

# The Roots of the Early Vocabulary in Infants' Learning From Speech

Daniel Swingley

University of Pennsylvania

---

**ABSTRACT**—Psychologists have known for over 20 years that infants begin learning the speech-sound categories of their language during the first 12 months of life. This fact has dominated researchers' thinking about how language acquisition begins, although the relevance of this learning to the child's progress in language acquisition has never been clear. Recently, views of the role of infancy in language acquisition have begun to change, with a new focus on the development of the vocabulary. Infants' learning of speech-sound categories and infants' abilities to extract regularities in the speech stream allow learning of the auditory forms of many words. These word forms then become the foundation of the early vocabulary, support children's learning of the language's phonological system, and contribute to the discovery of grammar.

**KEYWORDS**—language development; language acquisition; infant learning; phonology

---

Infants begin life hearing plenty of speech but grasping none of the words. At first, their understanding of language is limited to the universal songs of emotion in speech, like the graceful tones mothers produce to soothe (Fernald, 1992). To make progress in learning their own language, whether that be English or Ewe, infants must go beyond the broad melodies of speech and break down spoken language into its component parts, like the words that make up sentences and the consonants and vowels that make up words. A remarkable advance of 20th-century psychology was the discovery that, even before their first birthday, infants make considerable headway in one part of this analysis, the learning of their language's speech sounds. Some speech-sound differences infants can initially tell apart become indistinguishable if they are not used in the parents' language, while

infants' perception of sounds used in the parental language is enhanced (e.g., Kuhl et al., 2008). For example, 6-month-old English learners can discriminate similar-sounding Hindi consonants not found in English, but lose this ability by 12 months (Werker & Tees, 1984). Such changes reflect infants' ability to categorize speech and are viewed as beneficial, because they suggest that children are focusing on just the distinctions they need to distinguish words in their language.

Infants' discovery of phonetic (speech-sound) categories is impressive, particularly given the poor performance of computer systems designed to interpret spoken language. Demonstrations of infants' perceptual talents have been interpreted as showing that infants are perfectly adapted listeners, biologically predisposed to interpret the sounds of human languages (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). However, many have doubted the relevance of infant perception studies to the broader course of language acquisition. Once children start to talk, their pronunciation is often variable and inaccurate, and children do not always interpret speech the same way adults do (Nittrouer, 1996). While infants show excellent categorization skills in simplified laboratory contexts, these skills do not necessarily ensure accurate word learning in childhood.

To understand how infants' perceptual learning is related to vocabulary development, researchers needed to examine infants' and toddlers' linguistic knowledge more broadly—not just children's perception of isolated syllables but their perception and interpretation of words in context. Recent work along these lines has suggested a new view of the role of the infancy period in language learning in which infants learn not only speech sounds but also the auditory forms of words. These word forms help children build their vocabulary and support their discovery of the grammar of their language.

## THE ORIGINS OF THE CHILD'S VOCABULARY

Once researchers began studying infants' perception of words and sentences, it became clear that children start learning the

---

Address correspondence to Daniel Swingley, Department of Psychology, University of Pennsylvania, 3401 Walnut St. 302C, Philadelphia PA 19104; e-mail: swingley@psych.upenn.edu.

