

## **How to count words (or their parts): Measures of speech quantity and complexity in multilingual, translanguaging populations using word versus morpheme boundaries**

**Rui Qi Choo, Victoria Y. H. Chua, Vinitha Selvarajan, Sheetal Sahana Vimalraj, Annabel K. H. Loh, Shaza binte Amran, Jin Yi Wong, Fei Ting Woon & Suzy J. Styles**

Nanyang Technological University, Singapore

In child language research, counts of words/morphemes are commonly used as a measure of speech quantity, and in computations of metrics for speech complexity (e.g., MLU). As most prior research has been conducted in monolingual samples (Kidd & García, 2022), decisions about the unit of linguistic analysis tend to be straightforward, as word-building rules keep the relationship between word-count and morpheme-count stable within a language. However, in multilingual contexts, parents and children may speak in any of their languages, or translanguaging, making use of their linguistic resources together. Such practices mean the relationship between word-count and morpheme-count can vary greatly within a conversation, or even a single sentence. As morphemes are meaning-bearing units of speech, we propose that the morpheme is an optimal unit of analysis in multilingual/translanguaging samples, and we demonstrate the use of morpheme counts in parent-child conversations featuring four typologically distinct languages: English, Mandarin Chinese, Malay and Tamil, ranging from analytic to agglutinative. Spontaneous child-directed speech was elicited using a wordless picture book ('Little Orangutan: What A Scary Storm', Styles, 2020) in a large sample of parent-child dyads in Singapore (N = 146, child age range: 8 to 40 months). The corpus was transcribed by research assistants using a protocol designed to capture translanguaging. Counts of speech quantity were conducted using a) word boundaries and b) morpheme boundaries using automated morphemisers, a decision protocol for manual checking, and manual morphemising protocols for each language. In this paper, we 1. present a principled justification for the use of morphemes as a unit speech quantity/complexity in multilingual contexts, 2. describe a cross-linguistic morphemising workflow, and 3. demonstrate how the choice of word versus morpheme boundaries influences metrics derived from the full corpus. Implications for child language science and automated 'word count' estimates derived from English are discussed.