

Vocabulary Spurt: Are Infants full of Zipf?

Julien Mayor^{a,b} and Kim Plunkett^b

^a Basque Center on Cognition, Brain and Language, San Sebastian, Spain

^b Department of Experimental Psychology, University of Oxford, United Kingdom
(j.mayor@bcbl.eu, kim.plunkett@psy.ox.ac.uk)

Abstract

Infants do not learn words at a constant rate. During the second year of life, a dramatic increase in the speed of word learning is observed. Different mechanisms explaining this vocabulary spurt have been proposed, either through endogenous factors such as learning capacity or exogenous factors, such as frequency of word usage. We demonstrate that occurrence statistics alone is not sufficient to explain the acceleration in vocabulary growth, discuss other potential exogenous contributions such as phonological complexity and suggest that a change in word learning capacities is necessary. A model implementing an increased ease of learning is introduced and illustrates this endogenous approach by replicating the non-linear vocabulary growth characteristics of language acquisition.

Keywords: vocabulary spurt; mathematical modelling; word learning; learning mechanisms; Zipf's law; endogenous vs. exogenous factors

Introduction

Around their first birthday infants utter their first word and by their second birthday they learn on average one new word every waking hour. Between 18 and 24 months of age, an abrupt change in the speed of word acquisition is observed, called the vocabulary spurt or naming explosion (Bloom, 1973)¹. Two types of theories have been offered to explain the vocabulary spurt. One suggests that the vocabulary spurt corresponds to representational and/or maturational changes in the infant's brain. For example, researchers have suggested that infants start acquiring words at a faster pace when they understand that words refer to things and/or that things have names. On this view, the vocabulary spurt corresponds to a *naming insight* (Dore, Franklin, Miller, & Ramer, 1976; Reznick & Goldfield, 1992; McShane, 1979; Kamhi, 1986). Alternatively, word learning occurs at a faster pace when object concepts and categories become more detailed and refined (Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979; Gopnik & Meltzoff, 1987; Nazzi & Bertoncini, 2003). Other researchers have proposed that the spurt corresponds to linguistic refinements such as word segmentation (Plunkett,

¹We will use the terminology "vocabulary spurt" throughout the manuscript in the sense of a *supra-linear lexical growth*, characterised by slow learning in early development, followed by an increase in the speed of word learning later on. Even though an increase in the speed of word learning in the first years of human life is not questioned, its mathematical description is debated; should it possess a clear inflection point or is there a more gradual increase throughout early development, as suggested by Ganger and Brent (2004)? For the scope of the present manuscript, we use the term "vocabulary spurt" in its general – and milder – interpretation, whereby infants display slow initial learning followed by a faster rate of word learning, contrasting with a linear increase in which the rate of word learning would be constant during life.

1993), word retrieval capacities (Dapretto & Bjork, 2000), improvements in social cognition (Ninio, 1995) or changes in hemispheric specialisation (Mills, Coffey-Corina, & Neville, 1993). All of these hypotheses share the assumption that the vocabulary spurt reflects endogenous changes in the infant.

A second, contrasting, hypothesis has recently been introduced by McMurray (2007). He argued that under the reasonable assumptions that (i) words are learnt in parallel and (ii) some words are easier to learn than most words, a vocabulary spurt is inevitable and that "this distribution in difficulty derives from many factors, including frequency, phonology, syntax, the child's capabilities, and the contexts where words appear." (McMurray, 2007, p.631). Invoking the central limit theorem, he suggested that the individual contributions of the different factors sum to a Gaussian distribution of word difficulty. Later, using the logarithm of utterance statistics as a proxy for word difficulty, he showed that a time-to-acquisition growth curve yields a pattern of vocabulary development typical of infants during their second year. On the basis of this finding, he claimed that "acceleration in vocabulary growth could arise from occurrence statistics alone" (McMurray, 2007, p.631).

Our aim is to clarify the origin of this non-linear increase in the speed of lexical acquisition; whether this transition is the result of a change in the infant's mental representations or brain organisation (endogenous factors), or caused by the statistical nature of the input, such as phonological complexity or the frequency of word usage (exogenous factors). We show mathematically that word frequency cannot alone explain the acceleration in vocabulary growth. This demonstration also fits well with empirical findings that word frequency is not an entirely reliable proxy for word difficulty (Huttenlocher, 1991; Goodman, Dale, & Li, 2008). Instead, we suggest that changes in the infant's learning capacity are required to display the non-linear growth in the speed of word acquisition. These changes, such as the emergence of fast mapping (Carey & Bartlett, 1978), provide the basis for the unique learning capacities displayed late in the second year of human life.

