THE DEVELOPMENT OF CANONICAL PROPORTION CONTINUES PAST TODDLERHOOD

Kasia Hitczenko[1], Elika Bergelson[2], Marisa Casillas[3], Heidi Colleran[4], Margaret Cychosz[5], Pauline Grosjean[6], Lisa R. Hamrick[7], Bridgette L. Kelleher[7], Camila Scaff[1,8], Amanda Seidl[7], Sarah Walker[6] & Alejandrina Cristia[1]

[1]Laboratoire de Sciences Cognitives et Psycholinguistique, ENS, EHESS, CNRS, PSL University

[2]Duke University, [3]University of Chicago, [4]Max Planck Institute for Evolutionary Anthropology,

[5]University of Maryland, [6]University of New South Wales, [7]Purdue University, [8]University of Zurich

kasia.hitczenko@ens.psl.eu

ABSTRACT

A key aspect of the development of speech production, the emergence of vocalizations that combine consonants and vowels, is captured by a measure called *canonical proportion* (CP). Yet this measure has mainly been studied among children under 12 months old learning English. We study CP in naturalistic speech in a cross-linguistic sample of 129 children aged 1 to 72 months. We show that children's CP continues to grow well after 12 months, and that CP development may vary cross-linguistically/culturally. This study has implications for how we conceptualize and monitor the development of speech production and showcases how coarse, semi-automated approaches can be used to study cross-cultural speech development from natural production data.

Keywords: language acquisition; citizen science; child vocalizations; speech production; long-form recordings

1. INTRODUCTION

Children's speech production undergoes rapid development in the first years of life. Young infants only produce “non-canonical” vocalizations, consisting of just a vowel or just a consonant (e.g., “uh" or “mmm"), but at around 6 months, they begin combining consonants and vowels together, with increasingly fast, adult-like, consonant-vowel and vowel-consonant transitions (e.g., “ba" or “up") [1]. This development has traditionally been captured with a measure called "canonical babbling ratio", or the proportion of syllables that contain such transitions, which reaches 15% by 10 months of age among English learners [2]. Less is known about how frequently such canonical transitions occur outside of babbling, when looking at older children, as well as for more diverse populations, leading to two open questions that we aim to address here.

First, beyond 12 months, researchers have typically turned to other aspects of phonological development, leading to an incomplete understanding of the developmental timeline. Recent work has suggested a new definition [3]: "canonical proportion" (CP), defined as the proportion of vocalization sections that have adjacent consonants and vowels (without regards to transition speed). Thus defined, CP can be calculated in all speech-like vocalizations, both from meaningful ones and meaningless babble. At 10 months, CP values thus defined are 15%, matching previous work based only on babbles. By 36 months, CPs are only 40% [3]. While there is no research establishing CP in adults' speech, 40% intuitively feels small, suggesting that this aspect of children's vocalization development may continue beyond 36 months.

By studying CP in older children, we can both better understand the full trajectory of children's phonological development and potentially establish a stable metric of phonological development: one that can be used throughout the first years of life and can be easily calculated from naturalistic recordings collected from highly-diverse communities [3, 4].

Second, the literature on group differences in CP development is small, probably because most researchers relied exclusively on (time-consuming) manual annotation, which offers detailed phonological information about early production, but limits sample sizes. Phonological properties of the language being acquired impact other aspects of children's vocalizations [5, 6], and the same could be true here. For example, children learning languages with low syllabic complexity (i.e., all syllables are consonant-vowel, e.g., "ma") might learn faster than children learning languages with high syllabic complexity (i.e., languages that allow many different syllabic types, e.g., words like "striped" as well as "pea"), due to fewer targets to learn, more exposure to them, and no need to attempt difficult syllable types, which could lead to non-adult-like performance. As for other group differences, results are mixed, with some research documenting individual and group differences [7] in part as a function of caregiver responsiveness, which likely differs across rural (small-scale, subsistence-level) and urban (industrialized or post-industrial) communities [8]; whereas others report similarities [3]. In sum, more work is needed to establish whether CP varies across children learning typologically diverse languages and/or growing up in diverse communities.

This leads us to our two main research questions: **(Q1)** How does CP develop from toddlerhood to 6 years of age? **(Q2)** Can we observe statistically significant differences in CP development that could be due to diverse languages (e.g., maximum syllable complexity) and/or communities (e.g., rural versus urban)? We address these questions using an innovative method: long-form audio recordings of children's everyday lives analyzed semi-automatically, by combining machine learning algorithms and crowd-sourced labels.

