Submitted paper to be presented at INTERSPEECH 2017: Situated Interaction. Stockholm, Sweden. August 20-24.

Modeling Multiple Influences on Vocabulary Acquisition: Context, Symbol, and Association Learning

Dominic W. Massaro [1]

[1] Department of Psychology
University of California, Santa Cruz

Massaro@ucsc.edu

Abstract

The goal of this paper is to provide a theoretical taxonomy of the contribution of a variety of potential influences on children's vocabulary acquisition. The variables analyzed in this study include parental input frequency, articulation difficulty, iconicity, and concreteness (imageability). Within the framework of the Fuzzy Logical Model of Perception (FLMP), it is assumed that there are three important components in the acquisition of new vocabulary: contextual or stimulus learning, symbolic or response learning, and the association between the context and symbol. It is further assumed that parental input drives the learning process, concreteness influences context learning, articulation difficulty influences symbol learning, and iconicity influences learning the association between the context and symbol. Previous research on children's expressive vocabulary has shown strong effects of all four variables across the first four years of life. Parental input has the largest influence, articulation difficulty negative impacts word production, and the positive influence of iconicity and concreteness decrease systematically with increasing age. The relative independence of these influences is interpreted as support for the FLMP theoretical framework.

Index Terms: vocabulary acquisition, articulation difficulty, concreteness, iconicity, age of acquisition, parental input frequency, child directed speech

1. Introduction

In the course of our research, we have found the Fuzzy Logical Model of Perception (FLMP) to be a universal principle of perceptual cognitive performance that accurately simulates human pattern recognition [1] and [2]. People are influenced by multiple sources of information in a diverse set of situations. In most cases, these sources of information are ambiguous and any particular source alone does not usually specify completely the appropriate interpretation.

The goal of this paper is to evaluate whether the same FLMP processes also occur in language acquisition, not just in accomplished language users [1, Chapter 8]. Although multiple cues are functional [3], the relative influence of these cues must necessarily change across development and [4xx]. For example, research has shown that concreteness (imageability) and iconicity are strong influences early in vocabulary acquisition but their influences decrease dramatically with increasing vocabulary [5] and [6]. These cues and constraints are graded (not categorical), suggesting further that they must be combined to give a more reliable understanding of the input. Evidence to date indicates that this combination process is highly efficient or optimal, as described by a Bayesian-like process [7] and [8].

2. Influences on Vocabulary Acquisition

This framework sets the stage to evaluate various factors in vocabulary acquisition in terms of how and when they influence word learning. Taking an information processing approach, we try to locate specific influences at specific component processes. We assume that learning a word involves at least three different processes: context learning, symbol learning, and the association between the context and symbol. The goal is to locate various influences at one of these three component processes. By locating these influences on specific processes, we are better able to account for the nature of word learning across the first years of a child's life.

Parental input is a unique influence because it affords the actual word-learning event when the child experiences a word in a given context. The other variables under consideration are assumed to modulate the influence of this learning event. As a first attempt to circumscribe several influences, it is assumed that concreteness influences context learning, articulation difficulty influences symbol learning, and iconicity influences the learning of the association between context and symbol. Figure 1 illustrates the influence of these four variables. We describe each of these variables, beginning with parental input (or similarly input by caregivers and peers).

2.1. Parental Input (Child Directed Speech)

It is well established that many of the words children know are those that they have heard, especially in child directed speech [5]. The number of times a child hears the word improves the learning of that word. Researchers have operationalized this variable as parental input frequency and asked how word frequency of this child directed speech influences vocabulary acquisition. Increases in parental input frequency would correspond to increases in the number of learning trials for the child. Following the well-known finding of time on task or repeated practice, we expect that the more often a word is presented the better learned it will be. These presentation trials (with or without feedback) would improve learning so that each new experience would increase the likelihood of the child learning and remembering the vocabulary item.