Statement of the problem

For expository purposes, we make three simplifying assumptions; (i) infants only learn words when hearing them, (ii) word occurrence statistics follows Zipf's law and (iii) all words are equally difficult to learn. If these three criteria are satisfied, we demonstrate that vocabulary growth will be linear, unless a change in learning capacity takes place (as a function of time or as a function of the number of words al-

ready present in the lexicon²). In other words, a change in learning capacity is a necessary pre-requisite to drive a non-linearity in vocabulary growth. We justify this claim by both analytical considerations and through simulations. Later, we will show that (i) the assumption of online learning can be relaxed, (ii) that speech corpora used with real infants follow the same behaviour as Zipf’s law and we will suggest that (iii) phonological complexity of early words do not seem to play a prominent role in shaping the vocabulary spurt. We will suggest, therefore, that a change in the infants’ learning capacities is driving the naming explosion.

Let us first justify our initial assumptions. First, we argue that infants learn words when they are confronted with them and not by processing words off-line after accumulating evidence. Carey and Bartlett (1978) introduced the idea that infants are able to “fast map”, whereby infants demonstrate rapid mastery of the appropriate use of labels after a limited number of learning opportunities. Evidence of the infant’s ability to learn a new word after limited exposure was also explored by Woodward, Markman, and Fitzsimmons (1994), suggesting that novel words can be retained at least 24 hours after the infants have been exposed to them only 9 times, even for infants as young as 13 months of age. More recent evidence based on infant-caregiver interactions showed that the naming event needs to occur at the right moment in time when the infant is attending to the named object to be successful (Yu, Smith, & Pereira, 2008)³. These findings provide strong support for the claim that infants perform on-line word learning when exposed to them. Consequently, if infants only engage in online word learning, the raw statistics of word usage should be exploited and not, as in McMurray (2007), a logarithmic transformation of the word frequencies (a further comparison of our approach to McMurray, 2007, is discussed later). Moreover, we will show that even a relaxation of the assumption of online learning cannot explain an accelerated vocabulary growth.

Second, we adopt the perspective that infants are exposed to a distribution of word frequencies approaching Zipf’s law (Zipf, 1949), which states that, from any substantial corpus, the frequency of a word is inversely proportional to its rank. For example, the most frequent word is used twice as much as the second most frequent word and three times more often than the third most frequent word. A broad range of evidence suggests that spoken language essentially follows a Zipf distribution of word usage (Miller & Chomsky, 1963; Zipf, 1935; Beier, 1965; Dahl, 1979; Altmann, 2002). We will show, in a model with constant learning capacity and exposed to a corpus of speech used with real infants, that lexical growth fails to exhibit the characteristics of a vocabulary

spurt, even when the utterance statistics deviate slightly from Zipf’s law.

Analytical considerations

On average, an infant hears a word i having a frequency $f(i)$ within a time window $t(i) = 1/f(i)$. For example, a word uttered twice an hour will be heard on average every 30 minutes and a word uttered 4 times a month will be uttered every week or so. As a consequence, and to a first approximation, the time $T(i)$ to acquire a word i is inversely proportional to its frequency $T(i) \propto 1/f(i)$. The constant of proportionality depends on the number of times a word needs to be heard with respect to the threshold for learning it. Zipf’s law states that, from any substantial corpus, the frequency of a word is inversely proportional to its rank: $f(i) \propto 1/i$. This predicts a linear distribution of time to acquisition; $T(i) \propto i$, which in turn predicts a linear increase in the size of the lexicon. The (constant) speed at which infants increment their lexicon size would then be proportional to their (fixed) learning capacity, as defined by the number of times they need to hear a word in order to add it to their lexicon. In real word learning situations, words do not follow Zipf’s law deterministically. However, the fluctuations in everyday interactions can be modelled by drawing words probabilistically from Zipf’s distribution. Since analytical calculations become increasingly complex, we simulate this process in a stochastic model.

