

# A Fresh Look at Vocabulary Spurts

Frédéric Dandurand (frederic.dandurand@gmail.com)

Department of Psychology, Université de Montréal, 90 ave. Vincent-d'Indy  
Montréal, QC H2V 2S9 Canada

Thomas R. Shultz (thomas.shultz@mcgill.ca)

Department of Psychology and School of Computer Science, McGill University, 1205 Penfield Avenue  
Montreal, QC H3A 1B1 Canada

## Abstract

There is currently rather little agreement about the existence of, and explanation for, a vocabulary spurt in children during the second year. Here we apply a Functional Data Analysis-based technique called Automatic Maxima Detection to the problem of finding vocabulary spurts in a sample of 20 children. Even with considerable smoothing of the data, children were found to exhibit multiple vocabulary spurts of varying intensity and location. These results should provide a clearer target for researchers interested in detecting and explaining these deviations from linear growth.

**Keywords:** vocabulary spurts; functional data analysis; automatic maxima detection.

## Vocabulary Spurts

The psychological literature on vocabulary spurts in children is in an interesting state of turmoil. The spurt is usually taken to mean a sharp increase in vocabulary acquisition in the second year of life. There are at least eight different explanations of the vocabulary spurt with rather little consensus on which is the right explanation, and there is disagreement about whether a spurt even exists in most children. Here, we apply a new statistical methodology to the problem of detecting spurts and find evidence for a surprisingly larger number of vocabulary spurts in most children.

### Explanations of the Vocabulary Spurt

The assertion that there is a substantial and reliable vocabulary spurt during the second year has been repeated so often that most developmental psychologists readily accept it. This apparent consensus on an interesting phenomenon has led to a variety of explanations. Most of these explanations emphasize factors endogenous to the child, some based on sudden developmental changes and others based on leveraging of previous learning.

Among the sudden developmental changes are realizing that things have names (Dore, Franklin, Miller, & Ramer, 1976; Goldfield & Reznick, 1990; McShane, 1979; Reznick & Goldfield, 1992), ability to categorize (Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979; Gopnik & Meltzoff, 1987; Lifter & Bloom, 1989; Mervis & Bertrand, 1994; Nazzi & Bertoncini, 2003; Poulin-Dubois, Graham, & Sippola, 1995), pragmatic skill (Ninio, 1995), and hemispheric specialization (Mills, Coffey-Corina, & Neville, 1993).

Other endogenous factors emphasize leveraging techniques, such that known words facilitate learning of new words. These leveraging methods include mutual exclusivity (Markman, Wasow, & Hanson, 2003), syntactic bootstrapping (Gleitman & Gleitman, 1992), and word segmentation (Plunkett, 1993; Walley, 1993).

A third kind of explanation does not emphasize the child, but rather the statistical properties of word distributions in the child's language environment. Assuming due to the central limit theorem that word-learning difficulty is normally distributed and that words are learned in parallel, computer simulations show that an early vocabulary spurt is mathematically inevitable (McMurray, 2007). The same result was obtained with several other distributions of word difficulty (Mitchell & McMurray, 2008).

Most of these explanations have been disputed, including this last one. Under alternate assumptions that all words are equally difficult to learn and their frequencies are distributed under Zipf's law, simulations show that word acquisition is linear rather than spurt-like (Mayor & Plunkett, 2010). In a Zipf distribution, item frequency is inversely proportional to its rank (Zipf, 1949). This distribution is characteristic of word learning and many other phenomena such as city populations (Itti & Baldi, 2006).

### Methods of Spurt Detection

Four techniques have been used to assess the vocabulary spurt. The simplest is to calculate a ratio of vocabulary size to age and argue whether it is large enough to be a spurt (Schafer & Plunkett, 1998). This method is not particularly convincing as it does not assess change or rate of change. The most common approach is to specify a certain number of words that must be learned in a given time period (Goldfield & Reznick, 1990; Gopnik & Meltzoff, 1987; Lifter & Bloom, 1989; Mervis & Bertrand, 1994; Ninio, 1995; Poulin-Dubois, et al., 1995; Reznick & Goldfield, 1992). A third approach is to plot vocabulary growth over age and visually judge whether a spurt is present (Dromi, 1987).