2. METHODS

All code is available at github.com/khitczenko/canprop-by-age_icphs2023.

2.1. Participant Recordings

We study a cross-linguistic sample of 129 children aged 1-72mos, learning English (N=20; 5-19mos) [9, 10, 11, 4], French (N=10; 11-13mos) [12], Quechua/Spanish (N=3; 22-25mos) [13], Tseltal (N=10; 2-36mos) [14, 15], Tsimane' (N=30; 6-71mos) [16, 17], Yélî Dnye (N=41; 1-72mos) [18, 19], or a subset of 12 Austronesian languages, spoken in the Solomon Islands (N=15; 4-48mos) [20]. Children from these 8 different communities wore non-intrusive, light-weight audio recorders in a specialized vest/shirt as they went about a typical day. These recordings provide a large and ecologically-valid sample of children's language development [21].

2.2. Labeling Vocalization Types

We first randomly sampled 100–300 child vocalizations from each long-form recording (5-16h long), using speech technology algorithms (LENA [22, 23, 24] or Voice Type Classifier [25]).[1] For older children, these vocalizations may be meaningful words/phrases, in line with our goal of representing CP in *all* spontaneous child production.

Citizen science is a growing approach used to manage large quantities of scientific data where volunteers assist with research tasks online. This approach allowed us to study many more children than would be feasible otherwise, without compromising label reliability: CP calculated on the basis of the crowd-sourced citizen science labels we use are highly correlated (r = 0.93) with those calculated using traditional in-lab approaches [4].

To obtain citizen science labels, the 100–300 vocalizations from each child were first split into short 400–500ms clips. This step was taken to protect the privacy of those who were recorded because the clips were shared on the Internet via a citizen science platform. Then, 3–5 citizen scientists labeled each clip for vocalization type: Canonical, Non-Canonical, Laughing, Crying, or Junk (i.e., the clip does not contain voices, or there is so much overlap that it is difficult to make out properties of the vocalization).[2] Each clip's category was determined via "majority rule," or excluded if no majority was reached (see [3, 4] for more details).

2.3. Calculating Canonical Proportion

CP was calculated, for each child, as the number of clips that were labeled as Canonical divided by the number of clips that were labeled as Canonical or Non-Canonical [3]: 0 means all of the child's speech-like clips were labeled as Non-Canonical; 1 that all were Canonical.

3. RESULTS

Seeking a response to **(Q1)**, we compared two models: (1) a mixed effects logistic regression predicting the effect of age on CP, controlling for gender and including corpus in the random effects structure and (2) the same model but instead with age-squared as a predictor.

$$(1)\quad CP \sim age_{child} + gender_{child} + (0 + age_{child}|corpus)$$

$$(2)\quad CP \sim age^2_{child} + age_{child} + gender_{child} + (0 + age^2_{child} + age_{child}|corpus)$$