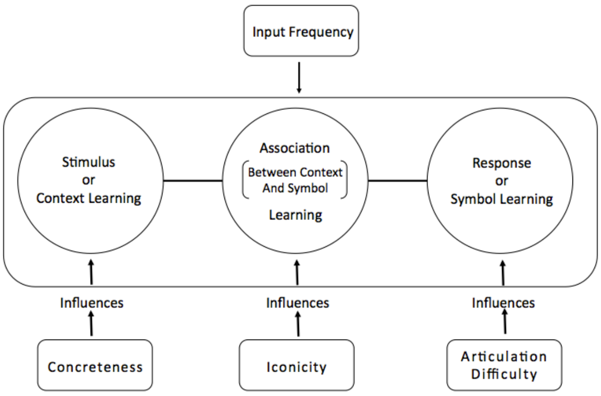

Figure 1*. Depiction of the three distinct processes involved in the acquisition of a vocabulary word along with the accompanying selective influence of concreteness, iconicity, and articulation difficulty. Word input initiates learning.*

2.2. Difficulty of Articulation (Symbol Learning)

A child's production of words has been shown to be related to the difficulty of articulation of segments of the words [9]. We [9] found that difficulty of articulation of words was negatively correlated with the likelihood of very young children between 8 and 18 months having those words in their expressive language.

A parsimonious description of the influence of articulation difficulty in vocabulary acquisition would be to locate the influence of this variable at the symbolic response learning stage. This means that words that are more difficult to articulate would be more difficult to learn because they are more difficult to pronounce and therefore we could expect that the perception and memory representation would be degraded relative to words that are easier to pronounce.

2.3. Iconicity (Association Learning)

Iconicity can be defined as a correspondence between the speech characteristics of a word and the characteristics of the object or event that the speech symbol represents [10] and [11]. Iconicity has been repeatedly demonstrated to influence vocabulary acquisition for both receptive and productive language [6]. Within the framework of the FLMP, we expect iconicity to have its influence in associating the context stimulus information with the symbolic response information.

2.4. Concreteness (Context Learning)

Behavioral science has repeatedly demonstrated that concreteness or imageability are easier to learn and remember. Research has also demonstrated an influence of these highly correlated variables in vocabulary acquisition. Like iconicity, the influence of concreteness diminishes with increases in a child's age or increases in vocabulary [6]. We expect that concreteness or imagineability would have an influence on the context or stimulus learning in vocabulary acquisition, as opposed to response learning or learning the association between stimulus learning and response learning.

3. Theoretical Model

Given the early stages of model development, the current presentation should be viewed as a simple framework of the time course of vocabulary acquisition and to better represent its various influences. Although the data that have been analyzed to date might be viewed as consistent with the model it is understood that no critical tests have yet been performed. What might be necessary is the independent variation of these variables in language acquisition studies.

The FLMP shown in Figure 2 is meant to describe the acquisition of new vocabulary for young children. The influences described by the model would be primarily determined by the characteristics of the vocabulary being learned. These characteristics correspond to at least the three variables we have described: concreteness, articulation difficulty, and iconicity, and their respective influences that they have on context learning, symbol learning, and learning the association between these two constituents.

First, however, it is necessary to describe how a word input influences its learning. Although researchers have not systematically studied the actual duration of a word presentation, it is important to stimulate how learning occurs across presentation time. Following perceptual acquisition in other domains, we assume that the learning of a word can be described by

$$s(W_i) = \alpha_i(1 - e^{-\theta t}) \qquad (1)$$

where $s(W_i)$ corresponds to the perceptual/memory strength of word i, α is the asymptotic strength, and θ indicates the rate of growth to that asymptote with increases in study time *t*.

This is the engine that drives the learning of a new word. The three components concreteness, articulation difficulty, and iconicity modulate this learning. Equation 1 would be applied separately for the multiple influences in word learning. In this case, concreteness, articulation difficulty, and iconicity would follow the time course of Equation 1 with unique α and θ parameters for each component of a given word.

Of course, a child will require multiple experiences in order to learn a new word. One way to implement the expected improvement with repeated experiences of a given word would be to allow α to grow with each new experience. We know that the improvement is unlikely to be linear but more roughly logarithmic as given by Equation 2.

$$\alpha_n = \alpha_{(n-1)}(1 - e^{-\gamma n}) \qquad (2)$$

where n is the n^{th} trial, and ϒ is a rate of growth factor. In this case, α grows in such a manner that each additional presentation of a word increases α but at a decreasing amount with increasing trials. The α in Equation 1 would be determined by Equation 2. Equations 1 and 2 are highly analogous except that they are based on processing time in a given presentation and repeated presentations, respectively.

