# How Data Drive Early Word Learning: A Cross-Linguistic Waiting Time Analysis

OPEN MIND

Discoveries in Cognitive Science

an open access journal

# How Data Drive Early Word Learning: A Cross-Linguistic Waiting Time Analysis

**Francis Mollica**[1] **and Steven T. Piantadosi**[1]

[1]Brain & Cognitive Sciences, University of Rochester

**Keywords:** word learning, rational construction, waiting time models

**ABSTRACT**

The extent to which word learning is delayed by maturation as opposed to accumulating data is a longstanding question in language acquisition. Further, the precise way in which data influence learning on a large scale is unknown—experimental results reveal that children can rapidly learn words from single instances as well as by aggregating ambiguous information across multiple situations. We analyze Wordbank, a large cross-linguistic dataset of word acquisition norms, using a statistical waiting time model to quantify the role of data in early language learning, building off Hidaka (2013). We find that the model both fits and accurately predicts the shape of children's growth curves. Further analyses of model parameters suggest a primarily data-driven account of early word learning. The parameters of the model directly characterize both the amount of data required and the rate at which informative data occurs. With high statistical certainty, words require on the order of $\sim 10$ learning instances, which occur on average once every two months. Our method is extremely simple, statistically principled, and broadly applicable to modeling data-driven learning effects in development.

**Citation:** Mollica, F., & Piantadosi, S. T. (2017). How data drive early word learning: A cross-linguistic waiting time analysis. *Open Mind: Discoveries in Cognitive Science*, *1*(2), 67–77. https://doi.org/10.1162/opmi_a_00006

**DOI:**
https://doi.org/10.1162/opmi_a_00006

**Supplemental Materials:**
www.mitpressjournals.org/doi/suppl/10.1162/opmi_a_00006

**Received:** 13 May 2016
**Accepted:** 3 January 2017

**Competing Interests:** The authors declare no competing interests.

**Corresponding Author:**
Francis Mollica
mollicaf@gmail.com



The MIT Press

The first year of life is an incredibly productive time for language learners. Babies discover which sounds are in their language (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992), how speech is segmented (Saffran, Aslin, & Newport, 1996), what common words refer to (Bergelson & Swingley, 2012), and, toward the end of the first year, how to produce their first word (Brown, 1973; Schneider, Daniel, & Frank, 2015). This growth is a complex endeavor that requires relying on abilities in many domains—social and pragmatic understanding, conceptual representation, joint attention, and acoustic and motor systems. However, little is known about how the development of nonlinguistic factors influences language growth. For instance, is the timing of language growth locked to factors like the maturation of cognitive and motor systems (e.g., memory and attention), or to the growth of children's conceptual repertoire? Or, alternatively, is early language learning primarily limited by the amount of data that children receive about language itself?

Evidence for a data-driven view of the timing of language learning comes from studies showing the importance of linguistic input for early learning (Hoff, 2003; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Shneidman, Arroyo, Levine, & Goldin-Meadow, 2013; Weisleder & Fernald, 2013). However, there are complications for the view that data are all that matters. Maturational constraints are often thought to play an important role in language learning (Borer & Wexler, 1987; Newport, 1990). Many words like function words (e.g., "the") and number words (e.g., "two") are learned surprisingly late for their frequency,

suggesting that the number of times a word is heard by a child is not a definitive predictor of learning. This fact has motivated hypothetical processes, including maturational constraints on function words or syntax (Borer & Wexler, 1987; Modyanova & Wexler, 2007) and conceptual or linguistic constraints in the case of number words (Carey, 2009).

At the heart of data-driven accounts is an ambiguity about how much data are required. Experimental studies of word learning have revealed children's ability to acquire word meanings from single instances (Carey & Bartlett, 1978; Heibeck & Markman, 1987; Markson & Bloom, 1997; Spiegel & Halberda, 2011), as well as from the aggregation of word usage across multiple contexts (Smith & Yu, 2008). It is not known which of these regimes governs the majority of lexical acquisition: Are most words learned by aggregation of tens, hundreds, or thousands of examples, or from a single informative instance?

Here, we develop a novel data analysis of word learning across 13 languages in order to address two questions about early word learning: When does it begin and how much data does it require? These questions turn out to be interrelated—they are coupled together by quantitative predictions that they make about the *distribution* of ages at which children learn a word. To illustrate this, consider a simplified picture of learning: Suppose that a word is learned by age 2. This could occur under many different situations. Three illustrative examples are: (a) the child could start accumulating data at birth, require about 24 cross-situational examples of the word, and receive them about once a month; (b) the child could start accumulating data at birth, require 4 examples, and receive them on average once every 6 months; (c) the child could start accumulating data at 12 months, require 12 cross-situational examples, and receive them once a month.