Simulation results

Fig. 1 displays simulations using raw frequencies of word usage from Zipf’s distribution. As in McMurray (2007), a knowledge level is associated with each word and is incremented with each presentation. When this crosses a threshold, the word is learnt. The model reveals a regular increase of word acquisition, the absence of an early, slow learning phase and no inflection point in word learning; in other words, the absence of a vocabulary spurt. The different curves on

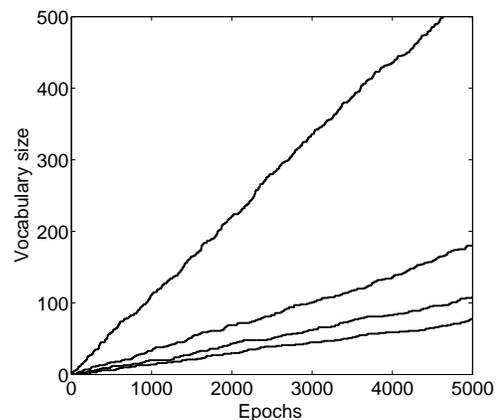


Figure 1: Vocabulary size as a function of time when a model with constant learning capacity is presented with a Zipf distribution of word usage (the different curves correspond to different numbers of words uttered per epoch).

²Mitchell and McMurray (2009) have shown that leveraged learning—the fact that knowledge of some words helps with the learning of others—does not create acceleration in word learning.

³In a recent experiment, Smith and Yu (2008) showed that infants were able to use cross-situational statistics to learn novel words. It remains to be shown, however, if these effects extend to longer time windows than used in the experiment, consisting of multiple presentations of each word-referent pair over the course of 4 minutes.

Fig. 1 correspond to different (constant) learning capacities, i.e., the number of presentations needed to acquire the word in the lexicon. Steeper curves correspond to better learning capacities. Note that, in the model, the absolute number of words uttered in an epoch of exposure also modulates the slope of the learning curves. Similar curves may correspond either to a proficient learner confronted to a low number of words or to a slower learner presented with a higher number of words in any time window. All combinations of learning capacities and absolute number of words uttered per epoch lead to linear increases in the lexicon size.

As a further control we ran simulations with actual word frequencies extracted from the CHILDES Parental Corpus, made out of the following 27 corpora; Bates, Belfast, Bernstei, Bliss, Bloom, Brown, Clark, Cornell, Demetras, Fletcher, Gatherco, Hall, Higginso, Howe, Kuczaj, Macboys (MacWhinney), Macros (MacWhinney), Peters, Post, Sachs, Snow, Suppes, Valian, Vanhout, Vankleec, Warren, and Wells. The Parental Corpus consists of 2.6 million word tokens (about 24,000 word types), and is a representative sample of the speech to which children are typically exposed (MacWhinney, 1991; Li & Shirai, 2000). Fig. 2 shows that for differing numbers of presentations needed to acquire a word (different learning capacities), vocabulary growth lacks both the long latent period of slow learning and subsequent rapid increase characteristics of the vocabulary spurt. Instead, the speed of acquisition is reduced after having learnt about 100 words, and converges to a constant rate of acquisition thereafter. The number of epochs used in this simulation is greater due to the large size of the corpus. We conclude that word

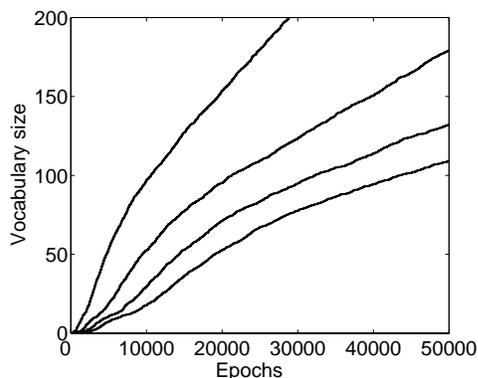


Figure 2: Vocabulary acquisition in a model with constant learning capacity, when presented with parental input. No vocabulary spurt is observed. Higher curves correspond to a lower threshold of learning.

statistics alone cannot be responsible for the vocabulary spurt observed at the end of the second year of life. Rather, as many researchers have suggested, the vocabulary spurt is driven by underlying changes in learning capacity arising from changes in mental representations and/or brain organisation. We next offer a conceptual implementation of this alternative view;

that a change in learning capacity *is required* in order to display an accelerated increase in word learning.

