

Précis of *The Birth of a Word*

Brandon C. Roy

January 15, 2014

Introduction

Sometime around their first birthday, children typically produce their first word, adding hundreds more to their vocabulary during their second year (Fenson et al., 1994). Learning words is a core task in the larger enterprise of acquiring one’s native language. Researchers often refer to this overall enterprise by the theoretically neutral term “language acquisition”, since there is much debate about where maturational processes end and learning begins. But there is little debate that words themselves are learned through experience. How is this accomplished? What is the trajectory of vocabulary development, and how do different environmental factors contribute to word learning?

We approach these questions with a particular theoretical interest in how the structure of a child’s everyday experience contributes to word learning. The primary scientific goal of this work is to study how the rich social interaction and activity structures of daily life ground everyday linguistic experience and support learning. We hypothesize that these structures provide useful learning constraints, and that more contextually constrained words will be more readily learned. More broadly, we aim to characterize the *environmental contributions* to the learning process, since learning requires not only a learner but also learnable structure. We seek to characterize this learnable structure both in the speech a child hears as well as the broader context in which that speech occurs, and to examine the relationships to word learning.

To pursue this scientific goal, we embarked on a naturalistic, observational study of one child’s early word learning, collecting and analyzing data about both the learning environment and what was learned. Our data consist of dense audio and video recordings of the home of a family with a young child, which we collected through the Human Speechome Project (Roy et al., 2006). Using a custom audio-video recording system embedded in the ceilings of every room in the family’s home, we initiated recording at the birth of the child and recorded roughly 10 hours per day until the child’s third birthday, collecting an unprecedented amount of data of one child’s experience and development. As with all observational case studies, there are numerous challenges in working with data collected “in the wild”, but to effectively leverage data at this scale required new methodological solutions.

Therefore, in addition to our scientific goal of studying the environmental contributions to word learning, our second goal is to develop methods for annotation and analysis of large-scale observational records. The challenges in working with this data are manifold. Audio and video are not easy to analyze directly but must be annotated, yet fully manual annotation is often exceedingly time consuming and expensive. We develop semi- and fully-automatic methods for general tasks such as speech transcription and speaker identification, and explore computational approaches to tasks such as identifying and labeling human activities.

With the combination of dense, naturalistic recordings and new methods for large-scale data annotation, we organized our analyses around predicting “word births”, a term we coined to identify

the first consistent use of a word by the child. The timeline of word births reveals interesting vocabulary growth dynamics, and using word births as an outcome measure permits evaluating different environmental factors on learning. Comparing the predictive strength of different factors may provide clues into the operation of learning mechanisms.

Thesis summary

Chapter 2: Continuous Naturalistic Recording of Life at Home

Word learning has long been a subject of speculation and investigation. St. Augustine (1961) provides an early and oft cited account, in which he “recalls” learning words by accumulating exposures to words and linking them to their referents using gesture, glances, tones of the voice and other social and contextual indicators of the speaker’s intent. Wittgenstein (2009) used St. Augustine’s account to motivate his philosophical inquiry into how language and meaning are linked through actions and situational context.

These accounts reflect a perspective of word learning and language that ties together communicative inference, action and situational context. This dissertation brings this perspective to a scientific inquiry into word learning by studying language use and learning in the context of everyday life. We describe a continuous thread of research on the Human Speechome Project (Roy et al., 2006), a study of early language development through dense, naturalistic, longitudinal recordings of a child’s early life. Using a custom recording system consisting of 11 video cameras and 14 microphones embedded in the ceilings of the child’s home, and with careful privacy safeguards in place, the family initiated recording at the child’s birth, recorded for an average of 10 hours per day, and concluded after the child’s third birthday. Altogether, roughly 90,000 hours of video and 120,000 hours of audio were recorded. From a technical standpoint, the recording system and tools for monitoring and visualizing multiple streams of data presented a significant engineering challenge of interest to researchers outside of cognitive science. From a scientific standpoint, this massive audio-video dataset contains not only a detailed record of linguistic development, but also the linguistic input from caregivers and the social, physical and situational context of the child’s early experience. But while the scale of this dataset is unprecedented, naturalistic observational case studies have a long and fruitful history in developmental psychology.