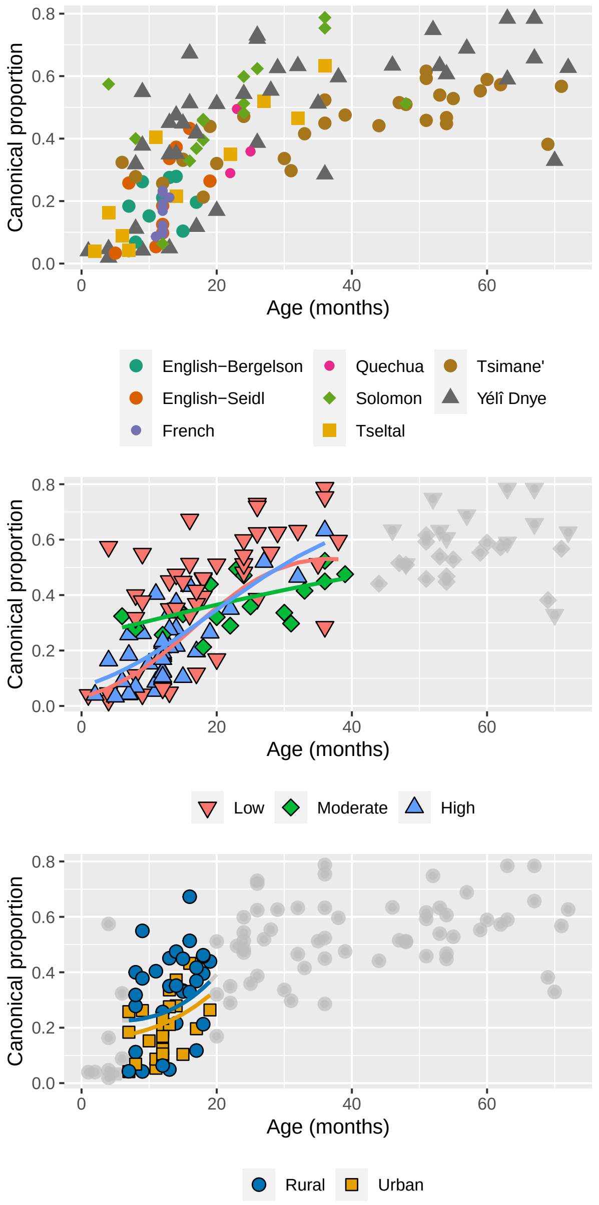

Figure 1: CP as a function of age, colored by corpus (top), syllabic complexity (middle), and rural/urban (bottom). CP continues to increase past toddlerhood, and may differ cross-linguistically/culturally. For subset analyses, we exclude (grayed-out) children in non-shared age ranges.

Fitting a quadratic effect of age allows us to model a non-linear relationship between canonical proportion and age. A log-likelihood chi-squared test comparing the two models (-1121.0 vs. -1029.7, $\chi^2(3) = 182.65$, $p < 0.001$), as well as comparison of AICs (2250.1 vs. 2073.4) and BICs (2261.5 vs. 2093.4), revealed that the quadratic model was a better fit, suggesting that while the measure continues to increase beyond toddlerhood, it may do so at a slowed rate with increasing child age.

Seeking a response to **(Q2)**, we first visually inspected CP by corpus (Figure 1). The corpora seem to differ in how quickly CP grows initially (e.g., it appears to grow faster for children learning Austronesian languages in the Solomon Island corpus, than in Tseltal, than in Tsimane'), as well as when we see a plateau (e.g., Yélî Dnye appears to have plateaued around 40 months of age at a high CP of 0.7-0.8, whereas Tsimane' has not reached this rate even by 72 months). However, assessing these differences statistically is challenging in view of marked differences in sample size (only 3 Quechua learners) and limited variability in age (all French learners are 11-13 months). Therefore, we next concentrated on two dimensions allowing us to pool across corpora, to study potential group differences as a function of language background and community, while controlling for age differences by subsetting to age ranges present in all levels of the factor of interest (1-40mos for syllabic complexity; 6-20mos rural vs urban), and including age in main and interaction terms.

To study the effect of syllabic complexity, we fit the following regression:

$$(3) \quad CP \sim age^2_{child} * syllabic\_complexity \\ \qquad + age_{child} * syllabic\_complexity$$

We found a significant main effect of syllabic complexity on CP ($\chi^2(2) = 96.19$, $p < 0.001$), and in interaction with both age ($\chi^2(2) = 169.73$, $p < 0.001$) and age^2 ($\chi^2(2) = 22.72$, $p < 0.001$), meaning that both the rates of learning and plateauing differed by whether children were learning a language with low (i.e., those that only allow consonant-vowel syllables; Solomon, Yélî Dnye), moderate (i.e., those that allow some consonant clusters and/or codas, but nonetheless have substantial restrictions on syllable types; Tsimane', Quechua), or high syllabic complexity (English, French, Tseltal). That being said, the fitted models are visually similar for low/high syllabic complexity languages.

Finally, we fit an analogous regression to (3) to test the relationship between community (urban vs. rural) and CP. We found a significant main effect of community on CP ($\chi^2(1) = 32.05$, $p < 0.001$), but no interaction effects with age ($\chi^2(1) = 0.87$, p=.35) or age^2 ($\chi^2(1) = 0.52$, p=.47), suggesting that children raised in rural environments may have higher CPs than those raised in urban environments, but show similar developmental patterns (i.e., slopes and plateaus).

4. DISCUSSION

The results of this study have broadened our understanding of canonical vocalization development and helped to map its trajectory in ecological settings, across a wider age range, and across diverse populations.