There is a somewhat obvious limitation in the descriptions provided by Equations one and two. These equations assume that the learning process is constant across development or increases in age. However it is well known that children improve in their perception, attention and memory as they increase in age. This improvement is not represented in the current formulation of the model. Somehow the variables in these equations would have to be modulated with increases in developmental abilities. The cue combination would remain constant [see 7, Chapter 5] but it would be modulated by the parameters in the equations.

Context learning. Context learning corresponds to some form of understanding of the situation in which the vocabulary item occurs. As stated previously, we expect this learning to be easier for those contexts or situations in which concrete words or high imageability words are appropriate. Thus a child more easily learns a concrete word than an abstract word not because of added symbol or associative information but rather because the situation corresponding to the concrete word is easier to understand, comprehend and learn than is the situation containing an abstract word.

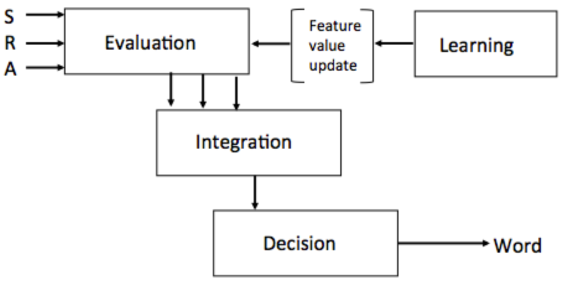

Figure 2. *Schematic representation of the three processes involved in understanding or producing a word utterance. The three processes are shown to proceed left to right in time to illustrate their necessarily successive but overlapping processing. These processes make use of prototypes stored in long-term memory. The different sources of information are stimulus context (S), response symbol (R), and the association between these two constituents (A). The evaluation process transforms these sources of information into psychological values. These sources are then integrated to give an overall degree of support, for each word alternative. The decision operation maps the outputs of integration into some word alternative.*

Articulation Difficulty. Articulation difficulty has been shown to negatively impact the learning of new vocabulary, particularly for productive vocabulary [5], [6], and [9]. Thus, *ceteris paribus*, words more difficult to articulate are at a disadvantage in the learning of new vocabulary. This disadvantage comes about because of symbol learning, and not context learning or the association learning.

Iconicity. Iconicity would have its influence primarily at the learning of the association between the context and symbol rather than the learning of these two components themselves. Highly iconic words would allow for a more fluid association between the context and symbol because of the cross model matching between these two dimensions [6].

Three processes involved in word perception and/or production are illustrated in Figure 2 and include evaluation, integration, and decision. These processes make use of word prototypes stored in long-term memory. A word prototype holds the appropriate contextual, symbolic, and associative information. For example, Deb Roy's son usually experienced the word water in the kitchen [12]. The evaluation process transforms these sources of information into psychological values indicating degrees of supports for various word alternatives, which are then integrated to give an overall degree of support for each vocabulary alternative.

The decision operation maps the outputs of this integration into some appropriate word alternative [13]. For example, previous research indicated that young children are capable of integrating auditory, visual, and gestural information in determining a word's referent [14]. Prototype representations within the FLMP meet Frank's [15] criterion "that models should be efficient compressions of input data at the desired level of analysis, and second, that models should include some bias towards parsimony in the representations they learn."

The assumptions central to the model are: (a) each source of information is evaluated to determine the continuous degree to which that source specifies various alternatives; (b) the sources of information are evaluated independently of one another; (c) the sources are integrated to provide an overall continuous degree of support for each alternative; and (d) perceptual identification and interpretation follows the relative degree of support among the various alternatives.

Given multiple sources of information, it is useful to have a common metric representing the degree of match of each feature. Features that define a prototype can be related to one another more easily if they share a common currency. To serve this purpose, fuzzy- truth values [16] are used because they provide a natural representation of the degree of match. Fuzzy-truth values lie between 0 and 1, corresponding to a proposition being completely false and completely true. With two possible alternatives, the value .5 corresponds to a completely ambiguous situation, whereas .7 would be more true than false and so on. Fuzzy-truth values, therefore, not only can represent continuous rather than just categorical information, they also can represent different kinds of information. The truth values for a vocabulary word would correspond to the likelihood of a word category given contextual, symbolic, and associative information. Truth values are also sensitive to the variability of the information signaling a specific word.

