Finding the Words:
How Young Children Develop Skill in Interpreting Spoken Language

Anne Fernald

Stanford University

Michael Frank

M.I.T.

To appear in: *Cambridge Handbook of Psycholinguistics*

Finding the Words:

How Young Children Develop Skill in Interpreting Spoken Language

Studies of spoken word recognition either by adults or by children all explore how listeners perceive and interpret strings of speech sounds. However, they focus necessarily on very different kinds of questions. One reason for this divide is that studies with mature, fluent language users can take it for granted that subjects know the meanings of the words they hear. Research on adult word recognition (reviewed by Samuel & Sumner, this volume) investigates a wide range of questions about factors influencing lexical access and ambiguity resolution in a system where it can be assumed that lexical representations are phonologically and semantically well established, that sentence processing is guided by mature syntactic knowledge, and that the cognitive capacities involved in processing are highly practiced and stable. With young language learners, the situation is crucially different. Although the “competent infant” has received a lot of press in recent years, based on many new findings revealing early skill in processing speech sounds (reviewed by Saffran & Sahni, this volume), learning a first language is a long, slow process that begins with gaining familiarity with initially meaningless sound patterns and only gradually moves toward fluent understanding. The young infant’s dawning awareness of language occurs over a period when the central nervous system is changing more rapidly than at any other time in postnatal life, a developmental trajectory profoundly influenced by early linguistic and social experience. Thus the child’s emerging linguistic knowledge and skill in interpreting spoken language are moving targets for the developmental researcher, influenced by many different endogenous and experiential factors and undergoing continual change on multiple levels from month to month.

If it is challenging for adults to make sense of potentially ambiguous strings of speech sounds as they fly by, it is even more so for the immature language learner, who must first build up initial representations of word forms from the input and then figure out how these word forms are used to convey meaning. Adults know tens of thousands of words in their dominant language, and all but a few of these are still “novel words” to the one-year-old hearing that language. If we thought about it at all, we would most often conclude we had understood every sentence heard in our native language that day, at least those that were clearly spoken and attended to, even if they were not directed to us. In contrast, young infants can understand very little of the speech they hear; thus potential ambiguities in strings of speech sounds pose a more fundamental problem. Even when children in the second year begin to show evidence of understanding more and more words, they have rudimentary semantic knowledge compared to that of an older child or an adult. Thus the young child’s task is not only to interpret familiar words correctly based on partial knowledge, but also to *discover* the unknown words in speech, discerning the meanings of these new words while also figuring out the grammatical rules that govern how these words can be combined and interpreted.

Our goal in this chapter is to identify central questions motivating current research on children’s spoken word recognition and early understanding, and to describe a few of the many recent contributions to this dynamic area of developmental inquiry. Following a brief historical overview, we review new findings showing how infants in the first year of life find the words in speech and begin to associate word forms with meanings, how children in the second and third year develop skill in recognizing familiar words and interpreting words in combination, and how young language learners across this age range deal with pervasive ambiguity as they continually encounter unfamiliar words in the speech they hear.

