

Phonological variation in child-directed speech is modulated by lexical frequency

Eon-Suk Ko¹ & Jongho Jun²

¹ Chosun University, ² Seoul National University

How does a child discover the underlying form of a morpheme under phonological variation? We address this question by investigating the production of Korean noun stem-final obstruents /t^h/, /c^h/, and /p^h/ before a vowel-initial suffix (Ko 1989; Kang 2003; Jun 2010) in child-directed speech (CDS) compared to adult-directed speech (ADS). The accommodation hypothesis of CDS (Ferguson, 1964; Kuhl et al., 1997; Fernald, 2000) might predict a higher ratio of canonical forms in CDS than their variants. Another factor influencing phonological variation, though largely overlooked in CDS research, is lexical frequency (Bybee, 2003; Gahl, 2008), which would predict a greater proportion of canonical forms in high frequency words in morpho-phonological phenomena. The present study conducts two experiments comparing the proportion of the canonical vs. variant forms ([s], [c^h], [t], [t^h], [c], [p]), and performs analyses to tease apart the effect of speech register and lexical frequency. Our results show that CDS contains a higher ratio of canonical forms, but that it is a fortuitous outcome driven by the adoption of high frequency word items in CDS rather than reflecting linguistic accommodation by caregivers.

In Experiment 1, 12 Korean mothers performed word-teaching tasks to their infants (M=0;10.3) and another adult using a picture book. The target words for CDS (*k'oc^h* 'flower', *pat^h* 'field', *sup^h* 'woods') and ADS (*koc^h* 'lynx', *sat^h* 'reed mat', *seop^h* 'brushwood') were first read and then explained. The target phonological alternations were extracted based on transcription. We fitted a mixed effect logistic regression model with the realization of the CANONICALITY as the dependent variable (canonical=1, variant=0). Fixed effects were REGISTER (CDS, ADS), CODA (/p^h, c^h, t^h/), VOWEL (/i, i/), and STYLE (reading, spontaneous). We also included by-item and by-subject random intercepts. The coefficient of REGISTER, our main interest, suggests that the rate of canonical form realization is higher in CDS than in ADS. The register effect, however, is confounded with lexical frequency since the target words in CDS are common words whereas those in ADS are rare, which calls for a follow-up experiment.

In Experiment 2, 12 college students produced 24 types of high- and low-frequency target nouns as determined by set thresholds in three tasks: reading, explaining semantic minimal pairs, and the map tasks (Anderson et al., 1991). A mixed effect logistic regression model was fitted to the phonological alternations with the same dependent and random variables as Exp 1, and the fixed effects including FREQUENCY (low, high), TASK (reading, minimal pair, map), SYLLABLE COUNT (monosyllabic, disyllabic), CODA (p^h, c^h, t^h) and VOWEL (i, i). The coefficient of FREQUENCY, our main interest, suggests that high-frequency words are more likely to be pronounced with the canonical form.

We conducted a further analysis comparing the proportion of canonical forms in the production of the two monosyllabic high frequency words, i.e. /k'ot^h/ 'flower' and /pat^h/ 'field', which are common target words across the CDS condition of Exp 1 and ADS in Exp 2. A mixed effect logistic regression was fitted with same dependent and random variables as Exp1 and Exp2, and the fixed effects REGISTER (cds, ads), TASK (reading, spontaneous), VOWEL (i, i, e), and CODA (p^h, c^h, t^h) (Table 3). The results show that the main effect of REGISTER is not significant. Further, its negative coefficient (-0.348) indicates a lower rate of canonical forms in CDS, contrary to what would be expected by the accommodation hypothesis.

Our results indicate that CDS does provide children with a simplified and enhanced phonological model, but that caregivers' use of a higher ratio of canonical forms is essentially a consequence of adopting high frequency word items in CDS. Reported lack of phonological enhancement in CDS in previous research (e.g. Buckler et al., 2018) could be attributed to the perceptual difficulty of disambiguating the canonical form from its variant (e.g.

gree[n]~gree[m] beans) even if a higher ratio of canonical forms had been provided. Our study thus highlights the importance of considering the usage factors and morpho-phonological phenomena in investigating the function and nature of CDS (Cychosz et al., 2021).