All three of these techniques use subjective and arbitrary specifications, and it is not too surprising that values are chosen in a way that guarantees that the expected spurt is found. Moreover, none of these methods distinguish a true spurt from a gradual continuous increase in words (Bloom, 2000).

The fourth, and more sophisticated, technique fit particular functions to vocabulary growth data (Ganger & Brent, 2004). If a logistic (S-shaped) function fit the data better than a quadratic (curvilinear or linear) function, then a spurt was considered to be present. Presence of a spurt was assessed by noting whether the root mean-squared residual for the quadratic function was more than twice that for the logistic function. Such residuals are smaller for better fits. Comparing the two values by dividing one by the other is known the log-likelihood ratio. Interestingly, only 4 of 20 children showed a vocabulary spurt by this measure, leading the authors to question the general existence of this spurt.

This kind of curve fitting is a definite improvement over the other three methods in objectivity and precision, but there are some limitations. Only two functions are tried, and it is not clear that these two functions can always differentially identify a spurt. For example, only a part of the logistic (just before the inflection point) can identify a spurt, and the rapidly increasing section of a quadratic could also resemble a spurt. As well, to optimize a function's fit, hundreds of parameter values for the functions had to be searched for and tried. Finally, this technique, like the other three, assumes that there is at most one spurt to find. None of these techniques are able to objectively identify the start, end, central location, or amplitudes of single or multiple spurts.

### The AMD Technique

Automatic Maxima Detection (AMD)<sup>1</sup> is a technique to automatically detect and measure statistically reliable maxima in functions of one variable or any of their derivatives (Dandurand & Shultz, 2010). Because growth spurts are characterized by local increases in the rate of change, AMD finds spurts as maxima in the first derivative of time-varying measures. In AMD, only two free parameters influence the number of spurts detected: (1) the smoothing parameter lambda ( $\lambda$ ), and (2) the  $p$ -value that determines the threshold used for statistical significance (Dandurand & Shultz, in press).

AMD takes as input a sample of data pairs  $(y_j, t_j)$  where  $y_j$  is a measure of interest (e.g., vocabulary size) and  $t_j$  are corresponding sampling times. AMD first uses Functional Data Analysis (FDA) (Ramsay & Silverman, 2005) to fit a spline-based smooth function to approximate the data sample, as well as the first three derivatives of this smoothed, continuous function to identify important function markers (see Figure 2).

The fitted function takes the form of a weighted linear combination of B-spline basis functions  $\phi_k$ ,  $k = 1, \dots, K$ :

$$x(t) = \sum_k^K c_k \phi_k(t)$$

FDA uses a roughness penalty approach to smoothing which limits or penalizes the size of some higher-order

<sup>1</sup> A Matlab implementation of AMD, bundled with FDA, can be downloaded from: <http://lnsclab.org/lib/AMD/>.

derivative of the smoothed function. Coefficients  $c_k$  are selected to minimize a penalized sum of squared errors (SSE) between the estimated function and observed data vector  $y$ :

$$PENSSE(y | c) = (y - \phi c)' W (y - \phi c) + \lambda c' R c$$

Where:  $c$  is a vector of coefficients  $c_k$ ;  $W$  is a symmetric positive definite weight matrix;  $\phi$  is the matrix of basis function values  $\phi_k(t_j)$ ;  $\lambda$  is a smoothing parameter; and  $R$  is a roughness penalty matrix, computed as follows:

$$R = \int \left[ \frac{d^2}{dt^2} \phi(t) \right] \left[ \frac{d^2}{dt^2} \phi'(t) \right] dt$$

Note that the fitted curve  $x(t)$  becomes increasingly smooth as lambda ( $\lambda$ ) increases; this smoothing value lambda is the only parameter in AMD that is manually set. There are techniques to automate selection of lambda (Dandurand & Shultz, 2010). However, existing techniques have important limitations, and so a careful manual selection is preferred (Ramsay, Hooker, & Graves, 2009).