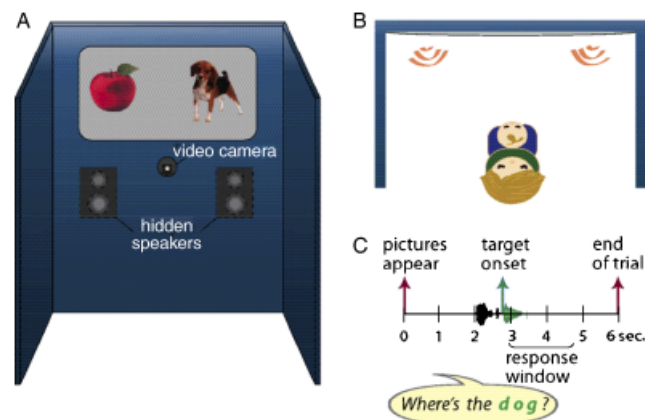
phonological (sound) forms of whole words by 8 months or even earlier—words like *baby*, *give*, and *little*, which may not mean anything to the infant but which are nevertheless stored in memory as familiar, recognizable bits of language (Jusczyk & Hohne, 1997). Early word-finding is an accomplishment partly because parents do not provide clear indications of word boundaries. How can infants determine where one word ends and another begins? We now know that infants have several tools that help in this, including attention to intonational changes like the pitch movements marking ends of clauses (Seidl, 2007). Infants also group together frequent syllables that usually occur together (so, *dar + ling* is a possible word because *dar* usually precedes *ling*, and *ling* usually follows *dar*). Tendencies like these yield a nascent lexicon of phonological forms. This “protollexicon,” in turn, exemplifies typical phonological properties of words in the language—properties that infants exploit to find yet more words (Thiessen & Saffran, 2003). For example, in English, pairs of syllables that tend to co-occur are predominantly *trochaic* (i.e., having first-syllable stress; Swingley, 2005a). Upon detecting this property of English words, 8-month-olds begin dividing continuous speech in a usefully biased way, picking out trochaic sequences as words and not stress-final sequences that are less likely to be words (Jusczyk, Houston, & Newsome, 1999).

Not all of the chunks of speech infants discover are actual words, but many are, so when children say their first word at around 12 months, they may already know the phonological forms of hundreds of other words. Many of these recognizable forms probably have little meaning to the child and wouldn't conventionally be considered true words. But they nevertheless aid vocabulary development. For example, if 1.5-year-olds are familiar with a word's form, they are better at learning what it means and better at differentiating it from similar-sounding words (Graf-Estes, Evans, Alibali, & Saffran, 2007; Swingley, 2007).

Thus, each word in the toddler's vocabulary has a developmental history that often begins with auditory learning in infancy. Word forms learned by infants are elaborated through the gradual addition of meaningful semantic and grammatical information, thereby becoming true words in the child's vocabulary. Although word learning is usually conceptualized (and studied) as a single event in which a novel word and novel concept are linked together, this one-trial learning or “fast mapping” is not characteristic of the young toddler's vocabulary.

### KNOWLEDGE OF SOUNDS IN WORDS

When infants learn the phonological form of a word, what do they retain in memory? Variability in children's own pronunciation of words might reveal incomplete knowledge (Ferguson & Farwell, 1975). If a child's realization of the word *baby* sounds like “ga,” it is hard to credit the child with an accurate memory of the word.



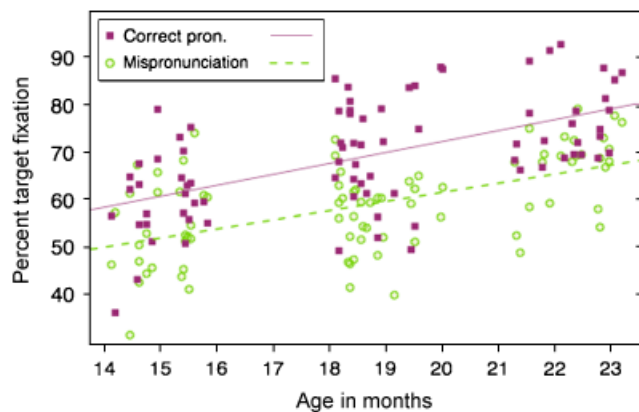
**Fig. 1.** Evaluation of word recognition using eye movements. The left figure (A) depicts the testing booth, with a trial in progress; the upper right figure (B) presents a bird's eye view showing the child on the parent's lap in the booth; and the graph (C) shows the timing of a single trial. Frequently, target fixation proportions are computed using only the data from a short *response window* immediately following the onset of the spoken target word, as shown.

But speech may be limited by articulatory abilities more than by perception.

