

INVITED COMMENTARY

# Uncovering the development of linguistic knowledge in lesser studied languages

Katherine DEMUTH<sup>1\*</sup> Francina MOLOI<sup>2</sup>, Litsepiso MATLOSA<sup>2</sup> and Mark JOHNSON<sup>1</sup>

<sup>1</sup>Macquarie University

<sup>2</sup>National University of Lesotho

\*Corresponding author. Email: [katherine.demuth@mq.edu.au](mailto:katherine.demuth@mq.edu.au)

(Received 21 October 2022; revised 01 November 2022; accepted 08 November 2022)

There has recently been an increased interest in studying the language development of non-western languages. This is not new - it began in 1960's and continued into the 1980's and 1990's. The current renewed interest is much welcomed, and will benefit from many of the experimental methods and theoretical insights developed over the past decades.

**Keywords:** longitudinal corpora; experiments; rural; language development

## Introduction

As Cristia et al. (this volume) rightly note, there is an enormous need for more research on lesser studied languages, many found in rural communities. A greater understanding of how languages in a diverse range are acquired is essential to our theoretical understanding of language, and critical to our understanding of what typical and atypical language development look like across populations. To this end, it is essential to develop the required knowledge base, tools and local personnel capable of assessing typical/atypical development across the world's languages. Such programs of basic and clinical research are beginning to be developed in South Africa, for example, on 11 lesser studied Bantu languages, all national languages of that country. Parental reports of Communicative Development Inventories (CDIs) are gradually being developed for each of these languages, in conjunction with training native-speaker speech pathologists, promising much future observational and experimental research on these languages in the coming years (cf. Pascoe & Jeggo, 2019). But this is only a start.

## The importance of collecting and annotating a longitudinal corpus

There is a serious need for converging evidence using multiple observational and experimental methods to fully understand how any language is acquired. The challenges of doing this for many lesser studied languages cannot be underestimated.

For less documented languages, where little is known about the language acquisition process, collecting a longitudinal corpus of children's developing language abilities can be an ideal place to start. This not only provides researchers with a rich sample of child spontaneous speech, but also daily interlocutor input from parents, grandparents, peers,

© The Author(s), 2023. Published by Cambridge University Press.

siblings, etc. It also provides much needed information about the range of linguistic structures that may be interesting to explore in future experiments. Such corpora can be probed in years to come, serving as a valuable future research and teaching tool. Such 'pilot' data can then greatly enhance future experimental study, and are ideal for local capacity building as well. Collection of such data by a research team that includes native speakers of the language ensures both ecological validity and provides excellent capacity building experience.

Morphological tagging and computerization of the corpora (e.g., Johnson, 2008) can then make it available to a much wider audience, and ensures future use for a wide range of new research questions, new collaborations, etc. Posting these transcriptions, complete with audio and video files, to various research archives (e.g., the CHILDES database, cf. MacWhinney, 2000), can ensure their future use.

Once a longitudinal corpus has been collected and annotated, the research team has a much better idea of 1) which structures are likely to be acquired when, 2) which structures pose challenges for acquisition (and possibly why), 3) at what ages different structures tend to be acquired, and 4) the nature of the input and how this may influence the acquisition process (see Demuth, Culbertson & Alter, 2006 for further discussion).

### Using a corpus to help inform experimental design

Building on the above, it is then much easier to design experiments to probe these issues more deeply. For example, using corpus data collected for another purpose, Demuth (1989, 1990) found that Sesotho-speaking children were spontaneously using syntactic passive constructions by the age of 2;8 years. This was surprising, since literature on English and other languages suggested these constructions, which involve syntactic movement, would be late acquired (e.g., around the age of 5). The Demuth Sesotho Corpus (Demuth, 1992) could then be examined more closely to design follow-up experiments to test the hypothesis that full syntactic passives, with a by-phrase, could be comprehended, elicited, and generalized to novel verbs. This was confirmed in Demuth, Moloji, and Machobane (2010). Importantly, the corpus provided abundant evidence not only for the types of lexical items/verbs that 3-year-olds would know, but also information about their lexical frequency. This played a critical role in designing the follow-up experiments, with all stimuli vetted by native speakers of the research team. Thus, a corpus can provide a wealth of lexical and other information for designing follow-up experiments that will more reliably tap what children know at various stages of develop. This kind of information is invaluable for designing experiments in languages where fewer language resources exist.