### Determining which spurts are significant

To determine which spurts are significant, AMD first estimates a confidence interval of the derivative (velocity) of the curve (see dotted lines surrounding the fitted curve in Figures 1 and 2) for the  $p$ -value provided (a standard value of .05 is used here). Width of confidence intervals (also called point-wise bands) is based on the variance of the fitted function:

$$Var[x] = \phi' Var[c] \phi$$

Where:  $\phi$  is a matrix of basis function values at the observation points and  $Var[c]$  is the variance of coefficients  $c_k$ , computed as follows:

$$Var[c] = (\phi' W \phi)^{-1} \phi' W \sum_e W \phi (\phi' W \phi)^{-1}$$

Where:  $W$  is a symmetric positive definite weight matrix; and  $\Sigma_\epsilon$  is the variance-covariance matrix of the residual vector  $\epsilon$ .

Second, AMD lists all local maxima in the velocity function as spurt candidates. Finally, for each spurt candidate, a null hypothesis is tested in which a straight line between the two local minima adjacent to the maximum in velocity is contained within the confidence band. This null hypothesis thus corresponds to an absence of spurt. A candidate is a genuine spurt when this null hypothesis has to be rejected, that is, the maximum of velocity is significant (see Figure 1). For a spurt to be considered significant, velocity not only has to significantly increase to indicate the beginning of a spurt, it must also significantly decrease to mark the end of the spurt (Dandurand & Shultz, 2010).

For example, in Figure 7, a spurt is not detected in the upper section of the curve because velocity keeps on increasing (that is, acceleration is always positive). This design decision reflects the fact that not all processes eventually decelerate or stop, like physical growth does. In

other processes, such as growth of national economies, the existence of such upper bounds is less clear. In such cases, an acceleration in growth may not mark a spurt per se, but instead a transition for a slower steady-state rate of increase to another, faster steady-state rate. The definition of spurts in AMD excludes the latter case.

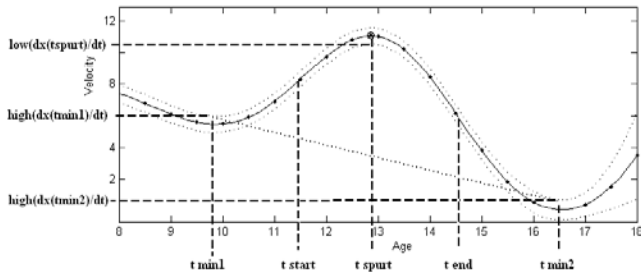


Figure 1: Example of a significant spurt. Dotted lines above and below the smooth function correspond to 95% confidence bands.

### Quantifying spurts

AMD also provides rigorous quantification of the important features of significant spurts: (1) when the spurt starts, (2) the point where it is most intense (maximal velocity), (3) the spurt amplitude and (4) the spurt duration. An example is given in Figure 2. A spurt starts when velocity is at an inflection point, acceleration is at a local maximum, and jerk crosses 0 at a negative slope. A spurt peaks when velocity is at a local maximum, acceleration crosses 0 at a negative slope, and jerk is negative. A spurt ends when velocity is at an inflection point, acceleration is at a local minimum, and jerk crosses 0 on a positive slope. Spurt amplitude is given by the vertical distance from acceleration at the start to acceleration at the end.

In previous work, AMD successfully detected and measured three important and well-known phenomena of physical growth of children: (1) An adolescent growth spurt in virtually all children; (2) an earlier age of onset for girls' adolescent growth spurts than for boys'; and (3) a smaller, pre-adolescent growth spurt in some children. Such spurts tend to be small and difficult to detect without techniques like AMD (Dandurand & Shultz, 2010).