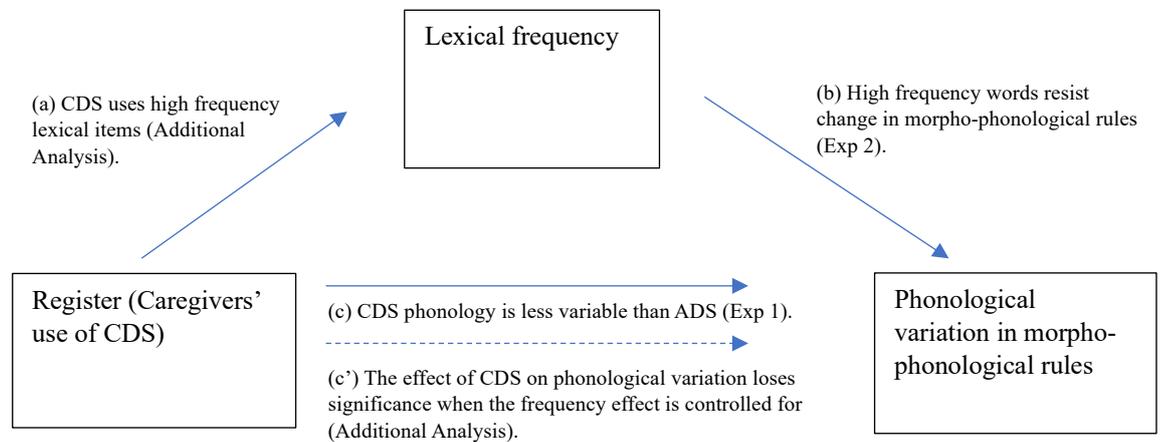


Figure 1. Mediation model. (a) represents the effect of register on lexical frequency (Additional Analysis, Lexical Frequency in CDS and ADS), (b) represents the effect of lexical frequency on phonological variation in morpho-phonological rules (Exp 2), (c) represents the direct effect of register on phonological variation (Exp 1), (c') represents the indirect effect of register on phonological variation, mediated by the frequency effect (Additional Analysis: Monosyllabic High Frequency Words in CDS and ADS, Common Lexical Items across CDS and ADS).

References

- Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S. & Weinert, R. (1991). The HCRC Map Task Corpus. *Language and Speech*, 34, pp. 351-366.
- Bybee, J. (2003). Mechanisms of change in grammaticization: The role of frequency. In B. Joseph, & R. Janda, (Eds.), *The Handbook of Historical Linguistics* (pp. 602–623). Oxford: Blackwell.
- Buckler, H., Goy, H. & E. K. Johnson (2018) What infant-directed speech tells us about the development of compensation for assimilation, *Journal of Phonetics* 66, 45-62.
- Cychosz, M., Edwards, J. R., Bernstein Ratner, N., Torrington Eaton, C., & Newman, R. S. (2021). Acoustic-Lexical Characteristics of Child-Directed Speech Between 7 and 24 Months and Their Impact on Toddlers' Phonological Processing. *Frontiers in Psychology*, 12, 712647.
- Ferguson, C. A. (1964) Baby talk in six languages, *American Anthropologist* 66 (6): 103–14.
- Fernald, A. (2000) Speech to infants as hyperspeech: knowledge-driven processes in early word recognition." *Phonetica* 57 (2–4): 242–54.
- Gahl, S. (2008) *Time and Thyme* are not homophones: The effect of lemma frequency on word durations in spontaneous speech." *Language* 84 (3): 474–96.
- Jun, J. (2010) Stem-Final Obstruent Variation in Korean, *Journal of East Asian Linguistics* 19 (2): 137–79.
- Kang, Y. (2003) Sound changes affecting noun-final coronal obstruents in Korean. In *Japanese/Korean Linguistics* 12, ed. W. McClure, 128-139. Stanford: CSLI
- Ko, K.-M. (1989) Explaining the noun-final change t>s in Korean. *Eoneohag* 11: 3-22.
- Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science (New York, N.Y.)*, 277(5326), 684–686.