A landmark observational case study of first language acquisition was led by Roger Brown (1973), who followed the language development of three children in their homes using portable tape recorders to collect samples of their speech. Dan Slobin recalled his wonder at first hearing these extensive, unedited audio samples and the richness of the data contained therein (Slobin, 1988). Later work by Bloom (1973), Nelson (1973), Bowerman (1978), Braunwald (1978), and Dromi (1987), to name only a few, focused on children’s early word use and communicative development with less emphasis on syntax. In many cases, the subjects of these studies included the researcher’s children up to about age two.

These samples of naturalistic, observational studies have important methodological similarities. Data collection often takes place in the comfortable, familiar environment of children’s homes and are generally not controlled experiments. Careful experimental work, with the ability to control and manipulate variables of interest, has contributed much to what we know about the mechanisms of early word learning. Yet children do not normally learn in a laboratory but rather the “clutter of life at home” (Bruner, 1983). Naturalistic, observational case studies have provided a complementary approach to studying children’s language acquisition. By observing and characterizing how children are exposed to language in context, and by tracking language outcomes, we can explore the basic relationship between a learner and his environment. McCall (1977) and Bronfenbrenner (1979)

make a strong case for the value of naturalistic data analysis to developmental psychology.

Much of the work in this dissertation, specifically, the ways in which nonlinguistic contextual factors are studied in relation to word learning, draws upon the ideas of Jerome Bruner (1983). Motivated by a social interactionist viewpoint and drawing from ideas such as Vygotsky’s zone of proximal development (Vygotsky, 1986), he argued that the games and routines of everyday life provide a “scaffold” into language. These stable interaction formats provided a structure for communication in which the prelinguistic child could participate with his caregivers, supported by the rich context of everyday life. Operationalizing some of these ideas and quantifying their effects is a substantial part of the work in this dissertation.

Chapter 3: Efficient Transcription of the Speechome Corpus

Chapter 3 describes our methods for data annotation, particularly, speech transcription. We chose to focus our annotation and analysis efforts on the child’s 9–24 month age range, a time period that typically includes both first words and the emergence of combinatorial speech. At the outset, we set a goal of transcribing as much of the recorded data as possible for this target age range. Though it predates the term, this work was very much an experiment in how “big data” approaches can contribute to developmental psychology, and what the picture of word learning looks like when densely sampled. These goals demanded new annotation approaches, and a number of tools were built and deployed for this research. The central pillar of all our annotation work was *BlitzScribe* (Roy and Roy, 2009), a semi-automatic speech transcription tool. *BlitzScribe* was designed to combine the complementary strengths of human and machine speech processing. In contrast to more typically used manual speech transcription tools such as CLAN (MacWhinney, 2000), which would be prohibitively time consuming to use at this scale (Reidsma et al., 2005; Tomasello and Stahl, 2004; Barras et al., 2001), *BlitzScribe* employs custom speech processing algorithms to identify and segment speech, presenting the resultant segments in a simple user interface for human annotators to transcribe. Although usable transcription accuracy from fully automatic speech recognition is still beyond the state of the art, our combination of automatic speech segmentation and human transcription yields significant performance gains at comparable accuracy to manual approaches. Most notably, *BlitzScribe* is roughly 5 times faster than several commonly used manual tools (Roy and Roy, 2009). In this chapter, we describe some of the details of the speech detection and segmentation system, sketching how multichannel audio is processed, the acoustic features extracted, and the classification algorithm. We also describe a separate, fully automatic system for speaker identification that is built using similar technology. We review the types of errors these systems make and characterize their performance.

Annotating the Speechome data was accomplished through the hard work of many human annotators. Although *BlitzScribe* was the central annotation tool, other custom annotation tools were also used. For example, TotalRecall (Kubat et al., 2007) was used to annotate video data in order to track the baby’s location over time and whether he was awake or asleep. These annotations were used to ensure that only “child available” adult speech would be transcribed. Child available speech, or speech that the child could have been exposed to, can be unambiguously defined and serves as a reasonable proxy for child directed speech (Vosoughi and Roy, 2012). A separate speaker ID annotation tool was developed and used to construct training and test sets for the fully automatic speaker ID system.