### Applying AMD to the vocabulary spurt

Our previous work also included some preliminary results for vocabulary spurts. In simulated data derived from a computational model of this spurt (McMurray, 2007), AMD flexibly found one large, global spurt under a low degree of sampling, and many local, mini-spurts under higher sampling.

In real vocabulary data from three English-speaking children (Corrigan, 1978), AMD found an average of 2.0 spurts per child. An example of a child who had 3 significant spurts is presented in Figure 3. These data had also been previously analyzed using FDA but without the

benefit of automated detection of significance (Shultz, 2003). Hence, identification of spurts and plateaus had to be performed manually. These early results suggested that children exhibit multiple vocabulary spurts, but the analysis was limited by the small sample size and limited number of observations. In the current project, we analyze a new set of data with a larger sample size and more densely sampled observations.

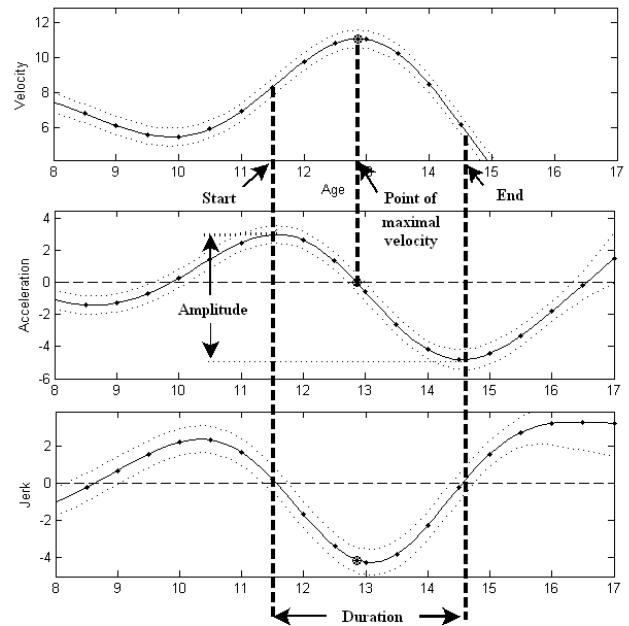


Figure 2: Example of quantification of spurt features: where it starts, where it is most intense, its amplitude and duration.

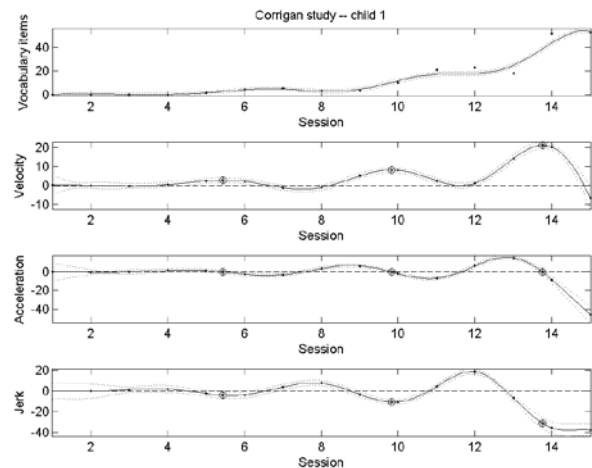


Figure 3: Example of an AMD analysis of vocabulary growth in a child over a few observations.

## Method

Here we apply the AMD method to data from each of 20 children from an online database (Ganger, 2004) whose vocabulary growth had been recorded daily over the second year of life (Ganger & Brent, 2004). These are the same 20 children studied in Ganger and Brent's (2004) Experiment 1. Parents had listed the words used by their child every day, even if the words had been used previously. Imitations had been excluded and context noted. As Ganger and Brent (2004) note, systematic parental reports on vocabulary enable large sample sizes, enhanced validity, and good reliability. We converted new words per day to cumulative vocabulary, and assumed a variation of zero when an observation value was not available (i.e., for missing data).