Interestingly, we found that CP continues to increase well into ages where children's vocalizations are thought to be driven by communication goals rather than vocal development. That is, even once children are producing words, which should dictate the consonant-vowel combinations they produce, we continue to see increases in CP. There are three, mutually compatible, possible explanations for this continued increase, which future work should disentangle. First, there could be a pragmatic explanation: as children mature, they begin inhibiting less advanced vocalizations that are not conversationally appropriate, i.e., meaningless non-canonical productions. However, it is unclear how this predicts development beyond the inhibition of babble, and specifically for the increase of canonical transitions. The second explanation is that vocal and/or phonological development is still ongoing. In particular, young children tend to simplify their syllables by e.g., dropping codas; the continued increase in CP could reflect a tendency to do this less and less over time, perhaps as children are able to automate articulatory processes (a speech motor development account), or learn the phonological importance of onsets/codas in their language (a phonological development account).

Finally, the continued increase could arise because children begin speaking faster or producing longer vocalizations, both of which could potentially lead to higher canonical rates in our methodological approach based on 500ms-long chunks. Future work should further validate this method across communities and qualitatively evaluate children's vocalizations to better understand what explains changes in CP across early childhood. Another open question for future work is what CPs are for adults, which would allow us to determine when children have reached adult-like status (there are non-canonical vocalizations, e.g., "yeah" and "hmm", that are appropriate at any age, so CP will likely not stabilize at 100%).

We also found preliminary evidence that CP may not develop universally across children, but may vary across linguistic and/or community settings. However, there were a couple of confounds and theoretical issues that make it premature to draw strong conclusions about what precisely drives these findings. For example, all of the children raised in urban environments were learning phonologically complex languages, and they were all younger than 20 months. Moreover, although our dataset is larger and more diverse than previous approaches, we still only included a couple of languages per syllabic complexity type. Finally, while we observed significant effects of syllabic complexity, the results seemed to be driven by languages with moderate syllabic complexity patterning differently from languages with low and high syllabic complexity (which patterned together), a surprising finding according to the hypothesis that syllabic complexity systematically relates to CP.

That said, while this set of results is ultimately inconclusive, they do raise the possibility that CP may not develop universally across children, which merits further dedicated study. To this end, we have submitted a registered report to undertake a confirmatory study on a larger, more diverse, and more balanced sample. This will also allow us to undertake a more nuanced analysis of the results to consider how other linguistic/phonological factors (e.g., word length, stress), which co-vary with syllabic complexity, may influence CP development.

Overall, these results provide some evidence that canonical vocalization development may be more protracted and potentially more variable across languages and/or settings than previously thought, which should be accounted for when using this measure, saliently in the context of potential applications. CP as studied here is promising because it can be calculated from naturalistic audio recordings collected from highly diverse communities, easily and without onerous and time-intensive transcriptions, and continues developing throughout early childhood. Future work could study its predictive and diagnostic value.

In all, our study shows the promise of coarse, semi-automated approaches towards studying early speech development. Combining long-form recordings, automated approaches, and citizen science approaches, we were able to study a large sample of children from around the world and broaden our understanding of language acquisition.

[1] Except for three of the corpora (Solomon, Yélî Dnye, Tseltal), where a subset of the corpora were labeled by trained, in-lab research assistants.

[2] Citizen scientists were provided with explanations and examples of each category, which they could access at any time and could contact the research team with any clarifications.

5. ACKNOWLEDGMENTS

This work was funded by the Agence Nationale de la Recherche, the J. S. McDonnell Foundation Understanding Human Cognition Scholar Award, and the European Research Council under the European Union's Horizon 2020 research and innovation programme. We thank Nadège Marin, Natalia Kuzminykh, and the Language Acquisition Across Cultures team for their assistance in preparing the data.

6. REFERENCES

[1] D. K. Oller, *The emergence of the speech capacity*. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.

[2] A. Nyman, S. Strömbergsson, and A. Lohmander, "Canonical babbling ratio–concurrent and predictive evaluation of the 0.15 criterion," *Journal of Communication Disorders*, vol. 94, p. 106164, 2021.

[3] M. Cychosz, A. Cristia, E. Bergelson, M. Casillas, G. Baudet, A. S. Warlaumont, C. Scaff, L. Yankowitz, and A. Seidl, "Vocal development in a large-scale crosslinguistic corpus," *Developmental Science*, vol. 24, no. 5, p. e13090, 2021.

[4] C. Semenzin, L. Hamrick, A. Seidl, B. L. Kelleher, and A. Cristia, "Describing vocalizations in young children: A big data approach through citizen science annotation," *Journal of Speech, Language, and Hearing Research*, vol. 64, no. 7, pp. 2401–2416, 2021.