The central idea of our approach is that although (a), (b), and (c) predict the same mean age of learning, they critically predict different distributions of ages at which acquisition succeeds due to the statistics of waiting for data (see Figure 1). Empirical measurement of the distribution shape could in principle distinguish these hypotheses, informing us about how data influence the process of word learning. For instance, if the distribution supported (b), we might infer that there are few early constraints on learning since data accumulation begins at birth, and that learning required few examples. If the data supported (c), we might infer that cognitive or maturational constraints delayed the accumulation of data substantially, and that word learning required aggregating information across contexts.

The logic of our approach is to formalize the process of learning by accumulating data. Following Hidaka (2013), we assume that learners successfully acquire a word after $k$ effective learning instances (ELIs), or instances of the word that contribute to the learner's accumulating an amount of information about the word and we assume that ELIs arrive with an average frequency of $\lambda$ per month.[1] However, unlike previous work, we also infer the age $s$ at which data accumulation begins and implement our analyses in a Bayesian data analysis that is capable of inferring the likely ranges of parameter values from children's data. This Bayesian approach comes with several distinct advantages (Kruschke, 2010; Wagenmakers, Lee, Lodewyckx, & Iverson, 2008), including the ability to determine all three variables simultaneously, with our uncertainty in each correctly influenced by uncertainty in the others. Thus, our inferences

[1] Hidaka (2013) compares three different generative models for AoA distributions including one with a changing rate. In this analysis, we extend on his best-fitting model for the greatest amount of words, which has a fixed rate. As this might seem counterintuitive, we summarize the models he suggested and justify our choice of model in Appendix A of the Supplemental Materials.

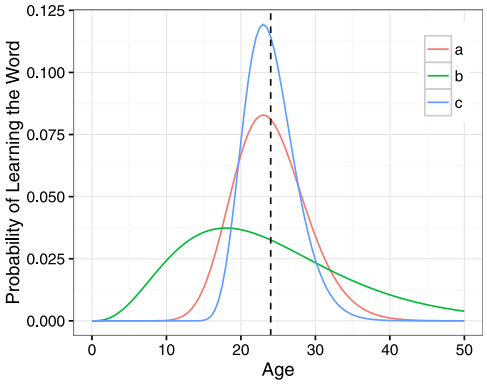


**Figure 1.** Example acquisition ages under 3 example assumptions: (a) children receive learning instances once a month from birth and require 24 total, (b) children require 4 examples and receive one every 6 months on average, (c) children require 12 instances, coming once every month, but only begin accumulating data at 12 months. Each predicts the same mean of 24 months (dotted line), but different shapes and variances in the timing of acquisition.

about the amount of data required to learn a word are statistically adjusted for our uncertainty over when learning that word began, and vice versa. The analysis also has the potential to reveal that the data are not informative about these variables, in which case we would find high uncertainty in the parameters given children's data. The advantage of our analysis compared to Hidaka's (2013) model comparisons is that we can confidently focus on interpreting the parameter estimates.

## PROBABILISTIC ASSUMPTIONS

Our model requires three primary assumptions: (i) age of acquisition (AoA) consists of two periods of time: a start time $s$ before learning a word begins and an accumulation time $t$, during which children are waiting for data; (ii) children learn a word after observing a number $k$ of ELIs of the word; and (iii) these ELIs occur stochastically, but at a fixed rate $\lambda$ (measured here in ELIs per month). For instance, $s = 0$, $k = 24$ and $\lambda = 1$ in example (a) above. Note that the model infers these parameters from learning curves, *not* from counting putative ELIs in child-directed data. It is likely that a constellation of factors are involved in determining whether any given instance contributes to learning (counts as an ELI). Similarly, start time $s$ could reflect several processes, including when children develop the ability to track and remember the data that they need to learn a word, or when their conceptual repertoire is ready to begin learning a word.