Figure 2 also illustrates how learning is conceptualized within the model by specifying exactly how the feature values used at evaluation change with experience. Learning in the FLMP can be described by the following algorithm [17] The initial feature value representing the support for an alternative is initially set to .5 (since .5 is neutral in fuzzy logic). A learning trial consists of contextual, symbolic, and associative information. Given this experience, the prototypes would be updated following Equations 1 and 2. Thus, the child is continuously modifying the prototype representations and these in turn will become better tuned to the informative components of the vocabulary being experienced. This continuous adjustment of memory representations is highly similar to many contemporary views of language acquisition [18].

The current contribution extends the FLMP to describe vocabulary learning. The model qualifies as a probabilistic learning mechanism within a recent taxonomy of computational models [15]. The FLMP instantiates probabilistic learning because it describes how children are able to acquire vocabulary even from observations that are individually ambiguous among a number of different hypotheses. As stated earlier, a successful receptive or productive use of the word requires at least three sources of information: contextual information, symbolic information, and an appropriate association between context and symbol.

In a two-alternative task with two word alternatives, the degree of support for one alternative (say *mommy*) can be represented by three sources of information. Let s_i, r_i, and a_i represent the support for situational, symbolic, and association information for mommy. Given just two word alternatives, the support for situational, symbolic, and association information for *daddy* would be given by s_j, r_j, and a_j, respectively. Thus, the probability of recognizing the word *mommy* would be

$$P(mommy) - \frac{s_i r_i a_i}{s_i r_i a_i + s_j r_j a_j} \quad (3)$$

The probability of recognizing the word for *daddy* would be simply one minus the support for mommy.

With multiple word alternatives, which is usually the case, then the FLMP predictions are very similar to Equation 3 in that the support for the various alternatives would be computed. In this case, the *P(mommy)* would simply have all competing alternatives in the denominator of Equation 3.

4. Empirical Investigation

To evaluate the contribution of concreteness, iconicity, articulation difficulty, and parental input frequency, we used the child production data from part of the Childes (2015) database [19]. This database consists of 5000 transcriptions of children's speech with 3,500,000 word tokens, and spans children from birth to 7 years of age [20] and [21]. Notably, we use word frequency in the ChildFreq database as a measure of vocabulary acquisition. The more often a word occurs in an age range, it is reasoned that the more likely that word had been acquired by that age. More specifically, this measure indexes how much children are using a word during a given age range.

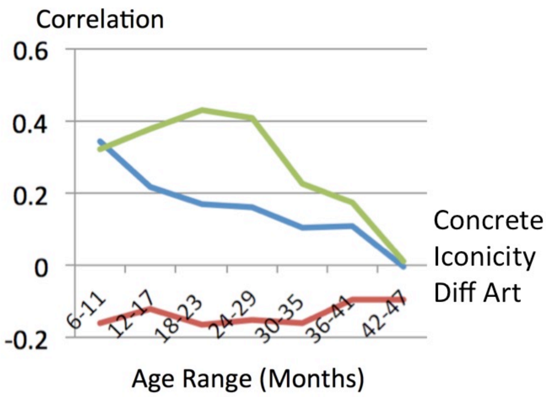

Figure 3. *Partial correlations the three independent variables iconicity, difficulty of articulation, concreteness carried out on the log 10 word child frequency at seven different age ranges (in months).*

We assessed the relationship among the four independent variables by correlating their measures on the 644 test words or some subset of the words when a measure was not available for some of the words. For our analyses, we used existing measures of concreteness [22], iconicity [6], articulation difficulty [9], and parental input frequency [23]. We computed the log 10 number of times each of the 644 words occurred across the seven age ranges 6-11,12-17,18-23, 24-29, 30-35, 36-41, and 42-47 months in the ChildFreq database. The first 6 months of life was not included because only one transcript was available. The raw frequency counts were normalized per one million occurrences.

Parental input frequency was negatively correlated with the other three variables: parents utter fewer words that are high in iconicity, concreteness, and difficult to articulate.

To evaluate the relationship between the four independent measures and children's word frequency, we report partial correlations. Parental input frequency had the largest influence but the other three valuables also had significant effects. As can be seen in Figure 3, both concreteness and iconicity contributed positively to word production and this contribution decreased systematically with increasing age. Articulation difficulty, on the other hand, impeded word production across all ages.