[5] B. de Boysson-Bardies and M. M. Vihman, "Adaptation to language: Evidence from babbling and first words in four languages," *Language*, vol. 67, no. 2, pp. 297–319, 1991.

[6] M. M. Vihman and B. de Boysson-Bardies, "The nature and origins of ambient language influence on infant vocal production and early words," *Phonetica*, vol. 51, no. 1-3, pp. 159–169, 1994.

[7] A. S. Warlaumont, J. A. Richards, J. Gilkerson, and D. K. Oller, "A social feedback loop for speech development and its reduction in autism," *Psychological Science*, vol. 25, no. 7, pp. 1314–1324, 2014.

[8] A. Cristia, "A systematic review suggests marked differences in the prevalence of infant-directed vocalization across groups of populations," *Developmental Science*, p. e13265, 2022.

[9] E. Bergelson, "Bergelson SEEDLings Homebank Corpus." [Online]. Available: https://doi.org/10.21415/T5PK6D

[10] E. Bergelson and R. N. Aslin, "Nature and origins of the lexicon in 6-mo-olds," *Proceedings of the National Academy of Sciences*, vol. 114, no. 49, pp. 12 916–12 921, 2017.

[11] E. Bergelson, A. Amatuni, S. Dailey, S. Koorathota, and S. Tor, "Day by day, hour by hour: Naturalistic language input to infants," *Developmental Science*, vol. 22, no. 1, p. e12715, 2019.

[12] A. Cristia, "PhonSES: A pilot study to measure socioeconomic status association with infants' word and sound processing," 2021. [Online]. Available: https://gin.g-node.org/LAAC-LSCP/phonSES-public

[13] M. Cychosz, "Cychosz Homebank Corpus," 2018. [Online]. Available: doi:10.21415/YFYW-HE74

[14] M. Casillas, P. Brown, and S. C. Levinson, "Casillas Homebank Corpus," 2017. [Online]. Available: https://doi.org/doi:10.21415/T51X12

[15] ——, "Early language experience in a Tseltal Mayan village," *Child Development*, vol. 91, no. 5, pp. 1819–1835, 2020.

[16] C. Scaff, J. Stieglitz, and A. Cristia, "Tsimane' daylong recordings collected with LENA in 2017-2018," 2018. [Online]. Available: https://doi.org/DOI10.17605/OSF.IO/6NEZA

[17] ——, "Excerpts from daylong recordings of young children learning Tsimane' in Bolivia," 2019. [Online]. Available: https://doi.org/DOI10.17605/OSF.IO/5869Q

[18] A. Cristia and M. Casillas, "LENA recordings in Rossel Island," 2019.

[19] M. Casillas, P. Brown, and S. C. Levinson, "Early language experience in a Papuan community," *Journal of Child Language*, vol. 48, no. 4, pp. 792–814, 2021.

[20] A. Cassar, A. Cristia, P. Grosjean, and S. Walker, "Long-form recordings in the Solomon Islands," 2021.

[21] M. Lavechin, M. de Seyssel, L. Gautheron, E. Dupoux, and A. Cristia, "Reverse engineering language acquisition with child-centered long-form recordings," *Annual Review of Linguistics*, vol. 8, pp. 389–407, 2022.

[22] H. Ganek and A. Eriks-Brophy, "Language ENvironment Analysis (LENA) system investigation of day long recordings in children: A literature review," *Journal of Communication Disorders*, vol. 72, pp. 77–85, 2018.

[23] A. Cristia, F. Bulgarelli, and E. Bergelson, "Accuracy of the language environment analysis system segmentation and metrics: A systematic review," *Journal of Speech, Language, and Hearing Research*, vol. 63, no. 4, pp. 1093–1105, 2020.

[24] A. Cristia, M. Lavechin, C. Scaff, M. Soderstrom, C. Rowland, O. Räsänen, J. Bunce, and E. Bergelson, "A thorough evaluation of the Language ENvironment Analysis (LENA) system," *Behavior Research Methods*, vol. 53, no. 2, pp. 467–486, 2021.

[25] M. Lavechin, R. Bousbib, H. Bredin, E. Dupoux, and A. Cristia, "An open-source voice type classifier for child-centered daylong recordings," *arXiv preprint arXiv:2005.12656*, 2020.