When data are observed stochastically with a rate $\lambda$ that is uniform in time, the number of ELIs actually received in a month will follow a Poisson distribution with rate $\lambda$. Under these assumptions, the distribution of times $t$ children must wait before receiving $k$ ELIs follows a Gamma distribution $\Gamma(k, \lambda)$ with density,

$$f(t; k, \lambda) = \frac{t^{k-1} e^{-t \cdot \lambda} \cdot \lambda^k}{\Gamma(k)}. \tag{1}$$

Thus, $f$ describes the distribution of time children must wait before observing enough data to learn a word. The curves in Figure 1 are Gamma distributions with the appropriate values

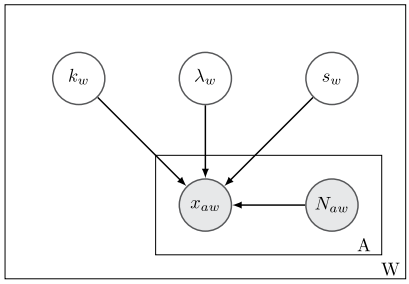


**Figure 2.** Graphical model notation for our model. Nodes denote variables of interest. Shaded nodes are observed variables. Plates denote groups of variables over age ($A$) and words ($W$). In the text, we provide equations for a single word and omit the subscript $w$.

for $k$ and $\lambda$. Note that in a Gamma, the mean scales linearly in the variance, meaning that if acquisition is driven by accumulating data, children's variance in learning times should scale with their mean learning time. Gamma-shaped learning time distributions should be taken as a hallmark of data-driven, constructivist accounts of learning (Xu, 2007; Xu & Kushnir, 2012) that applies to any theory of development in which accumulating data is the primary force advancing learners' knowledge.

## THE DATA ANALYSIS MODEL

Our data analysis model uses Bayesian techniques to recover $k$, $\lambda$, and $s$ from empirically measured learning curves. To do this, we require one data-analysis assumption that the population of children studied is relatively homogeneous, meaning that we may extend a word's single $s$, $k$, and $\lambda$ across children.[2] In this case, the proportion of children who know a word at accumulation time $T$ will approximate the cumulative distribution function of (1) at time $T$,

$$F(T;k,\lambda) = \int_0^T f(t;k,\lambda)\, dt. \tag{2}$$

Figure 2 shows a graphical model of the relationships between these variables and the observed data. At each age $a$, $N_a$ children were measured and $x_a$ of them reported having learned the word to either production or comprehension.[3] We model the number of children producing/comprehending the word $x_a$ as being drawn from a binomial distribution with $N_a$ trials and a probability of success equal to the proportion of children who know the word given by (2) at time $t = a - s$:

$$x_a \sim \text{Binom}(F(a-s,k,\lambda), N_a) \tag{3}$$

We assume uniform priors on these variables: $k \sim$ Uniform(0, 10,000) ELIs, $\lambda \sim$ Uniform(0, 10,000) ELI(s)/month and $s \sim$ Uniform(0, 1,000) months. Bayesian inference in this generative model allows us to take the empirical acquisition curves and determine posterior distributions for $k$, $\lambda$, and $s$ for each word in each language.

[2] Our conclusions hold even if we relax this assumption (see Appendix B of the Supplemental Materials).
[3] We fit the comprehension and production data separately.

## RESULTS

### *The Cumulative Gamma Matches Observed Word Learning Curves*

Figure 3 shows a general visualization of the model fit across a variety of English words. Despite its simplicity, the model closely accounts for the empirical learning trajectories across word types for both comprehension and production. Quantitatively, correlations between predicted values and the behavioral data are near 1.0 for each language (see Supplemental Figure S1 in our Supplemental Materials [Mollica & Piantadosi, 2017]) meaning that the model is able to capture the overall shape of acquisition across languages. More importantly, the model is able to more successfully *predict* learning than more standard alternatives: a probit (McMurray, 2007) and a logistic model. To test this, we divided the learning curve for each word into two halves, where we fit $k$, $\lambda$, and $s$ for each word on the first half and then computed the correlation between model and human data across words and ages on the full curves. The Gamma distribution fit quantitatively outperforms either the probit or the logit across most languages (see Figure 4).