The main unspecified parameter in AMD, as in most FDA techniques, is lambda, which controls the amount of smoothing that is applied to the raw data. We tried several lambda values (1, 100,  $1 \times 10^5$ , and  $1 \times 10^{10}$ ) and noticed the usual decrease in detected spurts with increases in lambda. For this paper, we concentrate on the results with a lambda of  $1 \times 10^{10}$ , which is a very large value. Our results are therefore conservative; even more spurts were found using less smoothing.

## Results

A plot of spurt locations and amplitudes for all 20 children at a lambda value of  $1 \times 10^{10}$  is presented in Figure 4. The number of spurts in this smoothing condition ranged from 1 to 6, with a mean of 3.0 and SD of 1.5. Three children showed only 1 spurt. Examples of representative individual cases are shown in Figures 5 and 6. Child 10 had 2 spurts, with the largest one being last at 554 days. Child 22 showed 4 spurts, the largest one being first, at 269 days. The variability in number, location, and amplitude of spurts is notable, all of which is obscured in a group plot of vocabulary spurts averaged across all 20 children (see Figure 7). Only 2 spurts appear in the averaged plot and they bear very little relation to results for any of the 20 individual children. The dynamics and variability of vocabulary growth are nearly completely obscured by group plots because spurts in individuals vary greatly in location, number, and intensity.

For comparison, the mean number of spurts at a smaller lambda value of  $1 \times 10^5$  was 6.5, with SD 2.8. The range was from 3 to 15 spurts.

## Discussion

There has been considerable disagreement about both the existence and causes of the vocabulary spurt, but until now virtually no consideration of the number of such spurts. The classical position has assumed one such spurt, during the second year. More recently, reviews of this evidence (Bloom, 2000) and results from more precise methods (Ganger & Brent, 2004) have cast serious doubts on the general existence of a vocabulary spurt. Earlier work had not taken the existence issue very seriously and had failed to

distinguish gradual linear vocabulary growth from a genuine spurt, defined as a sharp increase against a background of linear growth.

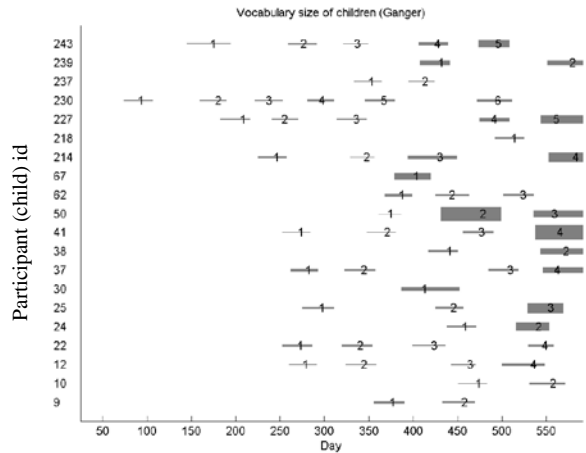


Figure 4: Spurts in vocabulary growth for 20 children ( $\lambda = 1 \times 10^{10}$ ). Spurt length is indicated by a horizontal line, spurt number by a number on the line, spurt amplitude by line thickness, and spurt peak by the location of the number.

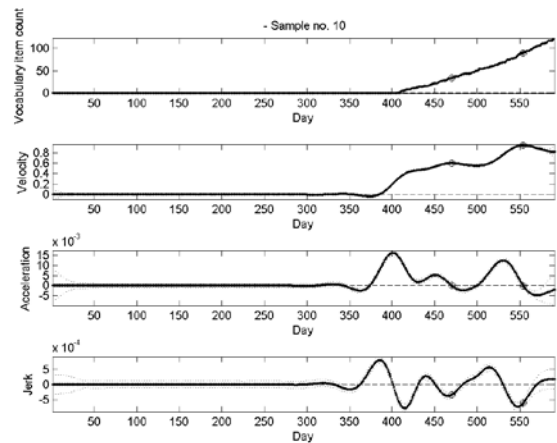


Figure 5: Vocabulary growth for child 10. The 2 spurt peaks are indicated by filled circles.