Leading a team in a 5 year long annotation effort also required tools for managing workflow, tracking quality, sharing information and maintaining annotation consistency. A set of management tools were built to help with these challenges. Our team ranged from 2 to 15 annotators at a time, with nearly 70 annotators over the 5 year period of active transcription (from 2007–2012). Even

basic tasks such as developing, sharing, and updating transcription conventions, user manuals and other materials depended on a streamlined process, largely supported by a private, collaborative Google Site. Transcribers, often undergraduates interested in cognitive science from MIT, Wellesley and other nearby schools, contributed to the project through their annotation work but also through discussions on their experience working directly with the data.

This chapter ends by summarizing the transcribed 9–24 month age range used in our study, which we refer to as the Speechome Corpus. Roughly 86% of the recorded 9–24 month period had been transcribed, yielding approximately 8 million tokens of child and child-available speech across approximately 2 million utterances, divided primarily among the four primary speakers: mother, father, nanny and child. The transcribed Speechome Corpus served as the foundation for our later analyses.

Chapter 4: The Child’s Productive Vocabulary

The Speechome Corpus captures a substantial portion of all vocalizations and speech produced by the child during a critical period in his early development. What words did the child know, and when did he first say them? The goal in this chapter is to answer these questions.

We begin by presenting a selection of prior word learning studies along with findings on children’s vocabulary size, growth rate and composition. Fenson et al. (1994) provided a wealth of information derived from many children on both productive and receptive vocabulary growth, while diary studies have often yielded more detailed accounts of the idiosyncrasies of children’s word use. In the Speechome Corpus, as in other case studies based on transcript data, assessing receptive vocabulary is particularly challenging. Our emphasis is the child’s productive word use, using similar criteria to (Dromi, 1987) when transcribing words in child speech.

We coin the term “word birth” to refer to the first (consistent) production of a word by the child; the sequence of word births describes both the child’s productive vocabulary as well as the timeline of growth. The notion of a word birth also suggests that words are not simply *acquired*, but rather have a gestation period prior to their first use. Characterizing a word’s gestation period is the subject of subsequent chapters. But first, detecting true word births in the sea of transcript data presents a significant challenge in its own right. Simply choosing the subset of transcripts attributed to the child and identifying the first use of each word leads to many spurious word births, since automatic speaker ID is imperfect and even human generated transcripts and speaker annotations are error prone. For example, a single adult utterance mislabeled as child speech (perhaps due to the child crying while the adult is speaking) could falsely attribute multiple words to the child’s productive lexicon if using this method. Instead, we present a series of statistical inference approaches to identify likely first word productions. These all incorporate, in some fashion, the false positive rate of the speaker ID in modeling child token counts. The best method models observations as arising from pre- and post-word birth regimes. Yet despite the efficacy of this approach, it is still crucial to our later analyses to have a high quality word birth timeline. Toward this end, a word birth annotation tool was developed in Java and several annotators checked (and revised) word births by hand.

The resultant vocabulary growth rate curve had a striking “shark’s fin” shape – a period of accelerating growth up to 18 months of age, followed by a precipitous drop in the number new word births per month. Although there has been much discussion of vocabulary growth accelerating (e.g. (Gopnik and Meltzoff, 1987; Goldfield and Reznick, 1990; Ganger and Brent, 2004; Bloom, 2000)), there is less discussion of decreasing growth rates. Dromi (1987) presents a similar growth rate curve and argues for a change in learning strategy; Clark (1995) sees a different growth trajectory in her subject, while Fenson et al. (1994) point out that averaging different growth functions may obscure

non-monotonic behavior. How shall we interpret the vocabulary growth curve? This chapter closes by exploring two ideas. We explore and ultimately reject a null model in which the observed curve results from an interaction between a steady or increasing “true” vocabulary and the decreased chance of observing any particular word as vocabulary size increases. The second exploration seems more promising, revealing that the number of new word *combinations* shows a marked increase just as vocabulary growth slows down. Combined with other results on the child’s average utterance length over time (Roy et al., 2009), this points toward a change in production style and complexity. The dynamics of vocabulary growth are complex and part of a larger developmental story, but how do basic external factors contribute to learning a word?

Chapter 5: Environmental Contributions to Word Learning

Children are remarkably robust learners, acquiring their native languages under a wide range of different circumstances. Yet circumstances do matter, from the most extreme cases of minimal linguistic input (Comrie, 2000; Windsor et al., 2007) to the variations of input even in “normal” situations (Hart and Risley, 1995). Greater linguistic input contributes not only to larger vocabularies, but also greater processing efficiency (Hurtado et al., 2008).

