk, A. & Müller, N. (2000) Bilingual first language acquisition at the interface between syntax and pragmatics. In *Bilingualism: Language and Cognition* 3 (3), 2000, 227-244. Cambridge University Press.
- Baker, A. & Woll, B. (Eds.) (2009). *Sign language acquisition*. Amsterdam: John Benjamins.
- Bogaerde, B. van den & Baker, A.E. (2005). Code-mixing in mother-child interaction in deaf families. *Sign language & linguistics*. 8(1-2), 151-174.
- Bogaerde, B. van den & Baker, A.E. (2009). Bimodal language acquisition in Kodas (kids of deaf adults). In M. Bishop & S.L. Hicks (Eds.), *Hearing, mother-father Deaf: Hearing people in Deaf families*, 99-131. Washington, DC: Gallaudet University Press.
- Emmorey, K., Borinstein, H.B., Thompson, R. & Golan, T.H. (2008). Bimodal bilingualism. *Bilingualism: Language and cognition*. 11(1), 43-61.
- Johnston, T. (in preparation). From archive to corpus: Transcription and annotation in the creation of signed language corpora, manuscript. Department of Linguistics, Macquarie University, Australia.
- Petitto, L. A., Katerelos, M., Levi, B., Gauna, K., Tetrault, K. & Ferraro, V. (2001). Bilingual signed and spoken language acquisition from birth: Implications for the mechanisms underlying early bilingual language acquisition. *Journal of child language*. 28(2), 453-496.