Figure 3 can also be interpreted to mean that abstract words low in concreteness and non-iconic words are more difficult to learn. Within the framework of the FLMP, this means that more parental input of those words are necessary for learning to take place. This means that these words are necessarily acquired at a more mature age than are concrete and iconic words. In terms of Equation 1, the support values s_i and r_i are smaller for abstract and non-iconic words relative to concrete and iconic words.

5. Potential Tests of the Model

The FLMP is usually tested in experiments that systematically manipulate several sources of information and evaluate how the different sources are used in pattern recognition and/or learning tasks. We could develop a task in which children have repeated learning trials of a set of pseudowords. The goal would be to simultaneously vary context, symbol, and association. Context could differ in terms of the stimulus transparency, symbol might differ in terms of its articulation difficulty, and iconicity could be varied to influence the association between context and symbol.

6. Discussion

In addition to the variables studied in this paper, there are a number of other variables that influence word learning, and these variables may also be primarily influential at the context symbol, and association level. The benefits of the speech embellishments inherent to so-called motherese are well-documented [24]. We expect mothese and other prosodic effects to have their influence primarily at the symbol learning level.

Children tend to associate the name of a novel word with the shape of an object rather than any of its parts or some other aspect of the object's context. We expect that this shape bias would modulate the context-learning component of word learning.

Children are also very adept at using mutual exclusivity to facilitate their word learning [26]. If a child experiences an unknown word when confronted with a number of familiar objects whose names are known and one unfamiliar object whose name is unknown, then the child will associate the new name with the new object. This result can be accounted for by the relative goodness implementation in Equation 3: there would be very weak support for the known alternatives, leaving the best support for the new word alternative.

Referential intention is another source of information that aids children word disambiguation and learning [27]. We expect referential intention to be a source of information that would influence context or stimulus learning..

Systematicity can be defined as a statistical regularity between sounds and meanings. Systematicity is measured as a correlation between form similarity and meaning similarity – that is, the degree to which words with similar meanings have similar forms [6] and [28]. If a child already has already learned one form of a word, then it should facilitate learning of another. The new related word might facilitate disambiguating the context, mastering the word, and associating the context with the word.

The FLMP has not been quantitatively tested and, thus, might not warrant joining the growing family of computational models of word disambiguation learning [29] and [30]. Mimicking Frank {15], however, the FLMP "may be the glue which holds these disparate kinds of information together and allows them to be used together in the service of learning words." [15].

A challenge to this cue integration perspective has been recently provided by Yurovsky and Frank [31]. They monitored looking behavior of young children in a word learning experiment that varied both the social cue of directed gaze and the salience of the object whose name was being learned. They proposed that the cue-combination predicts that perceptual salience and social information should have more weight early and later in development, respectively. However, both of these two variables had significant influences for both younger and older children. In our taxonomy, the perceptual cue of salience would influence stimulus learning whereas the social cue of looking at the object would influence learning the association between the word and the object. There is no a priori reason for children of different ages to be differentially influenced by these two cues, other than due to concomitant changes in perception, attention, and memory development.

7. Acknowledgements

The author acknowledges the helpful discussions with Bill Rowe for embellishing the theoretical illustrations and for his guidance in optimizing the signal in this message.

8. References

[1] D. W. Massaro (1987). Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry. Hillsdale, NJ: Erlbaum Associates.

[2] J. Movellan and J. L. McClelland, (2001). The Morton–Massaro law of information integration: Implications for models of perception. Psychological Review, 108, 113-148.

[3] [K. Hirsh-Pasek, R. M. Golinkoff, and G. Hollich (2000). An emergentist coalition model for word learning: Mapping words to objects is a product of the interaction of multiple cues. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. Smith, A. Woodward, N. Akhtar, M. Tomasello, & G. Hollich (Eds.), Becoming a word learner: A debate on lexical acquisition (pp. 136-164). New York, NY: Oxford University Press.

[4] L. K. Perry, M. Perlman, and G. Lupyan, (2015). Iconicity in English and Spanish and Its Relation to Lexical Category and Age of Acquisition. PLoS ONE, 10(9), e0137147

[5] D. W. Massaro (2016). Multiple influences in vocabulary acquisition: Parental input dominates. Proceedings of the 17th Annual Conference of the International Speech Communication Association (Interspeech 2016), pp. 878-882. ISSN 2308-457X. www.isca-speech/archive/interspeech_2016/pdfs/0037.PDFC. Fennell and S. R. Waxman (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. Child Development, 81(5), 1376–1383.