Because of this problem of interpretation, decades of research on children's speech provided no satisfactory answer to the simple question of whether children's knowledge of words was phonetically correct or ill-formed and vague. This finally changed with the advent of experimental methods yielding fine-grained measures of word recognition, even in children saying few words. One effective method has been eye tracking, sometimes called “looking while listening” or “language-guided looking.” On each of a series of trials, children see two pictures, such as one of an apple and one of a dog (Fig. 1). One object is named in a sentence (e.g., “Where's the dog?”), and children's eye movements are recorded. Children initially looking at the apple should, when hearing “dog,” look away from the apple and toward the dog, while a child already fixating the dog should persist. Performance is graded, not all-or-none. Two children who look at the dog above chance levels may differ in how quickly or reliably they respond. This variation in performance is correlated with age and vocabulary size.

To test details of children's phonological knowledge of words, researchers compare responses to experimenters' correct pronunciations and mispronunciations. Children who know that *dog* begins with the “d” sound should recognize the word most readily when it is pronounced correctly (starting with “d”) and less readily if the word is mispronounced (starting with a similar sound, like “t”). But if children are unsure of the details of speech sounds in words, changing one sound to a similar one should not impair recognition.

The first study to test children's phonological knowledge this way examined 1-year-olds ranging from 18 to 23 months (Swingley & Aslin, 2000). The six test words (e.g., *apple*, *baby*,



**Fig. 2.** Children's percentage of fixation (time spent looking) to the picture named in stimulus sentences when the target object was pronounced correctly (e.g., "dog"; filled red squares) versus when it was mispronounced (e.g., "tog"; green circles). Age in months is plotted on the *x*-axis. The lines drawn through the plot show a linear regression of fixation percentage on age for correct pronunciations (solid line) and mispronunciations (dashed line). Fixation performance improved with age, but the difference between correct-pronunciation and mispronunciation trials did not change with age.

*dog*), were either pronounced canonically or mispronounced (e.g., "opple," "vaby," "tog"—acoustically subtle but linguistically meaningful changes). Analysis of children's eye movements showed that fixation to the named (target) picture was substantially reduced when the word was mispronounced (average 61%) relative to when it was correctly pronounced (73%). Children could recognize the mispronunciations, as shown by their above-chance fixation, but children were slower to fixate the target and less likely to maintain fixation on it when the word was mispronounced. A follow-up study retested these mispronunciations (and others) on 14- and 15-month-olds (Swingley & Aslin, 2002). The results from both studies, for the same set of mispronunciations, are plotted in Figure 2.

Subsequent experiments have replicated these results with other mispronounced speech sounds, including word-medial consonants, word-final consonants, and vowels. Beyond demonstrating children's encoding of phonological details in words, this work has repeatedly shown that the effects of mispronunciation do not depend on children's age or on one measure of linguistic sophistication, vocabulary size. One might imagine that more linguistically advanced children would have both larger vocabularies and more fine-tuned speech-processing skills. Some theories propose that children refine their phonological representations of words only when they learn many words that sound similar to one another. But the lack of correlation between vocabulary size and children's sensitivity to mispronunciation is inconsistent with such theories. Toddlers who do not know any words that sound like *apple* or *baby* (apart from those words) still have trouble recognizing slight mispronunciations of them; children who do not know any words differentiated by "b" and "v" still consider "vaby" a poor instance of *baby*. Thus, children's encoding of how words sound

does not appear to depend upon explicit comparison of similar-sounding words like *bear* and *pear*.

At what age do children start encoding familiar words with this level of phonological detail? Several studies using other methods indicate better recognition of canonical pronunciations of words than of mispronunciations, including one striking demonstration featuring 6-month-olds (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005). At 11 months, infants would rather listen to good pronunciations of familiar words than to unfamiliar or made-up words, but this preference disappears if the familiar words are mispronounced (Swingley, 2005b). Thus, for at least some words, children's knowledge of phonological form appears to be accurate, as early as we can measure it.