### Building a research team

When conducting experiments, it is necessary to have access to a large sample of children of certain ages, a quiet place to test them, and often, parental consent. Again, having local research capacity is invaluable, if not essential, for ensuring that all these conditions are met. But it is not only the data collection phase that requires a team; a team is also needed for appropriate transcription, computerization, checking/verification and analysis, of both corpus data, and experimental data. For the Demuth Sesotho Corpus, this involved the mothers and grandmothers of the target child who would have been present along with the researcher during the data collection recording, and an independent native

speaker of the language who listened to recordings and independently verified the transcriptions. For subsequent experimental research, this involved 1) extensive consultation and verification of experimental design (procedure, lexical items/sentences/conditions used, etc.) before starting data collection, 2) gathering research consent and child age information from preschool heads and/or parents, 3) ensuring that research premises had a 'quiet' room for testing, etc. In this latter case – especially for 1 room preschools – this involved testing in the school while other children were outdoors/at recess. In other cases it involved testing in the teacher's 'office'. Since there was no electricity, all also had to be conducted before cold weather, with all equipment batteries recharged the night before. (see Demuth, Machobane, Moloi & Odato, 2005 for further discussion).

### Control group

One of the issues raised by Cristia *et al.* (this volume) is that of a control group. We always include an adult control group – even for studies of English, where the predictions may not be clear in advance, and as a sanity check for the experimental design. We have also done this for Sesotho, as in the case of exploring children's knowledge of word order in double object constructions (e.g., *John gave Mary the ball* vs. *John gave the ball to Mary*). Note that English uses the preposition 'to' in the second sentence: this is not available in Sesotho. Furthermore, the order of objects in Sesotho is influenced by Animacy, with the most animate object first (e.g., *Mary-ball*, *Mary-dog*, *Mary~child*). When would children learn this animacy restriction? As anticipated, illiterate adult controls (drawn from gardener/cleaning staff at the university) were at chance in the equal animacy condition, with children becoming adult-like by around the age of 12. Thus, having an adult control is sometimes essential for exploring aspects of language development.

### Rural vs. urban communities

One of the issues not directly addressed by Cristia *et al.* (this volume) is the notion of socio-economic status (SES). It is often assumed that rural communities will have low SES, and that children growing up in such communities will then have less rich language. However, rural areas with no books, electricity, etc. can be extremely rich in the use of language from many speakers – in both quantity and quality, with continual talking, story-telling, language games, etc. This may contrast with peri-urban settings, where both parents are at work, and pre-schoolers are left in various care situations during the day. Cognitive and language abilities may thus be more variable in the latter context (cf. Demuth, Machobane & Moloi, 2003).

### Conclusions

In sum, investigating the acquisition of lesser studied languages is essential to our understanding of human language. Carrying out such research is much more ecologically valid and rewarding when carried out as part of a team. Collecting a longitudinal corpus can provide the opportunity for first steps in capacity building to create such a team. This then ensures the longevity of further research in a community, where follow-up experiments can explore many aspects of language and how it is acquired.

## References

- Demuth, K. (1989). Maturation and the acquisition of the Sesotho passive. *Language*, 56–80.
- Demuth, K. (1990). Subject, topic and Sesotho passive. *Journal of Child Language*, 17(1), 67–84.
- Demuth, K. (1992). Acquisition of Sesotho. In D. Slobin (Ed.), *The Cross-Linguistic Study of Language Acquisition* (Vol. 3, pp. 557–638). Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality epenthesis and coda licensing in the early acquisition of English. *Language and Speech*, 49, 137–174.
- Demuth, K., Machobane, M., & Moloji, F. (2003). Learning animacy hierarchy effects in Bantu double object applicative constructions. In *Linguistic Typology and Representation of African Languages* (pp. 23–33). Africa World Press.
- Demuth, K., Machobane, M., Moloji, F., & Odatu, C. (2005). Learning animacy hierarchy effects in Sesotho double object applicatives. *Language*, 81(2), 421–447.
- Demuth, K., Moloji, F., & Machobane, M. (2010). 3-Year-olds' comprehension, production, and generalization of Sesotho passives. *Cognition*, 115(2), 238–251.
- Johnson, M. (2008). Unsupervised word segmentation for Sesotho using Adaptor Grammars. *Proceedings of the Tenth Meeting of the ACL Special Interest Group on Computational Morphology and Phonology*, p. 20–27, doi: 10.3115/16266328.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. 3rd ed. Mahwah, NJ: Lawrence Erlbaum Associates
- Pascoe, M., & Jeggo, Z. M. (2019). Speech acquisition in monolingual children acquiring isiZulu in rural KwaZulu-Natal, South Africa. *Journal of Monolingual and Bilingual Speech*, <https://doi.org/10.1558/jmbs.11082>.

---

Cite this article: Demuth K., Moloji F., Matlosa L., & Johnson M. (2023). Uncovering the development of linguistic knowledge in lesser studied languages. *Journal of Child Language* 50, 518–521, <https://doi.org/10.1017/S0305000922000691>