The present work applies a relatively new FDA technique to vocabulary spurts. This AMD technique automatically detects growth spurts from densely-recorded observations on an individual, ensuring that each detected spurt is statistically reliable, thus distinguishing spurts from linear growth. AMD had been successfully applied to several growth phenomena including longitudinally-measured physical growth (Dandurand & Shultz, 2010). As in previous AMD applications to vocabulary growth, we found multiple spurts in most children tested. And like in previous work applying FDA methods to growth (Shultz, 2003), the

number of spurts detected increased with less smoothing of the data.

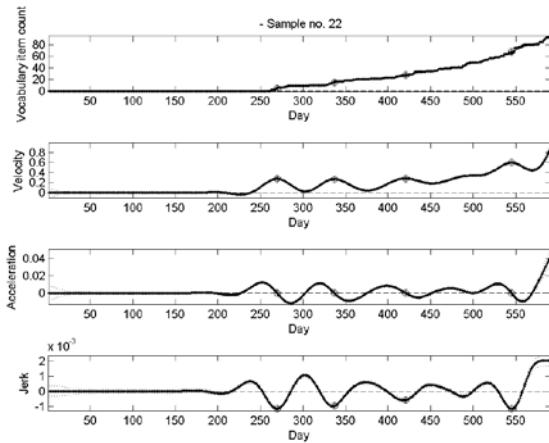


Figure 6: Vocabulary growth for child 22. The 4 spurt peaks are indicated by filled circles.

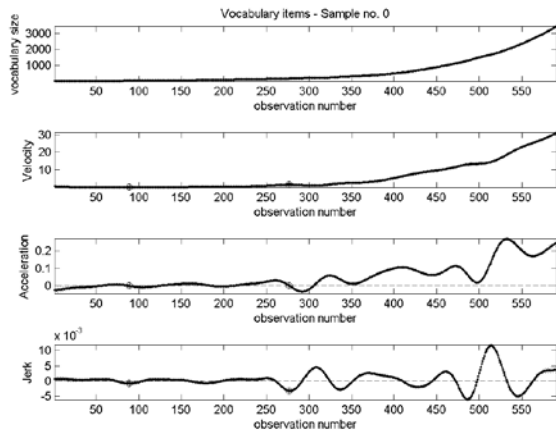


Figure 7: Mean vocabulary growth across all 20 children. The 2 spurt peaks are indicated by filled circles.

Smoothing is valuable for ignoring small, random, and uninteresting bumps in growth curves. In other words, smoothing allows AMD to focus only on the larger and more noticeable spurts. Even with an extremely large, and thus conservative, value of  $1 \times 10^{10}$  for the lambda smoothing parameter, multiple vocabulary spurts were found in the majority of children (17 of 20). With a still large and conservative lambda of  $1 \times 10^5$ , all children showed at least 3 spurts and a mean of 6.5 spurts. In future work, we plan to apply AMD to other available datasets on vocabulary and other kinds of growth.

Although AMD can detect statistically significant departures from linearity, interpretation of their theoretical importance requires expertise in the domain of the relevant study (Dandurand & Shultz, 2010). Interpretations of identified spurts may depend on the quality of the controls

included in the research design. For example, volubility of children could vary according to their health status, mood, and transient environmental stimuli. Empirical studies should control as much as possible for such effects, because, based on data alone, no statistical tool can distinguish reliable spurts caused by such effects from those due to cognitive and developmental processes.

Our findings could influence the literature on spurt causation, most of which has assumed only a single spurt at a particular age. Some of the proposed explanations, particularly those dealing with sudden developmental changes in other systems, may not enjoy being stretched to cover various multiple spurts at different ages, along with large individual differences in number, timing, and intensity of spurts. In any case, better documentation on the number and location of statistically reliable vocabulary spurts should provide researchers seeking explanations of these spurts with a clearer and more realistic target.