Relaxation of the assumption of ‘online’ learning

We have demonstrated that an acceleration in vocabulary growth cannot be expected when presented with word distributions following Zipf’s law, unless a change in learning capacity is implemented in the model or further variations in word difficulties are present. We have also shown that when the occurrence statistics deviates moderately from Zipf’s law, as exemplified through simulations using the Parental Corpus, a vocabulary spurt is still absent in the model. We now show that the assumption of online learning can also be relaxed.

Let assume that upon presentation of a word, a ‘memory trace’ is initiated. This memory trace would modulate over time the value of the knowledge variable associated with that given word. Let us discuss the potential behaviour of this memory trace. We have already discussed the case for which the memory trace remains constant: It corresponds to the case in which each presentation of a word leads to an increment in the knowledge variable associated with that word, until it crosses a threshold. We have demonstrated earlier that no acceleration in vocabulary growth is observed unless an improvement of learning capacity is implemented in the model. Moreover, frequent words are learnt very early on, thereby failing to reproduce the long latency period observed in early childhood. Alternatively, the memory trace could increase over time (Vlach, Sandhofer, & Kornell, 2008), mimicking consolidation of the word form or meaning during sleep (Dumay & Gaskell, 2007) or through rehearsal of that word. However, high frequency words would be learnt even more rapidly under these conditions than with the constant memory trace, resulting, again, in the absence of the long latency period observed early in life. This account would fail to exhibit the characteristic contrast observed in infancy between a slow initial learning followed by an acceleration in lexical growth. Finally, the memory trace could decay over time, reflecting the degradation in the representations of words in absence of a new utterance, as described by Horst and Samuelson (2008). In this case, low frequency words whose memory trace decays faster than the typical interval between successive word presentations would never be learnt. Although we do not suggest that no learning take place beyond the actual presentation of a word, dynamic memory traces associated with individual presentations of the word are not the ingredient needed to explain the supra-linear vocabulary growth. Decaying memory traces as in Horst and Samuelson (2008) or reinforcement (Vlach et al., 2008; Dumay & Gaskell, 2007) would merely modulate the vocabulary spurt, not create this acceleration.

Relationship to McMurray’s account

Our approach shares a similar goal to that of McMurray (2007): understanding the cause of the sudden increase in the speed of word learning observed during the second half of the second year of life. However, our approach differs in some important respects to both the original paper (Mc-

Murray, 2007) and subsequent implementations (Mitchell & McMurray, 2008, 2009). First, if infants only engage in on-line word learning, the raw statistics of word usage should be exploited and not, as in McMurray (2007), a logarithmic transformation of the word frequencies. In addition to a lack of psychological validity, such a transformation suffers from mathematical instability: depending of the lexicon size, the sum of log-frequencies may become negative, and/or words with a very low usage (frequency smaller than 1 in the time-scale used) would have a negative log-frequency, resulting in negative probability of occurrence. Thus, the vocabulary spurt described in McMurray (2007) is driven by a distribution of word frequencies that, due to its log-sampling, do not reflect the true nature of the statistics of word occurrences. Second, Mitchell and McMurray (2008) introduce a stochastic adaptation of the original model and show that a wide range of distributions can lead to a spurt-like behaviour. Crucially, Zipf's law belongs to the class of distributions that do not lead to a vocabulary spurt.

Finally, Mitchell and McMurray (2009) study leveraged learning in word learning. They explore different metrics for relating word difficulty to word frequency. In a first case, they scale difficulty as an additive function of frequency. In order to avoid the problem of very high frequency words having negative difficulty values, they add a constant value to the difficulty score. The second case, in which word difficulty is scaled to the inverse of frequency is the approach we have chosen: For example, a word that is heard twice as often is deemed to be exactly twice as easy to learn. However, words follow Zipf's law only at a stochastic level. Our analysis, beyond initial analytical considerations, provides a stochastic account of word learning, when infants hear words drawn either from Zipf's distribution or from a corpus consisting of speech to which infants are typically exposed. Mitchell and McMurray (2009) provide a non-stochastic implementation of Zipf's distribution and Mitchell and McMurray (2008) provide a stochastic implementation of non-Zipfian distributions. The critical combination of a Zipfian distribution with a stochastic implementation is absent from their account.⁴