But how does exposure to speech contribute to the uptake of *specific* words? We perform a series of regression analyses to study the relationship between variables characterizing caregiver word use and age of acquisition (AoA) of words in the child’s productive lexicon (i.e. word birth dates). The first such analysis considers the log usage rate of a word in caregiver speech prior to acquisition. Words with higher usage frequency are generally learned earlier ($r = -.19$), particularly for nouns, consistent with earlier research (Huttenlocher et al., 1991; Goodman et al., 2008). This finding supports an account in which each exposure to a word provides a new learning opportunity, and is consistent with both statistical word learning models (e.g. (Yu and Smith, 2007; Smith and Yu, 2008)) and hypothesis-testing models (Medina et al., 2011). Word frequency, however, measures word usage rates over months rather than minutes, which is closer to the timescales of both speech patterns and a child’s short-term memory and attentional limitations (Roy, 1999; Roy and Pentland, 2002). We thus measured the *recurrence* of each word in caregiver speech prior to the word birth, a variable that captures a word’s usage rate in short temporal windows where the word occurs at least once (Vosoughi, 2010; Vosoughi et al., 2010). Recurrence proves to be a better predictor of AoA than frequency, with higher recurrence words learned earlier ($r = -.30$).

Both frequency and recurrence are directly measurable from caregiver speech, but the child’s experience with language is linked to his experience in other modalities. Building on the work of Miller (2011), we characterize the spatial aspect of caregiver speech in the home for the words learned by the child. This is accomplished by processing video into low-dimensional motion activity vectors that indicate where there is motion during an utterance and aggregating over the appropriate subset of utterances (i.e. those utterances containing a target word prior to AoA). A frequency corrected version of KL-divergence (Cover and Thomas, 2006) is used to compare a word’s spatial distribution to the spatial distribution of overall language use, which summarizes the “distinctiveness” of a word’s spatial usage patterns. Words such as “mango” are highly spatially distinct (mostly occurring in the kitchen), while words such as “the” are more spread out. What is perhaps surprising is that this spatial distinctiveness measure is highly predictive of AoA, even when controlling for frequency and part of speech. Words that are more distinct are generally learned earlier ($r = -.42$), in close agreement with (Miller, 2011) who used earlier, slightly different versions of the Speechome data. Ongoing work has shown that spatial distinctiveness is not simply a proxy for imageability, concreteness, or simply part of speech. An interpretation of this result is that spatially distinct words are more strongly grounded in the child’s physical environment and

easier to decode. Large spatial distinctiveness also implies that such words deviate from the average and may thus be more salient. Another interpretation is that spatial distinctiveness serves as a proxy measure for a word’s link to particular activities, which may themselves be spatially localized. If this interpretation holds merit, it may also apply to recurrence. Words may be salient in a particular activity and thus used more frequently during that activity, leading to a high recurrence value.

Chapter 6: Language Use in Activity Contexts

In chapter 6, the final analysis chapter in the dissertation, we focus on characterizing the activities in the child’s home and quantifying a word’s contextual boundedness across these activities. This follows up on the suggestions in the previous chapter and is a direct attempt to operationalize some aspects of Bruner’s language acquisition support system (Bruner, 1983). We begin by proposing *activity contexts*, which are labels for *what is happening* in the data at the temporal granularity of minutes as a first step down this path. For our purposes, mealtime, bath time, and reading books are all examples of coherent activity contexts.

We pursued two approaches to obtaining activity contexts. The first was a fully manual approach in which transcribers were asked to use a modified version of BlitzScribe to label the activities occurring in their transcription assignments (which were 15 minute chunks of data.) After roughly 10 months of annotation, 10% of the transcribed data had been annotated with activity labels. We then explored activities in terms of their temporal, spatial, speaker and word usage distributions. For example, the `going_to_sleep` activity was temporally distinct from the overall temporal distribution, with peaks after lunch and at night. The transcript words during episodes containing this activity were also distinct (e.g. “dream”, “sleep”). The `reading` activity was highly spatial, linked to particular speakers, and also exhibited specific words in the transcripts (e.g. “book”, “page”). In order to study the relationship to word learning, we then characterized each word’s pre-AoA distribution across activities relative to the baseline activity distribution, analogous to the method used for the spatial distinctiveness predictor. We found that this measure of a word’s activity distribution was predictive of AoA ($r = -.34$), with words more focused in fewer activities (and with greater distinctiveness) learned earlier.

