Negotiating wordhood in elicitation and transcription

While the literature discusses wordhood issues in terms of mismatches between different criteria resulting from professional linguistic analysis (see e.g. Schiering et al. 2010), less attention has been paid to what happens in the elicitation and transcription processes, which can be considered the places where the most basic wordhood decisions take place (see e.g. Himmelmann 2018). In elicitation, researchers ask speakers for names for concepts, and speakers have to make a decision as to what they consider the appropriate chunk size for an answer. With some exceptions, the chosen chunk will be word-like, as we will briefly further discuss in the presentation. It is the word shapes that emerge in this process that provide the stepping stone for the further analysis of wordhood in a newly documented language.

Similarly, when asked to transcribe a recording, speakers have to make decisions as to how they chunk larger units. Chunks emerging from this will typically also be word-sized with some variation, where the variation is often indicative of analytical problems related to wordhood that will appear at later stages of the analysis.

In this talk we present some observations on these processes of negotiating wordhood in early stages of a language documentation. Data comes from Iha (Trans New Guinea, eastern Indonesia), where there is no long writing tradition for transcribers to fall back on, and speakers transcribing their language need to make decisions on the fly about where to draw word boundaries (which in fact they also do when writing Indonesian, the language they acquired basic literacy with). Our main example pertains to the inter- and intra-transcriber variation observed in transcriptions of the phrase *mi du-rik* '3 CLF.AN-two' plus (what we after a year of analysis consider) clitics (*=ke* 'FOC', *=ya* 'TOP', *=yɔ* 'PRX', *=ma* 'GEN'). Six instances are shown in (1).

(1)	Transcriber 1	Transcriber 2	Transcriber 3	Researcher analysis
a	mi durik yoyah	midurik oya	mi du rik ɔ ya	*mi du-rik=yɔ=ya*
b	mi duruk yoh	midurik	mi durik ɔ	*mi du-rik(=yɔ)*
c	mi duruk yoma	midurik oma	mi durik ɔ ma	*mi du-rik=yɔ=ma*
d	miduruk keh	midurike	mi du rik ge	*mi du-rik=ke*
e	mi durik keh	midurike	mi durik a	*mi du-rik=ya/mi du-rik=ke*
f	midurikyo	midurigo	mi durik ɔ	*mi du-rik=yɔ*

Our observations have implications for documentation practices. Based on the findings, we recommend that transcription be undertaken by many members of the community to avoid the influence of a single speaker's intuition. Transcription conducted together with a researcher should be recorded where possible. We argue that the speech community's collected metalinguistic knowledge of a language can be a rich source of data (revealed through transcription) when making analytic decisions. That is, the different 'words' in (1) are examples of primary data, and such data can and should be taken into account when defining the 'word' as a descriptive analytical unit. We show that, while it inevitably already is, this should be done consciously and then made explicit when presenting analyses. Increasing our knowledge of what happens in the transcription process and increasing transparency about how we use primary data from transcription to define words can improve both the quality and comparability of our analyses.

References

Himmelmann, N. P. 2018. Meeting the Transcription Challenge. In B. McDonnell, A. L. Berez-Kroeker, and G. Holton (eds.) *Reflections on Language Documentation 20 Years after Himmelmann 1998*. Honolulu: University of Hawai'i Press. Pp. 33–40.

Schiering, R., B. Bickel & K. A. Hildebrandt. 2010. The prosodic word is not universal, but emergent. *Journal of Linguistics* 46:657–709.