### PHONOLOGICAL INTERPRETATION

This does not mean that children have the phonology of adults. In linguistics, *phonology* refers to the sound structure of language as a system, not simply accurate memory for spoken words. An essential linguistic generalization concerning phonological systems is *contrast*: two utterances having different phonological descriptions convey different meanings ("dog" is not *bog*); and, excepting homonyms, two messages having the same phonological description are linguistically the same, conveying the same meaning ("dog" is *dog* whether declaimed or muttered). The developmental course of this principle of phonological contrast has not been elucidated, though some revealing trends are emerging.

One trend is that there are changes in the way infants handle variation in the realizations of words. At 7.5 months old, infants respond differently to words they have just heard several times than they do to nonfamiliarized words, an effect that enables us to evaluate when infants consider two instances of words to be the same. Infants familiarized to "cup" in a story then prefer hearing "cup" over "feet," showing retention of "cup." This preference is not found in infants familiarized to a "near miss" of "cup," "tup." One might consider this evidence that preverbal infants recognize the phonological principle: "cup" and "tup" count as different. But subsequent research revealed similar effects of *nonphonological* changes in words. For example, 7.5-month-olds do not show recognition of the same word spoken by a man and a woman, nor of the same word uttered with joyful and neutral intonation, even though the words consist of the same sequences of consonants and vowels (Singh, Morgan, & White, 2004). By 10.5 months, however, infants recognize words despite changes in talker and intonation. This suggests development in how much infants consider phonological and nonphonological variation to be relevant for matching words.

The shift in infants' interpretation may result from lexical development. The many word forms that infants learn provide a database over which they can draw phonological generalizations. By 10 months, infants may recognize that the word forms familiar from one talker (like their father) are virtually identical

to those from another (their mother), leading infants to conclude that each talker's vocal characteristics are not properties of the words themselves. This kind of abstraction process is crucial for learning the grammar of language and may operate at the phonological level even in infancy.

The functional importance of this development concerns children's interpretation of the meanings of words and sentences. Understanding language depends on using language's conventions, including its phonological properties. Recent research looking specifically at phonological interpretation has focused on 18-month-olds, many of whom are just beginning to learn the meanings of many words. This work shows 18-month-olds to be in transition, interpreting phonological variation appropriately in some contexts but not others. For example, although children fixate a canine more upon hearing "dog" than upon hearing "tog," the phonetic similarity of these words still hinders learning of *tog* as a meaningful word for a novel object, probably because *dog* is so well entrenched (Swingley & Aslin, 2007). When two similar-sounding words are equally unfamiliar to children, they can learn both. A recent study used this ability to test children's understanding of their language's phonological characteristics. Dietrich, Swingley, and Werker (2007) found that Dutch 18-month-olds readily learned the novel words *tam* and *taam*, which varied only in vowel duration; but English-learning 18-month-olds did not think these words were different. This pattern is explained by phonological properties of the languages: Dutch has vowel pairs distinguished largely by duration, as in *man* (man) versus *maan* (moon), whereas in English, parents may make vowels very long ("That's a kiiiity!") without making the vowel ambiguous. Though children learning each language can hear the difference between long and short vowels, they interpret this difference according to the rules of the language while learning words.

## CONCLUSIONS

How has this new work changed the study of language acquisition? First, the fact that infants begin learning words months before their first birthday makes children's rapid progress in language acquisition less mysterious. Children amaze parents and researchers alike when their first spoken words are so quickly followed by many others, and when single words become multiword sentences. Children can do this because of "underground" learning in infancy that is not expressed in day-to-day behavior but that can be detected using laboratory tests. Infants learn the forms of many words and phrases, apparently with substantial phonological accuracy, and they gather information about the linguistic and situational contexts in which these forms are used. This knowledge provides the foundation of toddlers' vocabularies. Toddlers build upon this foundation when they learn more about what words mean and when they use their phonological knowledge to recognize familiar words and identify novel ones.

Second, new techniques that have enabled us to broadly characterize the beginnings of language acquisition offer considerable promise for clinical assessment and the study of individual variation. Although audiologists have evaluated infant hearing for many years using discrimination methods, it is only now that we have procedures to reliably gauge 1- and 2-year-olds' *interpretation* of speech sounds and words. Indeed, recent studies show sizeable correlations between young children's performance in word recognition and later language achievement (Marchman & Fernald, 2008). Testing very young children's ability to interpret spoken language, whether by identifying novel words as novel or by comprehending sentences, may prove a more sensitive predictor of children's language outcomes than simpler tests of speech-sound categorization. This is an important focus of current research.