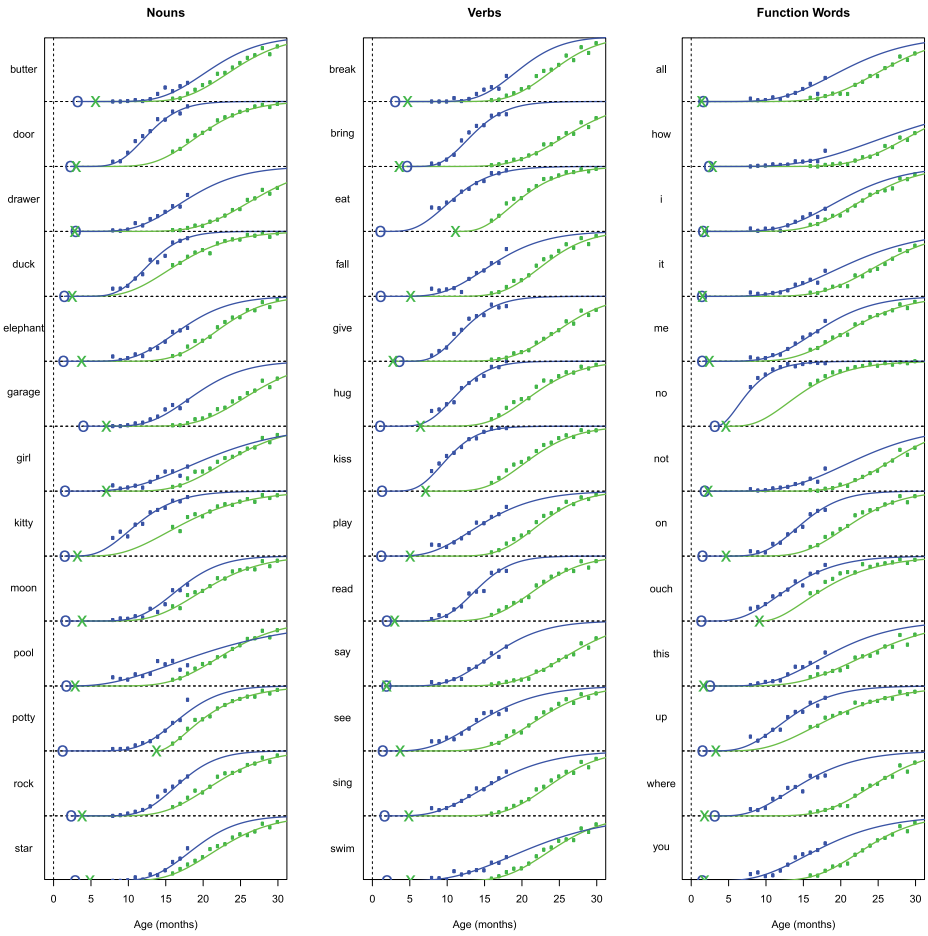


**Figure 3.** Points shows the proportion of English-speaking children ($y$-axis) who know a word at each age ($x$-axis) as measured by comprehension (blue) and production (green). Lines show the posterior mean parameters in the model (2), and X and O show the posterior mean start time of data accumulation for each word. This generally shows good model fits, early start times for comprehension, and somewhat later times for production.

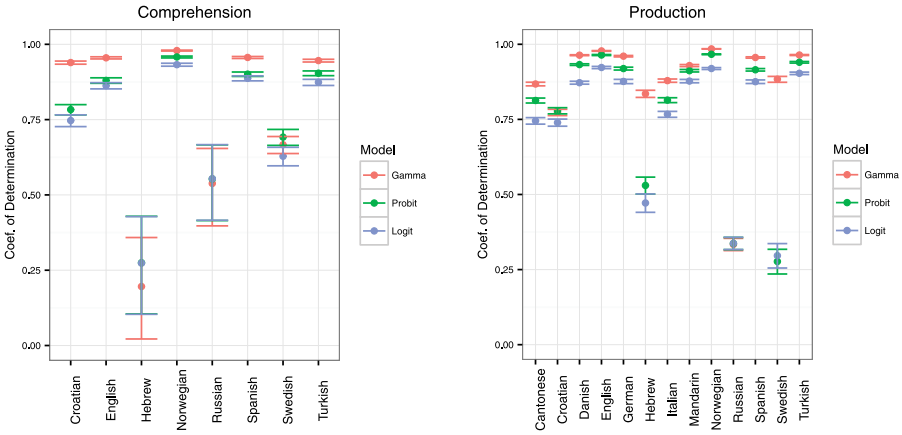


**Figure 4.** Model comparison of the Logit, Probit, and Gamma models when trained on the first half of comprehension and production learning curves and tested on the full trajectory. Across words and languages, the correlations between observed data and model predictions for the full curve are close to 1 with the Gamma model showing the best fit.

### *On the Order of 10 ELIs Are Needed to Learn a Word*

The order of magnitude of the estimated parameters are informative about the underlying mechanisms of learning, as they characterize when learning starts ($s$), how many ELIs are needed ($k$), and how frequently they occur ($\lambda$). Figure 5 shows the mean values of $k$, $\lambda$, and $s$ for each language. The box plots for English further broken down based on MacArthur-Bates Communicative Development Inventory (MCDI) semantic category are similar (see Supplemental Figure S2 in our Supplemental Materials [Mollica & Piantadosi, 2017]).