In the spirit of this large-scale, data driven study of early word learning, we explored a second approach to identifying activity contexts. The basic idea was to see whether activities could be viewed as hidden variables to be discovered by latent variable methods. Although manually labeled activities exhibited structure in four modalities (temporal, spatial, speaker and word distributions), we focused only on using the transcripts to try to automatically infer latent activities. We employed a standard unsupervised topic modeling method, latent Dirichlet allocation (LDA) (Blei et al., 2003), examining whether the resultant topics reflected interpretable activities. Indeed, some topics (though not all) could be interpreted as activities such as `diaper_change`, `mealtime` and so on, whether through inspection of their distributions across modalities or by directly correlating topics with manually labeled activities. Using the resultant topics as proxies for activities, we calculated each word’s activity context distinctiveness in the same manner as for manually labeled activities. This distinctiveness measure strongly correlates with AoA ($r = -.37$). Earlier work using word-topic entropy yielded similar findings (Roy et al., 2012); that research and ongoing work show that this result is robust across word classes and holds in combined regression models with other non-contextual predictors such as word frequency.

Our interpretation of these results follows a similar storyline to that of the spatial context predictor, and indeed these results seem to be convergent. The child’s exposure to words in caregiver speech is embedded in the larger physical and activity context of everyday experience. Words that

are strongly linked to spatial contexts or activity structures may be more accessible to the young learner and easier to decode. Words strongly tied to an activity may primarily co-occur with only a limited range of actions, objects and sensations providing more focused co-occurrence statistics for associative learning (Yu and Smith, 2007; Smith and Yu, 2008), or if the child is forming and testing hypotheses about word meanings (Medina et al., 2011; Trueswell et al., 2013), context-bound words may have a limited hypothesis space.

Conclusion

There are two parallel threads in this dissertation. The first is a study of one child’s early word learning and the contribution of natural environmental factors. The aim of this account is to examine the hypothesis that contextual constraints support word learning. It provides operational definitions of variables that could be used to investigate social interactionist theories of word learning. More generally, this work provides a descriptive account of lexical acquisition and the role played by different environmental factors. The second thread in this dissertation is methodological and more technical in nature, offering a variety of new tools and approaches that may be useful for other large-scale data annotation and analysis projects. This thread may be of interest to engineers as well as computer and data science researchers. Researchers in artificial intelligence and machine learning may be interested in how computational tools, such as LDA, can contribute to fields outside their original application domains. Whatever the case, we hope this work makes a meaningful contribution to our understanding of human language acquisition, and serves as a useful guide for future research.

References

- Barras, C., Geoffrois, E., Wu, Z., and Liberman, M. (2001). Transcriber: Development and use of a tool for assisting speech corpora production. *Speech Communication*, 33(1-2):5–22.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022.
- Bloom, L. (1973). *One word at a time: The use of single word utterances before syntax*, volume 154. Mouton, The Hague.
- Bloom, P. (2000). *How Children Learn the Meanings of Words*. The MIT Press.
- Bowerman, M. (1978). The acquisition of word meaning: An investigation into some current conflicts. *The development of communication*, pages 263–287.
- Braunwald, S. (1978). Context, word and meaning: Toward a communicational analysis of lexical acquisition. *Action, gesture and symbol: The emergence of language*, pages 485–527.
- Bronfenbrenner, U. (1979). *The Ecology of Human Development: Experiments by Nature and Design*. Harvard University Press.
- Brown, R. (1973). *A first language: The early stages*. Harvard University Press.
- Bruner, J. (1983). *Child’s Talk: Learning to Use Language*. WW Norton.
- Clark, E. (1995). *The lexicon in acquisition*, volume 65. Cambridge University Press.