Early research on early understanding: A historical perspective

How infants begin to make sense of speech is an intriguing question with a long history. Reflecting on his own experience as a young child, St. Augustine concluded that “by constantly hearing words, as they occurred in various sentences” in conjunction with gaze and gestural cues from his elders, he “collected gradually for what [these words] stood” (398/1961, p.11). Some 1500 years later, the German philosopher Tiedemann published the first scholarly observations of early language learning, describing his infant son's gradual progress in understanding speech. At 8 months, his son appeared to recognize names of a few familiar objects, turning to search when he heard the word, and by 14 months he could articulate a few words but did not yet appear to use them intentionally. Tiedemann concluded that at this age "words awakened in him their proper images and ideas, but not conversely, images of objects and desire of them, any concept of the corresponding word; primarily because children begin by learning words more for the sake of understanding the intention of others than in order to impart their own" (Tiedemann, 1787/1927, p. 221). This early insight anticipates a very modern point of view, reflected in the recent research literature on the role of infant mind-reading skills in word learning (Tomasello, 2001). A century later, the diary studies of Hippolyte Taine and Charles Darwin also anticipated questions that have proven to be of enduring scientific interest. Taine (1877) noted that when his 11-month-old daughter was asked "Where's Mama?", she always turned toward her mother, an example similar to an earlier observation by Tiedemann: when asked to "Make a bow" or "Swat the fly", his 8-month-old son also made appropriate gestures. However, their interpretations differed: While Tiedemann asserted that his son had "learned to comprehend" simple sentences, Taine was more cautious, suggesting that "there is nothing more in this than an association". By the age of 12 months, however, Taine thought his daughter did demonstrate true comprehension of the word *bébé*. Although the child's understanding did not coincide with the conventional meaning, Taine claimed that *bébé* had "a general signification" for her beyond a limited association between a sound pattern and a gestural response. Such perceptive parental observations reflected curiosity and wonder about the origins of children's understanding, long before language learning was viewed as a legitimate object of scientific inquiry. It is easy to underestimate the originality of these early diary studies given their lack of methodological rigor, but they were the first to raise challenging questions that continue to motivate research on the development of spoken language understanding: When do infants first begin to respond distinctively to words as familiar sound patterns? And when and how do young children learn to apprehend meaning in these sounds?

When experimental research on spoken word recognition by adults began to emerge in the 1950's, studies addressing similar questions in children soon followed. For example, psychoacoustic research on contextual factors influencing the intelligibility of speech in noise (Miller, Heise, & Lichten, 1951) led to parallel studies with children, often motivated by clinical concerns (see Mills, 1975). Over this same period, young children's responsiveness to spoken language was increasingly of interest to linguists and psychologists working in quite different traditions. In the newly emerging field of language acquisition, Roger Brown (1957; 1973) used informal testing methods and naturalistic observation to explore the early development of comprehension. And psychologists studying perceptual development began to devise ingenious new experimental procedures for assessing discrimination and categorization of visual and auditory stimuli by young children. The introduction of an habituation procedure that could be used to investigate categorical perception of speech sounds by 2-month-olds (Eimas, Siqueland, Jusczyk, & Vigorito, 1971) enabled and inspired numerous studies of infants' abilities to discriminate and group vowels and consonants (see Aslin, Pisoni, & Jusczyk, 1983). However, it was not until the 1980's that work in these different traditions began to converge, crossing disciplinary boundaries to explore deeper questions related to language learning. As researchers in the field then known as "infant speech perception" realized how many interests they had in common with some of their colleagues working in the separate field of "language acquisition", there was a significant shift in emphasis: studies of early speech processing began to focus on the *discovery procedures* infants use to identify words and higher order elements in spoken language.

Finding the words in the first year

Even as newborns, infants are more interested in listening to speech than to other engaging forms of auditory stimulation (Vouloumanos & Werker, 2007). But to begin finding meaning in spoken language, infants must detect patterns and regularities in the sequences of sounds produced by speakers of the language they are hearing. Hundreds of experiments on speech perception in the first year of life have shown that infants become attuned to characteristic sound patterns in the ambient language months before understanding or speaking a single word, (see Jusczyk, 1997; Kuhl, 2004;

Saffran & Sahni, this volume). Here we describe just a few of many recent studies exploring how infants become acquainted with complex distributional patterns in spoken language and how these implicit learning strategies enable them to identify potential words using multiple sources of information available in the sound patterns of continuous speech.