Figure 5a and 5d show that, across languages, the order of magnitude of $k$ is around 10 for production, with slightly lower values for comprehension. It is important to focus on the order

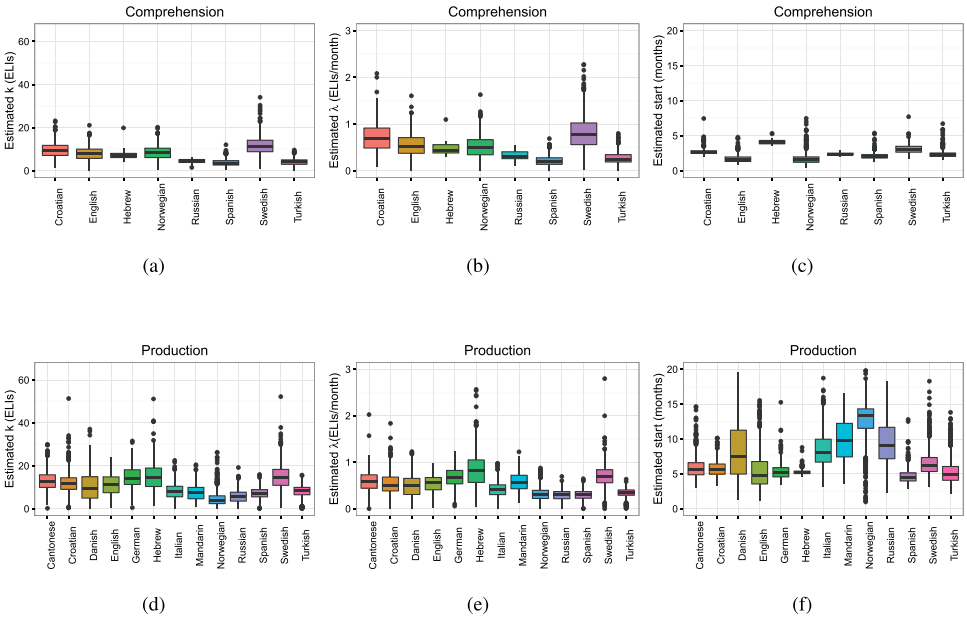


**Figure 5.** Box plots of the distribution of $k$, $\lambda$, and $s$ across words in each language.

of magnitude, not the exact numerical values, because the order of magnitude of our parameter estimates are robust to noise (see Appendix B of the Supplemental Materials). The important issues in language development can still be distinguished based on order of magnitude. We primarily interpret Figure 5 as showing that languages agree in order of magnitude of their estimates.[4] Thus, children do not require hundreds or thousands of instances of a word to learn, even for words that may be very frequent, nor do they learn from a single instance. Instead, learning is likely focused around ten critically informative learning instances. These findings demonstrate the importance of cross-situational statistics over single examples and is consistent with the finding that children do not retain fast-mapped meanings (Horst & Samuelson, 2008).

#### *ELIs of a Word Occur Roughly Every Two Months*

The variable $\lambda$ characterizes the estimated rate at which ELIs of a word occur. Figures 5b and 5e show that ELIs of a word occur once every two months ($\lambda \approx 0.5$), indicating that ELIs are relatively infrequent for an individual word. However, because children learn many words simultaneously, ELIs of any word may in fact be quite frequent. For instance, if children track statistics on 1,000 early words, and observe an ELI for *each* word on average once every two months, they will receive around 17 ELIs per day.

#### *Data Accumulation Starts Around Two Months*

The start times in Figures 5c and 5f show that learning begins early: approximately by two months in the case of comprehension measures. The starting age is somewhat later when curves are fit to production measures, possibly because production may require motor and speech systems to be working before production can progress. This may indicate that although maturational factors play little role in learning as measured by comprehension, production depends on the development of other cognitive or motor systems.