- Comrie, B. (2000). From potential to realization: An episode in the origin of language. *Linguistics*, 38(5):989–1004.
- Cover, T. and Thomas, J. (2006). *Elements of information theory*. John Wiley & Sons.
- Dromi, E. (1987). *Early Lexical Development*. Cambridge University Press.
- Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D. J., Pethick, S., Tomasello, M., Mervis, C. B., and Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59:1–185.
- Ganger, J. and Brent, M. R. (2004). Reexamining the vocabulary spurt. *Developmental Psychology*, 40(4):621.
- Goldfield, B. A. and Reznick, S. J. (1990). Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of Child Language*, 17(1):171–183.
- Goodman, J., Dale, P., and Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, 35:515–531.
- Gopnik, A. and Meltzoff, A. (1987). The development of categorization in the second year and its relation to other cognitive and linguistic developments. *Child Development*, 58:1523–1531.
- Hart, B. and Risley, T. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Brookes Publishing Company, Baltimore, MD.
- Hurtado, N., Marchman, V., and Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, 11(6):F31–F39.
- Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., and Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, 27(1236-248).
- Kubat, R., DeCamp, P., Roy, B., and Roy, D. (2007). TotalRecall: Visualization and semi-automatic annotation of very large audio-visual corpora. In *ICMI*.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Lawrence Erlbaum Associates, third edition.
- McCall, R. B. (1977). Challenges to a science of developmental psychology. *Child Development*, 48(2):333–344.
- Medina, T., Snedeker, J., Trueswell, J., and Gleitman, L. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences*, 108(22):9014.
- Miller, M. (2011). Semantic spaces: Behavior, language and word learning in the human speechome corpus. Master’s thesis, Massachusetts Institute of Technology.
- Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, pages 1–135.
- Reidsma, D., Hofs, D., and Jovanović, N. (2005). Designing focused and efficient annotation tools. In *Measuring Behaviour, 5th International Conference on Methods and Techniques in Behavioral Research*, Wageningen, The Netherlands.

- Roy, B. C., Frank, M. C., and Roy, D. (2009). Exploring word learning in a high-density longitudinal corpus. In *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*.
- Roy, B. C., Frank, M. C., and Roy, D. (2012). Relating activity contexts to early word learning in dense longitudinal data. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*.
- Roy, B. C. and Roy, D. (2009). Fast transcription of unstructured audio recordings. In *Proceedings of Interspeech*, Brighton, England.
- Roy, D., Patel, R., DeCamp, P., Kubat, R., Fleischman, M., Roy, B., Mavridis, N., Tellex, S., Salata, A., Guinness, J., Levit, M., and Gorniak, P. (2006). The human speechome project. In *Proceedings of the 28th Annual Meeting of the Cognitive Science Society*, pages 2059–2064, Mahwah, NJ. Lawrence Erlbaum Associates.
- Roy, D. K. (1999). *Learning Words from Sights and Sounds: A Computational Model*. PhD thesis, Massachusetts Institute of Technology.
- Roy, D. K. and Pentland, A. P. (2002). Learning words from sights and sounds: a computational model. *Cognitive Science*, 26(1):113–146.
- Slobin, D. (1988). From the Garden of Eden to the Tower of Babel. In Kessel, F. S., editor, *The Development of Language and Language Researchers: Essays in Honor of Roger Brown*, pages 9–22. Lawrence Erlbaum Associates.
- Smith, L. and Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106(3):1558–1568.
- St. Augustine (398/1961). *The Confessions of St. Augustine*. Random House.
- Tomasello, M. and Stahl, D. (2004). Sampling children’s spontaneous speech: How much is enough? *Journal of Child Language*, 31(1):101–121.
- Trueswell, J., Medina, T., Hafri, A., and Gleitman, L. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, 66(1):126–156.
- Vosoughi, S. (2010). Interactions of caregiver speech and early word learning in the Speechome Corpus: Computational Explorations. Master’s thesis, Massachusetts Institute of Technology.
- Vosoughi, S., Roy, B. C., Frank, M. C., and Roy, D. (2010). Contributions of prosodic and distributional features of caregivers’ speech in early word learning. In *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*.
- Vosoughi, S. and Roy, D. (2012). An automatic child-directed speech detector for the study of child language development. In *Proceedings of Interspeech*, Portland, Oregon.
- Vygotsky, L. (1986). *Thought and Language*. The MIT Press.
- Windsor, J., Glaze, L., Koga, S., et al. (2007). Language acquisition with limited input: Romanian institution and foster care. *Journal of Speech, Language and Hearing Research*, 50(5):1365.
- Wittgenstein, L. (1953/2009). *Philosophical Investigations*. Wiley-Blackwell, 4th edition.
- Yu, C. and Smith, L. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, 18(5):414.