## Acknowledgments

This work is supported by a grant to TRS and a post-doctoral fellowship to FD, both from the Natural Sciences and Engineering Research Council of Canada. Marie Lippeveld, Kyler Brown, Caitlin Mouri, and Vincent Berthiaume each provided helpful comments on an earlier draft.

## References

- Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy*. New York: Academic Press.
- Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.
- Corrigan, R. (1978). Language development as related to stage 6 object permanence development. *Journal of Child Language*, 5, 173-189.
- Dandurand, F., & Shultz, T. R. (2010). Automatic detection and quantification of growth spurts. *Behavior Research Methods*, 42(3), 809-823.
- Dandurand, F., & Shultz, T. R. (in press). Automatic Maxima Detection: A graphical user interface and a tutorial. *Tutorials in Quantitative Methods for Psychology*.
- 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.
- Dromi, E. (1987). *Early lexical development*. Cambridge: Cambridge University Press.
- Ganger, J. (2004). Data from Ganger & Brent 2004 Retrieved 24 January 2011 <http://www.pitt.edu/~jganger/spurtdownloads.html>
- Ganger, J., & Brent, M. (2004). Reexamining the vocabulary spurt. *Developmental Psychology*, 40, 621-632.

- Goldfield, B., & Reznick, J. S. (1990). Early lexical acquisition: Rate, content, and vocabulary spurt. *Journal of Child Language*, 17, 171-183.
- Gopnik, A., & Meltzoff, A. N. (1987). The development of categorization in the second year and its relation to other cognitive and linguistic developments. *Child Development*, 58, 1523-1531.
- Itti, L., & Baldi, P. (2006). Bayesian surprise attracts human attention. In Y. Weiss, B. Schölkopf & J. Platt (Eds.), *Advances in Neural Information Processing Systems* (Vol. 18).
- Lifter, K., & Bloom, L. (1989). Object knowledge and the emergence of language. *Infant Behavior & Development*, 12, 395-423.
- Markman, E. M., Wasow, J. L., & Hanson, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, 47, 241-275.
- Mayor, J., & Plunkett, K. (2010). Vocabulary spurt: Are infants full of Zipf? In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 836-841). Austin, TX: Cognitive Science Society.
- McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, 317, 631.
- McShane, J. (1979). The development of naming. *Linguistics*, 17, 879-905.
- Mervis, C. B., & Bertrand, J. (1994). Acquisition of the novel name-nameless category (N3C) principle. *Child Development*, 65, 1646-1662.
- 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.
- Mitchell, C. C., & McMurray, B. (2008). A stochastic model for the vocabulary explosion. In B. C. Love, K. McRae & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1919-1924). Austin, TX: Cognitive Science Society.
- Nazzi, T., & Bertoncini, 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. *Childrens Language*, 8, 103-124.
- Plunkett, K. (1993). Lexical segmentation and vocabulary growth in early language acquisition. *Journal of Child Language*, 20, 1-19.
- Poulin-Dubois, D., Graham, S., & Sippola, L. (1995). Early lexical development: The contribution of parental labeling and infants' categorization abilities. *Journal of Child Language*, 22, 325-343.
- Ramsay, J., Hooker, G., & Graves, S. (2009). *Functional data analysis with R and MATLAB*. New York: Springer.
- Ramsay, J., & Silverman, B. (2005). *Functional data analysis* (2nd ed.). New York: Springer.
- Reznick, J. S., & Goldfield, B. A. (1992). Rapid change in lexical development in comprehension and production. *Developmental Psychology*, 28, 406-413.
- Schafer, G., & Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development*, 69, 309-320.
- Shultz, T. R. (2003). *Computational developmental psychology*. Cambridge, MA: MIT Press.
- Walley, A. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review*, 13, 286-350.
- Zipf, G. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.