#### *Early Word Learning Is Primarily Data-Driven*

The model assumes that AoA is the sum of two time periods: start time $s$ and accumulation time $t$. There are two measures we derive from these parameters to quantify the extent to which early word learning is data-driven: the percent of total AoA time spent accumulating data, and the percent of variance in AoA explained by variance in accumulation times. If early word learning is primarily constrained by maturation, the majority of acquisition time should not be spent accumulating data and the majority of the variance in acquisition times should be explained by the variance in start times $s$. On the other hand, a data-driven account of early word learning would expect the majority of acquisition time to be spent accumulating data and the majority of the variance in acquisition times to be explained by variance in accumulation times $t$. Figure 6 shows the proportion of total acquisition time and the variance in acquisition times that is due to $t$ (accumulating data) rather than $s$ (start times). We find that generally the majority of acquisition time is spent accumulating data and the variance in accumulation times explains the majority of the variance in acquisition times. Taken together, this indicates that data-driven factors are the primary drivers of early word learning.

[4] We suspect that the greater uncertainty around estimates for Hebrew and Swedish is due to data sparsity (see Supplemental Figure S4 in our Supplemental Materials [Mollica & Piantadosi, 2017]).

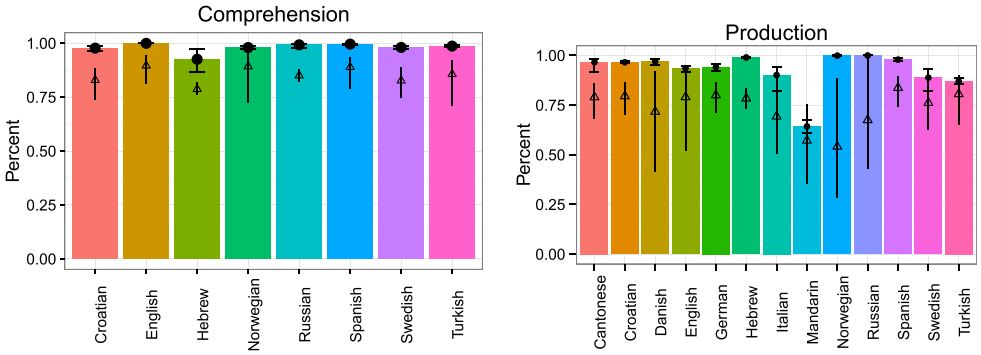


**Figure 6.** The bar plot shows percent of the variance in age of acquisition (AoA) times explained by accumulation time (suggesting data-driven learning). The triangular points shows the percent of AoA time spent accumulating data. Error bars and point ranges represent bootstrapped 95% confidence intervals. Outliers ($<$ 2.5% of the data) were removed for this analysis (see Methods section).

### *Learning Instances Are Weakly Correlated With Log Frequency*

Under a simple view that most usages of a word are informative about its meaning, our estimates of $k$ and $\lambda$ should be surprising; word frequencies vary over several orders of magnitude (Zipf, 1949), yet the inferred $k$ and $\lambda$ values do not. This means that ELIs cannot be very strongly correlated with frequency. Most of the time a frequent word is used, it is not an ELI. One possibility is that a single ELI for a word like *tiger* might be an entire visit to the zoo.

To investigate the relationship further, we computed the correlation between the estimated $k$, $\lambda$, and $s$ values for each word in English and the log frequency as measured in CHILDES (MacWhinney, 2000). For comprehension, there is only a small correlation between the estimated $k$ parameter and frequency ($k : r = -.14, p = .01$). For production, there is a modest correlation ($k : r = .19, p < .001$; $\lambda : r = .32, p < .001$; $s : r = -.22, p < .001$) as observed by Hidaka (2013). But what is notable is the *weakness* of the correlation (see

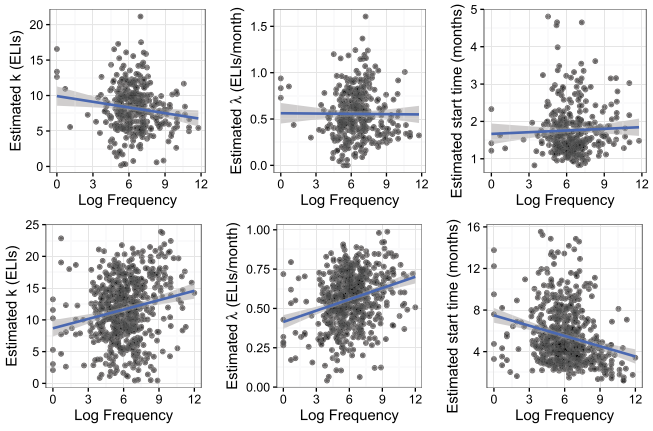


**Figure 7.** Correlations between CHILDES frequency for words in English and estimated parameter values. Top row: For comprehension, there is a small correlation between frequency and $k$ and no correlation between frequency and $\lambda$ and frequency and $s$. Bottom row: For production, the correlations between frequency and $k$, frequency and $\lambda$, and frequency and $s$ are very weak and only significant when frequency is log transformed.

