

Exploring the Impact of Syllable Complexity on Canonical Proportion in Children: Insights from a Multilingual and Cross-cultural Study

Kai Jia Tey¹, Sarah Walker², Amanda Seidl³, Camila Scaff^{1,4}, Loann Peurey¹, Bridgette L. Kelleher⁵, Kasia Hitczenko⁶, William N. Havard⁷, Lisa R. Hamrick,⁸ Pauline Grosjean², Margaret Cychosz⁹, Heidi Colleran¹⁰, Marisa Casillas¹¹, Elika Bergelson¹² & Alejandrina Cristia¹

¹Laboratoire de Sciences Cognitives et de Psycholinguistique, ENS, EHESS, CNRS, PSL University, ²University of New South Wales, ³University of Delaware, ⁴University of Zurich, ⁵Purdue University, ⁶The George Washington University, ⁷University of Orléans, ⁸University of South Carolina, ⁹University of California, Los Angeles, ¹⁰BirthRites Lise Meitner Research Group, Max Planck Institute for Evolutionary Anthropology, ¹¹University of Chicago, ¹²Harvard University

One sign of early phonological development is the increasing prevalence of canonical syllables (consonant+vowel; Oller et al., 1998). A recently proposed metric is canonical proportion (CP): the proportion of a child's speech-like vocalisations containing clear consonant-vowel transitions (Cychosz et al., 2021). Initial analyses of 129 children suggested that CP relates to age non-linearly, continuing to develop well beyond the appearance of children's first words; and that it varies as a function of the ambient language structure (Hitczenko et al., 2023). Here we investigate CP further, considering potential effects of multilingualism (i.e., being exposed to 2+ languages). With the help of citizen scientists, we crowdsourced the annotation of 256,842 clips extracted from speech-like vocalisations by 371 children (2-77 months: 178 boys). The resulting dataset represents children from Bolivia (n=44), France (10), Mexico (10), Papua New Guinea (46), Solomon Islands (198), Vanuatu (40), and the USA (California 3, Indiana 10, New York 10). Children's CP appears to depend on age, mono-/multilingualism, and ambient language complexity (Figure1). First, a generalised linear model was fit to the monolingual data, declaring age in interaction with complexity (as in Hitczenko, languages were categorised as allowing only simple, moderately complex, or complex syllables following Maddieson, 2013). Second, a Levene Test confirmed significant difference in variance of CP between monolinguals and multilinguals ($F = 11.47$, $p < .005$). Our first analysis confirmed Hitczenko's observation that CP develops more slowly in languages that allow more complex syllables, possibly due to the challenge posed by learning the complex syllables. Previous research suggests that children often mirror the characteristics of their ambient language in their canonical babbling (Andruski et al., 2014). In a subsequent analysis, significant differences in variance between monolinguals and multilinguals were observed, and the reasons for these differences will be discussed.