[6] D. W. Massaro and M. Perlman (2017). Quantifying Iconicity's Contribution During Language Acquisition: Implications for Vocabulary Learning. Frontiers

[7] D. W. Massaro (1998). Perceiving Talking Faces: From speech perception to a behavioral principle. Cambridge, MA: MIT Press.

[8] D. W. Massaro and D. G. Stork, D.G. (1998). Sensory integration and speechreading by humans and machines. American Scientist, 86, 236-244.

[9] D. W. Massaro and B. Rowe (2015). Comprehension outscores production in language acquisition: Implications for Theories of Vocabulary Learning. Journal of Child Language Acquisition and Development – JCLAD, Vol: 3 Issue: 3 121-152, 2015, September ISSN: 2148-1997.

[10] L. K. Perry, M. Perlman, and G. Lupyan, (2015). Iconicity in English and Spanish and Its Relation to Lexical Category and Age of Acquisition. PLoS ONE, 10(9), e0137147.

[11] Perlman, M. & Cain, A. A. (2014). Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. Gesture, Vol. 14:3, pp. 320–35.L. J. Gogate and G. Hollich (2010). Invariance detection within an interactive system: a perceptual gateway to language development. Psychol. Rev. 117:496–516

[12] Brandon C. Roy, Michael C. Frank, Philip DeCamp, Matthew Miller and Deb Roy. (2015). Predicting the Birth of a Spoken Word. Proceedings of the National Academy of Sciences of the United States of America (PNAS). http://www.pnas.org/content/early/2015/09/15/1419773112G.

[13] D. W. Massaro (1998). Perceiving talking faces: From speech perception to a behavioral principle. Cambridge, Massachusetts: MIT Press.

[14] L. A. Thompson, and D. W. Massaro (1986). Evaluation and integration of speech and pointing gestures during referential understanding. Journal of Experimental Child Psychology, 42, 144-168.

[15] M. C. Frank (unpublished). Computational models of early language acquisition.

[16] L. A. Zadeh (1965). Fuzzy Sets. Information and Control, 8, 338-353.

[17] S. N. Kitzis, H. Kelley, E. Berg, E., D. W. Massaro, D.W., and D. Friedman (1998). Broadening the tests of learning models. Journal of Mathematical Psychology, 42, 327-355.

[18] B. McMurray, J. Horst, J., and L. Samuelson, (2012) Word learning emerges from the interaction of online referent selection and slow aasocialive learning. Psychological Review, 119, 831-877.

[19] Childes (2015). http://childes.psy.cmu.edu/

[20] Baath, R. (2014). ChildFreq: An online tool to explore word frequencies in child language. http://childes.psy.cmu.edu/topics/childfreq.pdf

[21] ChildFreq (2015). http://childfreq.sumsar.net/

[22] Brysbaert, M., Warriner, A.B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. Behavior Research Methods, 46, 904-911.

[23] http://childes.psy.cmu.edu/derived/parentfreq.cdc

[24] A. Weisleder and A. Fernald (2013). Talking to Children Matters. Early Language Experience Strengthens Processing and Builds Vocabulary. Psychological Science, November 2013 vol. 24, no. 11, 2143-2152

[25] S. S. Jones, L. Smith, L., and B. Landau (1991). Object properties and knowledge in early lexical learning. Child development, 62 (3), 499–516.

[26] E. M. Markman (1989). Categorization and naming in children: Problems of induction. Cambridge, MA: MIT Press.

[27] D. Baldwin (1993). Early referential understanding: Infants' ability to recognize acts for what they are. Developmental Psychology, 29, 832–843.

[28] P. Monaghan, R. C. Shillcock, M. H. Christiansen, and S. Kirby (2014). How arbitrary is language? Philosophical Transactions of the Royal Society B, 369 20130299. DOI: 10.1098/rstb.2013.0299

[29] M. C. Frank, N. D. Goodman, and J. B. Tenenbaum (2009). Using speakers' referential intentions to model early cross-situational word learning. Psychological Science 20, 578-585.

[30] F. Xu and J. B. Tenenbaum (2007). Word learning as Bayesian inference. Psychological Review 114, 245-272.

[31] D. Yurovsky, D., and M. C. Frank (2015). Beyond naïve cue combination: Salience and social cues in early word learning. Developmental Science, 1-17.