

Wordful: Tracking Early Productive Vocabulary Growth with Smartphones

S. Meylan^{1,2}, M. Braginsky¹, B. deMayo³, A. Sanchez³, C. Schonberg⁴, M. Srinivasan⁵, H. Vlach⁴, G. Lupyán⁴, T. Griffiths⁶, M. Frank³

¹MIT ²Duke ³Stanford ⁴U.W. Madison ⁵U.C. Berkeley ⁶Princeton

44th Annual Meeting of the Boston University Conference on Language Development

Parental report, e.g., the MacArthur-Bates Communicative Development Inventory (CDI) [1], has long been used to gather information on the development of children's receptive and productive vocabulary. Beyond estimating the size of children's productive and receptive vocabulary, CDI data are increasingly used to study input-uptake relationships and item-specific variability [2,3,4]. While "short-form" instruments can provide robust estimates of children's scores [5,6], creating reusable datasets requires dense longitudinal sampling using either multiple long-form CDI administrations or diary methods.

We have developed a smartphone app, Wordful, to address this need. Caregivers report production or comprehension by swiping a card labeled with a word left (no) or right (yes), a common user interface. The logic underlying which words to query, which can be customized depending on the specific study, may reflect previous data from that child and population norms. A "word journaling" interface allows caregivers to report information on an open vocabulary. The app allows multiple caregivers to contribute data for the same child, and each caregiver to contribute data for multiple children. The app also provides visualizations of a child's early lexicon.

We conducted a pilot study with 118 U.S. parents recruited through Facebook, focusing on expressive language among children between 16 and 30 months. The study consisted of two online CDI administrations (Word and Sentences instrument), with 3-4 weeks of app usage in between. Caregivers received "push notifications" every other day to supply 20 yes/no card-swiping judgments; caregivers were also free to use the app at will. Tracked words included 680 words from the CDI, plus 193 supplementary items selected from CHILDES [7,8]. Items were selected for presentation each session by drawing from a Gaussian centered at the child's age from a list of words ordered by 50% producing (computed from Wordbank for CDI items, and predicted from a linear model for the supplementary items). Items where parents reported "no" were eligible for re-presentation after 3 days.

The 97 parents who completed the 2nd CDI provided 44,546 observations regarding production (18,020 yes / 26,526 no) from 1,122 sampling days with data. There was a strong by-item correspondence between app and CDI-based production data, with similar by-item correlations between the app and CDI and between the two CDI administrations (Figure 2). Further, app-collected data showed a high degree of consistency with word production trajectories from aggregate CDI data in Wordbank (Figure 3, left; mean per-item Pearson's r across words = .73) and provided item trajectories for 193 newly-tracked words (Figure 3, right). Estimated AoAs for new words were correlated with adult-provided AoA estimates ($r=.60$) [9], but the AoA estimates from Wordful are arguably more realistic.

This strong correspondence with CDI-collected data suggests that Wordful is a robust data-gathering tool for early vocabulary knowledge. Further, the modular architecture allows us to ask new research questions that require dense sampling (e.g., irregular verb forms). We hope Wordful will provide a valuable new tool for tracing vocabulary development and other phenomena using densely-sampled parental report. (500 words)

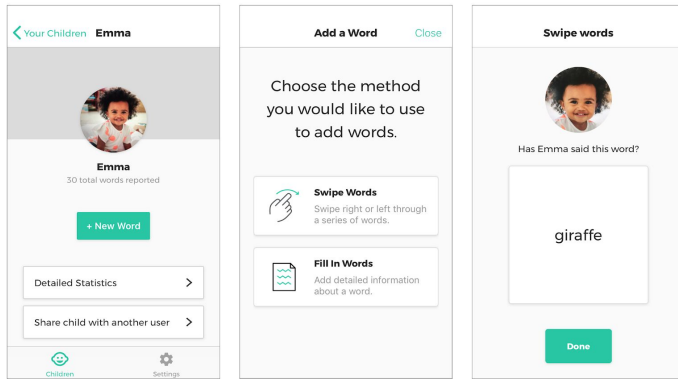


Figure 1. Example screenshots from the Wordful app: From left to right: Parent dashboard, available data gathering interfaces, and the “swipe words” interface.