An alternative account

Since a Zipf distribution of word usage is insufficient to capture the vocabulary spurt, we simulate an alternative account where the capacity of learning a word is not kept constant during early life. As infants only learn words on the basis of raw exposure, the model is presented with words drawn from a Zipf distribution and, for each presentation, the model has an increasing probability of learning that word. We presented 10,000 words per "day" in the simulation⁵, out of

⁴"[...] it is important to remember that frequency is not a property of the word [...], it is an estimate of how often it occurs (stochastically) in the child's environment. Thus, our model may be limited in its ability to handle frequency, and a stochastic model may be a better approach for dealing with it (e.g., Mitchell & McMurray, 2008)" (Mitchell & McMurray, 2009, p.1519)

⁵Hart and Risley (1992) reported that, on average, 10- to 18-month-old infants hear 1275 words per hour. Assuming that this

a 40,000 word lexicon distributed with Zipf's law. Words that were presented on average less than once per day, were sampled according to their probability of occurrence within a day. The developmental time course of this probability is implemented as a non-linear function of time, in order to mimic the emergence of fast mapping and increased learning capacity, observed during the second year of life. In the model, the probability of learning a word increases with time; $p(t) = (t/20000)^3$. Any non-linear increase in the probability of learning a word would result in a non-linear developmental trajectory of word learning. Such a change in the parameters would only result in a quantitatively different path to word learning, not a qualitative change⁶. Note that this model is equivalent to a modified version of McMurray's model, in which increment size increases with time. From this perspective, many presentations of a word are needed for successful learning early in development whereas later in the second year, just a single presentation may be sufficient for learning that word, due to the emergence of fast mapping. Fig. 3 depicts the developmental trajectory simulated with the model. The curve of vocabulary acquisition possesses a clear non-linearity separating the early slow learning and the late fast learning regimes, similar to the naming explosion.

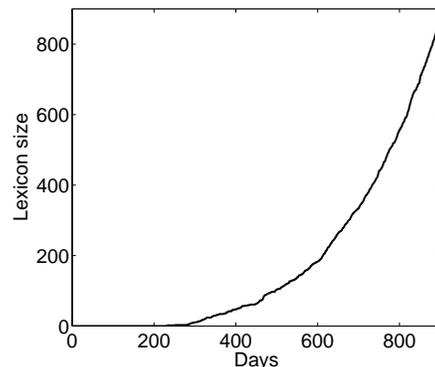


Figure 3: Acquisition in the present model, where learning capacity changes over time.

Discussion

Two contrasting hypotheses have been proposed in order to explain the rapid increase in the speed of word learning occurring in the second year of life. On the one hand, researchers have argued that the vocabulary spurt is driven by changes in the infant's learning capacities, such as the emergence of a naming insight or via maturational changes in the brain. We refer to this view as the *endogenous* hypothesis. In con-

level of exposure is maintained for 8 hours per day, then infants hear about 10,000 words a day.

⁶Since in the present simulations we did not simulate the system for more than 20000 epochs, the probability is always smaller than 1. One could alternatively choose a non-linear function of time that saturates at 1 (or close to 1) for increasing time, so as to mimic a smooth and continuous improvement in learning capacities.

trast, a second hypothesis highlights environmentally-based factors that contribute to the difficulty in learning words, such as frequency, phonological complexity, etc. On this view, the vocabulary explosion is a by-product of variability in word difficulty. We refer to this hypothesis as the *exogenous* hypothesis.

We have argued that simple analytical considerations demonstrate that a linear increase in the size of the lexicon is expected when presented with word frequencies distributed with Zipf's law. Moreover, simulations with a stochastic sampling of words following Zipf's law, as well as with samples of speech to which infants are exposed, confirmed that the type of distribution of word frequencies found in natural language would fail to induce a naming explosion. Mitchell and McMurray (2008) have shown that a wide range of mathematical distributions of word difficulties predict a non-linear growth of the infant lexicon. We have demonstrated that word occurrences following Zipf's law and speech typically heard by infants does not belong to this family of mathematical distributions.⁷

