From doggy to dog: Developmental shifts in children's use of register-specific words

Kennedy Casey & Marisa Casillas, University of Chicago

Background

Child-directed language (CDL) includes register-specific words (e.g., *doggy*, *night-night, tummy*) that are uncommon in adult-directed language (ADL).[1]

Past research explains why CDL variants are overrepresented in children's early vocabularies—features such as diminutivization, reduplication, and onomatopoeia are linked with early learnability.[2,3]

We do not yet know how or when children switch to using primarily ADL variants—*dog, goodnight, stomach*.

Standard vocabulary measures (e.g., CDI) typically collapse across lexical variants.

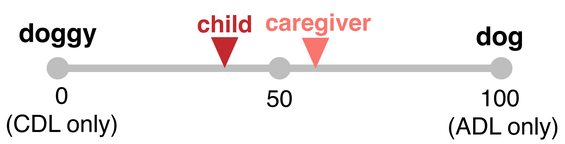
We investigate (1) when the CDL-to-ADL vocabulary shift occurs, and (2) what features of children's linguistic input may support this shift.

Caregiver Survey

Age range = 1;0 - 7;0
N = 120 English-hearing children
(20 per 1-year age bin; cross-sectional design)

For 15 CDL/ADL pairs:

Rate the relative frequency of use of CDL vs. ADL variants



Corpus Analysis

CHILDES[4]

Age range = 0;1 - 7;0
N = 980 children
North American English
Range of sampling strategies

Language Development Project[5]

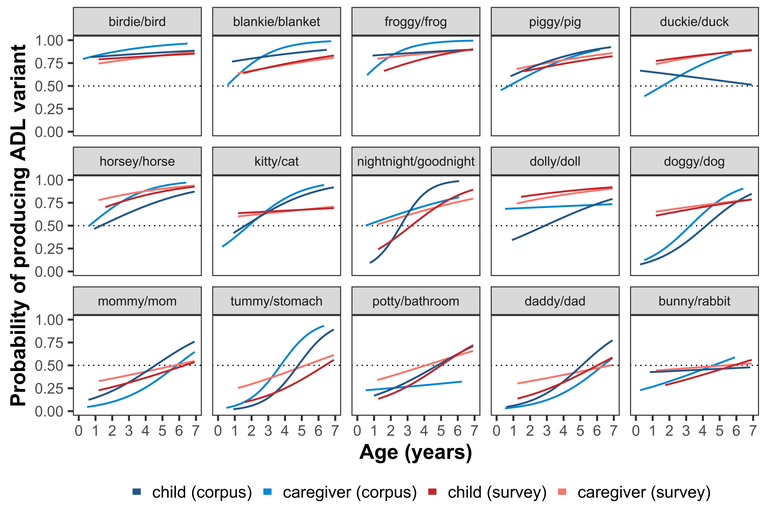
Age range = 1;4 - 4;10
N = 64 children
North American English
Longitudinal sampling
(90-minute home sessions every 4 months)

64,852 child utterances

65,079 caregiver utterances

4 linguistic features calculated at the utterance level:

- Lexical complexity (average AoA from adult ratings[6])
- Lexical rarity (relative frequency of occurrence in CHILDES)
- Utterance length (number of words)
- Syntactic complexity (number of verb phrases)



When do children and their caregivers switch to using primarily ADL variants?

Children and their caregivers increase ADL variant production over time.

Individual CDL/ADL pairs show varying shift trajectories.

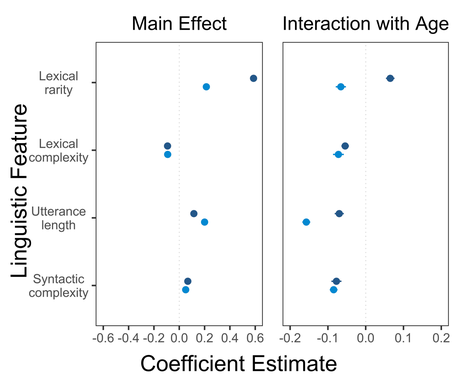
The average age of switch (i.e., when ADL variants are used >50% of the time) varied across speakers and methods:

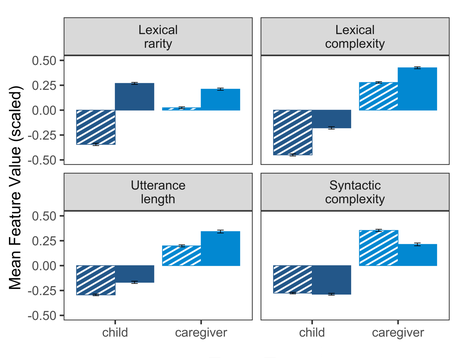
Speaker	Method	*β*	SE	Avg. switch age
child	corpus	0.47***	0.09	2.6 years
child	survey	0.37***	0.003	1.9 years
caregiver	corpus	0.57***	0.08	2.3 years
caregiver	survey	0.25***	0.003	1.2 years

***$p < 0.001$; glmer(variant (CDL = 0 vs. ADL = 1) ~ age (numeric; scaled) + (1 + age) | pair, family = "binomial")

Are CDL vs. ADL variants used in reliably different linguistic contexts?

All 4 linguistic features were significant independent predictors of variant type (ps < 0.001).





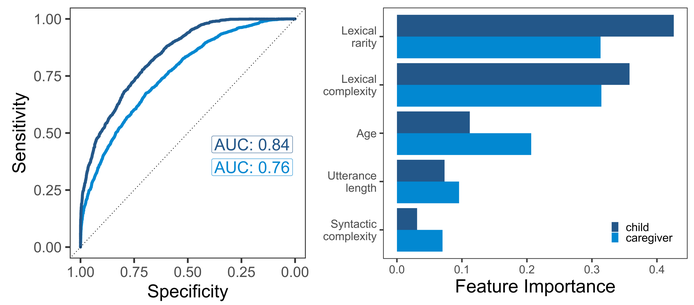
glmer(variant (CDL = 0 vs. ADL = 1) ~ feature (numeric; scaled) * age (numeric; scaled) + 1 | pair + 1 | speaker)

CDL/ADL Classifier

Extreme gradient boosting (XGBoost) tree-based algorithm[7,8]

Classify utterances with CDL vs. ADL variants using 4 linguistic features + age

Train on 90% of utterances
Test on remaining 10% of utterances



Discussion

Vocabulary development encompasses more than word accumulation.

Developmental shifts in children's CDL vs. ADL variant use reflect their emerging understanding that language should be adjusted to the current interactional context[9], including *who* is involved.

CDL-to-ADL vocabulary shifts may be supported by both changes in variant frequencies in children's input and reliable linguistic cues to register association.

References

[1] Ferguson, C. A. (1964). Baby talk in 6 languages. *American Anthropologist, 66*(6), 103–114.
[2] Laing, C., E. (2017). A perceptual advantage for onomatopoeia in early word learning: Evidence from eye-tracking. *Journal of Experimental Child Psychology, 161*, 32–45.
[3] Ota, M., Davies-Jenkins, N., & Skarabela, B. (2018). Why *choo-choo* is better than *train*: The role of register-specific words in early vocabulary growth. *Cognitive Science, 42*(6), 1974–1979.
[4] MacWhinney, B. (2000). *The CHILDES project: The database* (Vol. 2). Psychology Press.
[5] Goldin-Meadow, S., Levine, S. C., Hedges, L. V., Huttenlocher, J., Raudenbush, S. W., & Small, S. L. (2014). New evidence about language and cognitive development based on a longitudinal study: Hypotheses for intervention. *American Psychologist, 69*(6), 588–599.
[6] Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*, 978–990.
[7] Bang, J. Y., Kachergis, G., Weisleder, A., Marchman, V. A. (2022). An automated classifier for child-directed speech from LENA recordings. In Y. Gong, & Kpgogo (Eds.), *Proceedings of the 46th Annual Boston University Conference on Language Development* (pp. 48–61). Cascadilla Press.
[8] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM.
[9] Clark, E. V. (1997). Conceptual perspective and lexical choice in acquisition. *Cognition, 64*(1), 1–37.

Contact: kbcasey@uchicago.edu

OSF poster | GitHub data

Acknowledgements: Special thanks to Chatter Lab RAs Alexander Stern, Carla Suarez-Soto, Emily Chan, Lizzie Mickiewicz, and Sarah Sommer for their help with data collection and manual annotation.