Figure 7)—it is not as though doubling the quantity of input will double the number of ELIs. This finding is compatible with findings of frequency effects in word learning (Ambridge, Kidd, Rowland, & Theakston, 2015; Hoff, 2003; Huttenlocher et al., 1991; Shneidman et al., 2013; Weisleder & Fernald, 2013), but suggests that frequency will be less important than the frequency of ELIs (see also Hoff, 2003).

## DISCUSSION

We view the Gamma model not as a mechanistic learning account, but instead as a scientific *tool* for understanding the basic forces in early language acquisition. Unlike characterizations in terms of mean acquisition ages, the parameters $s$, $k$, and $\lambda$ are *psychologically meaningful* in terms of a causal process that likely supports part of word learning, data accumulation (Hidaka, 2013). Our analysis of empirical learning curves strongly suggests that data accumulation begins very early, that production may be delayed due to maturational factors, and that typical words take on the order of $\sim$ 10 ELIs to learn, not hundreds of occurrences and not a single occurrence or two. The model also suggests that the *informative* data points for word learning occur relatively infrequently, about once every two months, and that these occurrences are not strongly related to a word's overall frequency. Moreover, the mechanisms of data accumulation not only provide the best quantitative fit to learning curves, they explain *nearly all* of the variance in when children learn a word.

This analysis has capitalized on the existence of large corpora of acquisition trajectories across children. In particular, the key variables of interest, data amounts, data rates, and the time at which data are first considered, are discovered entirely from children's acquisition trajectory—not from recordings of children's input. While it may seem tempting to address these questions of acquisition with an intensive home recording study (Roy et al., 2006) or an evaluation of child-parent interactions (MacWhinney, 2000), these approaches come with the challenge of delineating which instances of a word concretely contributed to learning. For example, a word use might only aid acquisition if the child is attentive and receptive, and the referent is clear, which might not be observable in those datasets. Given that we have found that overall frequency is a weak predictor of the rate of ELIs, the detailed measurement of just parental productions will not fully clarify the relevant data sources for learning. Instead, our work takes a different tack, looking to find evidence of data-driven effects writ large in the *distribution* of learning times for words.

This work leaves open a central question: what makes a usage of a word an ELI? The weak correlation between the parameters and word frequency suggests that ELIs are rare—and perhaps even intentional. It is likely that children actively decide what stimuli they engage and deeply process (Kidd, Piantadosi, & Aslin, 2012, 2014), which could place an internal yoke on the rate of ELIs. Extrinsic factors probably also play a role though, as seen by the correlations with frequency. Analogously, these analyses raise the question of what determines differences in $k$ and $\lambda$ across words and languages. Future research should attempt to characterize the impact of external factors, such as semantic content (Jones, Johns, & Recchia, 2012) and phonotactic probability (Storkel, 2001), on $k$ and $\lambda$. Our framework provides the initial step at connecting such factors to the data accumulation process that implicitly supports all existing models of word learning.

It is also important to note the limitations of the MCDI data and our model. First, we restrict all of our conclusions to the early learned words covered by the MCDI. It will be important to extend this model beyond the age range of the existing MCDI. Children are flexible learners and it is probable that an older child adopts a variety of strategies, which may influence

the data-driven process. For example, older children might be able to bootstrap from their existing vocabulary/syntactic constructions or their intuitive theories of the world. Additionally, the lack of variability in the MCDI words constrains the empirical testing of many hypothesized constraints on vocabulary acquisition (e.g., Markman, 1990). Applied to the appropriate data, our approach is a suitable tool to evaluate these constraints at the computational level. Further, we chose to encode maturation as a constant offset from birth to address our main questions. This is an appropriate operationalization but a coarse distinction, and future research should address this.

## METHODS

We fit $k$, $\lambda$, and $s$ within individual words and languages on data retrieved on June 16, 2015, from Wordbank (Frank, Braginsky, Yurovsky, & Marchman, 2016), a repository for MCDI instruments (Fenson et al., 2007). This yielded cross-sectional data from 13 languages (see Supplemental Figure S3 in our Supplemental Materials [Mollica & Piantadosi, 2017] for further description). For each word in each language, $k$, $\lambda$, and $s$ were fit using JAGS (Plummer, 2003) and corresponding R packages, `rjags` and `runjags`. For every word, four chains were run for a total of 1.25 million steps with a thin of $1,000$ steps between each saved step. The chains converged ($\widehat{R} < 1.2$) for all $2,397$ words in the comprehension and $9,420$ words in the production measure. For our data vs. maturation analyses, we removed outliers ($< 2.5\%$ of the data) that were all syntactic constructions as opposed to lexical items. The forward predicting model was trained on the first half of the data using the same method. In these runs, 88 words failed to converge for comprehension and 78 words failed to converge for production and were excluded from further analysis. Code and parameter estimates are available from the first author and our lab's webpage.