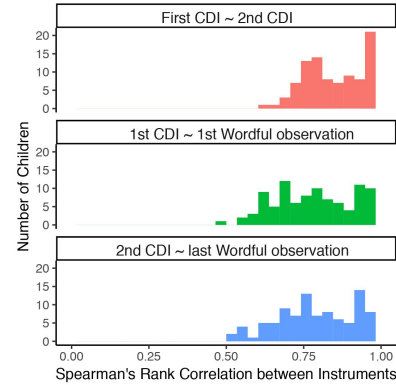


Figure 2. Correlation between produces/does-not-produce judgments by item between CDI administrations and Wordful.

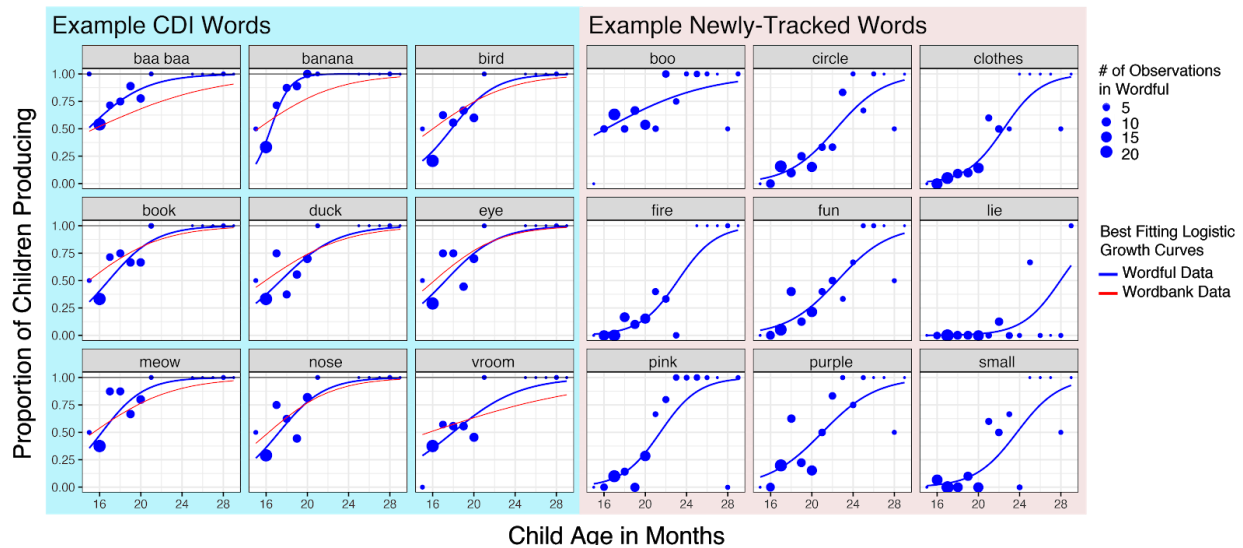


Figure 3. Proportion of children from Wordful producing words on the CDI (left) or newly-tracked words (right). In the first set, trajectories obtained from Wordful (blue logistic growth curves) can be compared to aggregated CDI data from Wordbank (red). Wordful AoA estimates ($\approx 50\%$ producing) correlate with Wordbank estimates ($\rho = .74$, $n = 680$); this exceeds the correlation between AoA estimates from naive adults [9] and Wordbank [2], ($\rho = .56$, $n = 563$).

1. Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., & Reznick, J. S. (2014). *MacArthur-Bates Communicative Development Inventories: user's guide and technical manual*. PB Brookes.
2. Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, 44(3), 677-694.
3. Braginsky, M., Yurovsky, D., Marchman, V. A., and Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *OpenMind: Discoveries in Cognitive Science*.
4. Beckage, N., & Colunga, E. (2013). Using the words toddlers know now to predict the words they will learn next. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 35, No. 35).
5. Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur communicative development inventories. *Applied Psycholinguistics*, 21(1), 95-116.
6. Mayor, J., & Mani, N. (2018). A short version of the MacArthur–Bates Communicative Development Inventories with high validity. *Behavior Research Methods*, 1-8.
7. MacWhinney, B. (2014). *The CHILDES project: Tools for analyzing talk, Volume II: The database*. Psychology Press.
8. Sanchez, A., Meylan, S. C., Braginsky, M., MacDonald, K. E., Yurovsky, D., & Frank, M. C. (2018). childes-db: a flexible and reproducible interface to the Child Language Data Exchange System. *Behavior Research Methods*, 1-14.
9. Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44(4), 978-990.