A great deal remains to be discovered. Infants learn their language's phonetic categories and dozens of its word-forms at the same time, but how these learning processes interact is unknown. In addition, future work carefully testing children's own speech and their interpretation of language may reveal connections that have not been uncovered using gross measures of productive ability like vocabulary size. Finally, close examination of the phonetics of infant-directed speech will permit the quantitative modeling that is essential for understanding the learning problem most children solve so well.

---

## Recommended Reading

- Dietrich, C., Swingley, D., & Werker, J.F. (2007). (See References). An empirical paper describing how children from two different language environments interpret the same speech-sound variation in different ways, showing how phonological knowledge guides language comprehension during word learning.
- Jusczyk, P.W. (1997). *The Discovery of Spoken Language*. Cambridge, MA: MIT Press. A readable and thorough discussion of how infants begin to discover words from speech, with accessible descriptions of empirical methods and results.
- Jusczyk, P.W., & Aslin, R.N. (1995). Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology*, 29, 1–23. The first laboratory demonstration of infants' ability to find words in continuous speech.
- Soderstrom, M., White, K.S., Conwell, E., & Morgan, J.L. (2007). Receptive grammatical knowledge of familiar content words and inflection in 16-month-olds. *Infancy*, 12, 1–29. A good review of infants' grammatical knowledge, and four studies examining children's sensitivity to properties of English word order and grammatical inflections.
- 

**Acknowledgments**—This work was supported by research grants from the National Institutes of Health (National Institute of Child Health and Human Development HD049681) to D.S. and the National Science Foundation (HSD-0433567) to Delphine Dahan and D.S.

## REFERENCES

- Bortfeld, H., Morgan, J.L., Golinkoff, R.M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science, 16*, 298–304.
- Dietrich, C., Swingle, D., & Werker, J.F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences, USA, 104*, 16027–16031.
- Eimas, P.D., Siqueland, E.R., Jusczyk, P.W., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*, 303–306.
- Ferguson, C.A., & Farwell, C.B. (1975). Words and sounds in early language acquisition. *Language, 51*, 419–439.
- Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In J.H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The Adapted Mind* (pp. 391–428). New York: Oxford University Press.
- Graf Estes, K., Evans, J.L., Alibali, M.W., & Saffran, J.R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science, 18*, 254–260.
- Jusczyk, P.W., & Hohne, E.A. (1997). Infants' memory for spoken words. *Science, 277*, 1984–1986.
- Jusczyk, P.W., Houston, D.M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39*, 159–207.
- Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society of London, B, 363*, 979–1000.
- Marchman, V.A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science, 11*, F9–F16.
- Nittrouer, S. (1996). Discriminability and perceptual weighting of some acoustic cues to speech perception by 3-year-olds. *Journal of Speech, Language, and Hearing Research, 39*, 278–297.
- Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language, 57*, 24–48.
- Singh, L., Morgan, J.L., & White, K.S. (2004). Preference and processing: the role of speech affect in early spoken word recognition. *Journal of Memory and Language, 51*, 173–189.
- Swingle, D. (2005a). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology, 50*, 86–132.
- Swingle, D. (2005b). 11-month-olds' knowledge of how familiar words sound. *Developmental Science, 8*, 432–443.
- Swingle, D. (2007). Lexical exposure and word-form encoding in 1.5-year-olds. *Developmental Psychology, 43*, 454–464.
- Swingle, D., & Aslin, R.N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition, 76*, 147–166.
- Swingle, D., & Aslin, R.N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science, 13*, 480–484.
- Swingle, D., & Aslin, R.N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology, 54*, 99–132.
- Thiessen, E.D., & Saffran, J.R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology, 39*, 706–716.
- Werker, J., & Tees, R. (1984). Cross-language speech perception evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49–63.