Since we have demonstrated that word frequency cannot account for the vocabulary spurt, it is reasonable to ask whether other exogenous factors that influence word difficulty could be the source of the non-linear vocabulary growth. For example, McMurray (2007) points out that phonological complexity contributes to word difficulty. It is not straightforward to measure the impact of phonological complexity during early word learning since the basis of infant's lexico-phonological representations is not yet well understood. However, as a first approximation, we might consider word length as a proxy for phonological complexity and hence word difficulty. In a recent review, Juhasz (2005) identified contributing factors in picture naming tasks. All reviewed studies (13) showed a correlation between age of acquisition and latency measures, suggesting that latency in picture naming tasks is a reliable way of determining when the word was acquired. In contrast, word length was found to be a significant variable in only 3 studies, whereas 9 studies found it to be non-significant. Phonological complexity, therefore, like frequency may not be a suitable candidate for predicting vocabulary acceleration as "an unavoidable by-product of variation in difficulty". Whereas many factors can impact the distribution of difficulty in learning a word, such as word length or word frequency, it remains to be proven that they play a primary role in determining the shape of the vocabulary spurt. Nevertheless, it is important to highlight that other exogenous factors are likely to contribute to differences in word difficulties. Many researchers would argue that words are not learnt in isolation, and the context in which they

⁷An anonymous reviewer pointed out that a caregiver's word usage may vary over time, despite following Zipf's law at a global scale. As a result, fragments of a caregiver's speech may deviate from Zipf's law, resulting in a vocabulary spurt. The analysis of a biased stochastic sampling of words from a Zipf distribution would be an interesting avenue for further research. However, a random sampling from a Zipf distribution failed to display a spurt-like pattern of word learning.

appear may affect directly the set of potential interpretations of the words, through referential uncertainty. Computational models have shown that word learning in a sentential context can display a spurt-like pattern in the learning curve (Siskind, 1996; Fazly, Alishahi, & Stevenson, 2008) and experimental studies have shown that context diversity and within-context ambiguity can override the role of word frequency (Kachergis, Yu, & Shiffrin, 2009). Nevertheless, Hayes and Ahrens (1988) have shown that there is a positive correlation between a caregiver's mean length of utterance and the age of the infant. As a consequence, young infants are exposed frequently to words in isolation or in short utterances.

We propose, instead, that endogenous factors are primarily responsible for the vocabulary spurt. Among them, the emergence of fast mapping can explain the increase in the ease of acquisition late in the second year of life (Carey & Bartlett, 1978). Further evidence for a change in learning capacity is that word familiarity impacts the distribution of brain regions involved in word learning, reflecting an increased efficiency in the manner in which infants process familiar and novel words across the vocabulary spurt (Mills, Plunkett, Pratt, & Schafer, 2005). It is, however, important to note that neither maturational changes in the brain, nor the application of innate or domain-specific constraints are required to explain a change in learning capacity. For example, Mayor and Plunkett (2008) showed that no specialised mechanisms are needed to explain the vocabulary spurt, as a simple general learning mechanism can lead to the spontaneous emergence of fast mapping. A change in learning capacities, not mechanisms, drives the rapid onset of vocabulary acquisition observed late in the second year of life. Hence, a word that seems difficult for a 15-month-old may be acquired almost instantaneously by a 21-month-old. Is the vocabulary spurt compatible with Zipf's law? The answer is clearly "yes" provided we allow the listener to develop her learning capacities.

References

- Altmann, G. (2002). Zipfian linguistics. *Glottometrics*, 3, 19–26.
- Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy*. New York: Academic Press.
- Beier, E. (1965). *Analysis of word frequencies in spoken language of children*.
- Bloom, L. (1973). *One word at a time: The use of single word utterances*. The Hague: Mouton.
- Carey, S., & Bartlett, E. (1978). Acquiring a new word. *Papers and Reports on Child Language Development*, 15, 17–29.
- Dahl, H. (1979). *Word frequencies of spoken American English*. distributed by Gale Research Co Detroit, Mich.
- Dapretto, M., & Bjork, E. (2000). The development of word retrieval abilities in the second year and its relation to early vocabulary growth. *Child Development*, 635–648.