## ACKNOWLEDGMENTS

The authors thank Dick Aslin, Elika Bergelson, Celeste Kidd, and anonymous reviewers for comments on early drafts of this article.

## AUTHOR CONTRIBUTIONS

FM and STP designed the model, FM implemented the model, and FM and STP analyzed the data and wrote the article.

## REFERENCES

Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*(02), 239–273.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253–3258.

Borer, H., & Wexler, K. (1987). *The maturation of syntax*. Dordrecht, Netherlands: Springer.

Brown, R. (1973). *A first language: The early stages*. Oxford, England: Harvard University Press.

Carey, S. (2009). *The origin of concepts*. Oxford, England: Oxford University Press.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, *15*, 17–29.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303–306.

Fenson, L., Bates, E., Dale, P. S., Marchman, V. A., Reznick, J. S., & Thal, D. J. (2007). *MacArthur-Bates Communicative Development Inventories*. Baltimore, MD: Brookes.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2016). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*. Advance online publication. doi:10.1017/S0305000916000209

Heibeck, T. H., & Markman, E. M. (1987). Word learning in children: An examination of fast mapping. *Child Development*, *58*(4), 1021–1034.

Hidaka, S. (2013). A computational model associating learning process, word attributes, and age of acquisition. *PLOS ONE*, *8*(11), e76242.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1368–1378.

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*(2), 128–157.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*(2), 236.

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *66*(2), 115.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLOS ONE*, *7*(5), e36399.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The goldilocks effect in infant auditory attention. *Child Development*, *85*(5), 1795–1804.

Kruschke, J. K. (2010). Bayesian data analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(5), 658–676.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Hillsdale, NJ: Lawrence Erlbaum.

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*(1), 57–77.

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature*, *385*(6619), 813–815.

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, *317*(5838), 631.

Modyanova, N., & Wexler, K. (2007). Semantic and pragmatic language development: Children know "that" better. In *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition–North America (GALANA 2)* (pp. 297–308). Somerville, MA: Cascadilla Proceedings Project.

Mollica, F., & Piantadosi, S. T. (2017). Supplemental material for "How data drive early word learning: A cross-linguistic waiting time analysis." *Open Mind: Discoveries in Cognitive Science*, *1*(2), 67–77. https://doi.org/10.1162/opmi_a_00006

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*(1), 11–28.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing* (Vol. 124, p. 125). Retrieved from https://sourceforge.net/projects/mcmc-jags/

Roy, D., Patel, R., DeCamp, P., Kubat, R., Fleischman, M., Roy, B., . . . Gorniak, P. (2006). The Human Speechome Project. In *Symbol grounding and beyond* (pp. 192–196). Berlin, Heidelberg: Springer.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Schneider, R. M., Daniel, Y., & Frank, M. C. (2015). Large-scale investigations of variability in children's first words. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 2110–2115). Austin, TX: Cognitive Science Society.

Shneidman, L. A., Arroyo, M. E., Levine, S. C., & Goldin-Meadow, S. (2013). What counts as effective input for word learning? *Journal of Child Language*, *40*(03), 672–686.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

Spiegel, C., & Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *Journal of Experimental Child Psychology*, *109*(1), 132–140.

Storkel, H. L. (2001). Learning new wordsphonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, *44*(6), 1321–1337.

Wagenmakers, E.-J., Lee, M., Lodewyckx, T., & Iverson, G. J. (2008). Bayesian versus frequentist inference. In *Bayesian evaluation of informative hypotheses* (pp. 181–207). New York, NY: Springer.

Weisleder, A., & Fernald, A. (2013). Talking to children matters early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152.

Xu, F. (2007). Rational statistical inference and cognitive development. *The Innate Mind: Foundations and the Future*, *3*, 199–215.

Xu, F., & Kushnir, T. (2012). *Rational constructivism in cognitive development* (Vol. 43). Waltham, MA: Academic Press.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. New York, NY: Addison-Wesley.