- Dore, J., Franklin, M. B., Miller, R. T., & Ramer, A. L. H. (1976). Transitional phenomena in early language acquisition. *Journal of Child Language*, 3, 13-28.
- Dumay, N., & Gaskell, M. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, 18(1), 35.
- Fazly, A., Alishahi, A., & Stevenson, S. (2008). A probabilistic incremental model of word learning in the presence of referential uncertainty. In B. Love, K. McRae, & V. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society*. Cognitive Science Society.
- Ganger, J., & Brent, M. (2004). Reexamining the Vocabulary Spurt. *Developmental Psychology*, 40, 621-632.
- Goodman, J., Dale, P., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, 35(03), 515-531.
- Gopnik, A., & Meltzoff, A. (1987). The development of categorization in the second year and its relation to the other cognitive and linguistic developments. *Child Development*, 58, 1523-1531.
- Hart, B., & Risley, T. (1992). American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. *Developmental Psychology*, 28(6), 1096-1105.
- Hayes, D., & Ahrens, M. (1988). Vocabulary simplification for children: A special case of "motherese". *Journal of Child Language*, 15(2), 395-410.
- Horst, J., & Samuelson, L. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, 13(2), 128-157.
- Huttenlocher, J. (1991). Early Vocabulary Growth: Relation to Language Input and Gender. *Developmental Psychology*, 27(2), 236-48.
- Juhász, B. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin*, 131(5), 684.
- Kachergis, G., Yu, C., & Shiffrin, R. (2009). Frequency and Contextual Diversity Effects in Cross-Situational Word Learning. In N. e. a. Taatgen (Ed.), *Proceedings of the 31st annual conference of the cognitive science society*. Cognitive Science Society.
- Kamhi, A. (1986). The elusive first word: the importance of the naming insight for the development of referential speech. *Journal of child language*, 13(1), 155.
- Li, P., & Shirai, Y. (2000). *The acquisition of lexical and grammatical aspect*. Walter de Gruyter.
- MacWhinney, B. (1991). *The CHILDES project : Tools for analyzing talk*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Mayor, J., & Plunkett, K. (2008). Learning to associate object categories and label categories: A self-organising model. In B. Love, K. McRae, & V. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society*. Austin, TX: Cognitive Science Society.
- McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, 317(5838), 631.
- McShane, J. (1979). The development of naming. *Linguistics*, 17, 879-905.
- Miller, G., & Chomsky, N. (1963). Finitary models of language users. In R. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 419-491). Wiley, New York.
- Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1993). Language acquisition and cerebral specialization in 20-month-old infants. *Journal of Cognitive Neuroscience*, 5, 317-334.
- Mills, D. L., Plunkett, K., Pratt, C., & Schafer, G. (2005). Watching the infant brain learn words: Effects of language and experience. *Cognitive Development*, 20, 19-31.
- Mitchell, C., & McMurray, B. (2008). A stochastic model for the vocabulary explosion. In B. Love, K. McRae, & V. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society*. Austin, TX: Cognitive Science Society.
- Mitchell, C., & McMurray, B. (2009). On Leveraged Learning in Lexical Acquisition and Its Relationship to Acceleration. *Cognitive Science*, 33(8), 1503-1523.
- Nazzi, T., & Bertoni, J. (2003). Before and after the vocabulary spurt: two modes of word acquisition? *Developmental Science*, 6(2), 136-142.
- Ninio, A. (1995). Expression of communicative intents in the single-word period and the vocabulary spurt. 8, 103-124.
- Plunkett, K. (1993). Lexical segmentation and vocabulary growth in early language acquisition. *Journal of Child Language*, 20, 1-19.
- Reznick, J. S., & Goldfield, B. (1992). Rapid change in lexical development in comprehension and production. *Developmental Psychology*, 28, 406-413.
- Siskind, J. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61(1-2), 39-91.
- Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106(3), 1558-1568.
- Vlach, H., Sandhofer, C., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, 109(1), 163-167.
- Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology*, 30, 553-566.
- Yu, C., Smith, L., & Pereira, A. (2008). Grounding Word Learning in Multimodal Sensorimotor Interaction. In B. Love, K. McRae, & V. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society*. Austin, TX: Cognitive Science Society.
- Zipf, G. (1935). *The Psychobiology of Language: An Introduction to Dynamic Biology*. MIT Press, Cambridge, Massachusetts.
- Zipf, G. (1949). *Human behavior and the principle of least effort*. Cambridge, Mass.: Addison-Wesley Press.