The majority of experiments on speech processing by infants in the first year have investigated their capacity to recall isolated sound sequences, rather than examining associations formed between words and objects or individuals. Much of this research has focused on the task of word segmentation, exploring how infants identify particular sequences of sound within a larger body of fluent speech. In an influential study of early segmentation abilities, Jusczyk and Aslin (1995) asked when infants first showed evidence of being able to identify repeated word forms embedded in fluent speech. Using a headturn-preference-procedure, infants were first familiarized with multiple repetitions of a single word such as *cup* or *dog* and then tested in an auditory preference procedure with passages that either did or did not contain the familiarized word. While 6 month-old infants showed no preference for the passages containing the words to which they had been familiarized, 7.5 month-olds did show a significant preference, indicating that they recognized the word forms heard earlier in isolation, even in the context of continuous speech. However, when infants were familiarized with non-words such as *tup* and *bawg* that differed by only one or two phonetic features from the words that were presented at test, they did not show a preference. This finding suggested that infants at this age did not confuse *tup* with *cup or* with *bawg* with *dog*, indicating a sensitivity to phonetic details. In a follow-up study, Jusczyk and Hohne (1997) found that 8-month-olds exposed to recordings of stories over a 10-day period showed a preference for lists of words from the stories they had heard, even two weeks after familiarization. These and other studies revealed that by the second half of the first year infants are beginning to be able to segment recurrent acoustic patterns corresponding to words from fluent speech, and that they are capable of long-term storage of these segmented word forms.

During this period infants also begin to show sensitivity to a variety of speech cues used by adults to recognize words in fluent speech, including distributional cues, lexical stress, phonotactic constraints, and allophonic variation (e.g., Mattys, White, & Melhorn, 2005). Saffran and colleagues made the important discovery that 7-month-olds make use of distributional information to segment highly predictable syllable sequences from fluent speech (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996). Other studies have asked when infants become aware of characteristic regularities in the ambient language such as phonotactic patterns (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk (1993) and lexical stress patterns (Jusczyk, Cutler, & Redanz, 1993), and how they make use of such language-specific cues in identifying potential word boundaries (e.g. Johnson & Jusczyk, 2001). For example, by the age of 10.5 months infants can use their knowledge of the typical “trochaic” stress pattern of English to segment words with an initial stressed syllable. They can also use subtler cues such as allophonic variations (e.g., the difference in aspiration between the /t/ in *nitrates* and the /t/ in *night rates*) as well as phonotactic probabilities or constraints (e.g., that the sequence /kt/ does not appear word initially in English, though it does in Russian) to identify potential word boundaries (Jusczyk, Hohne, & Bauman, 1999; Jusczyk, Houston, & Newsome, 1999). Moreover, infants can also learn *new* phonotactic patterns after only minimal exposure (Chambers, Onishi, & Fisher, 2003), and then quickly exploit such newly learned regularities as cues to identify the boundaries of novel words (Saffran & Thiessen, 2003). Thus although infants in the second half of the first year already show a strong commitment to the particular sound patterns absorbed from hearing their native language, early speech processing remains a highly dynamic process, and infants remain open to new experience as they build on prior learning. Just as recent research on spoken language processing by adults has focused on how listeners integrate probabilistic information from numerous sources (e.g., Seidenberg, 1997), developmental researchers are also now exploring how infants integrate multiple sources of information to find potential words in strings of speech sounds. (e.g., Curtin, Mintz, & Christiansen, 2005; Mattys, et al, 2005; Thiessen & Saffran, 2003, 2004).

In the studies exploring early skill in segmentation described so far, infants were familiarized with the stimulus words through repeated exposure during the experimental procedure. Other studies have asked when infants begin to recognize familiar sound sequences they have encountered frequently outside of the laboratory. The earliest evidence for sensitivity to familiar words comes from Mandel, Jusczyk, and Pisoni (1995) who found that 4-month-olds tested in an auditory preference procedure listened longer to their own name than to a distractor name. Toward the end of the first year, infants demonstrate familiarity with the sound patterns of a broader range of frequently heard words. Research with infants learning French (Halle & de Boysson-Bardies, 1994) and Dutch (Swingley, 2005) found that 10- to 11-month-olds listened longer to words likely to be familiar in the speech they were hearing, as compared to words less common in infants' experience. Parallel studies using electrophysiological measures confirm the findings based on behavioral measures, and also provide information about the time course of infants' responses to familiar and unfamiliar words (Koojiman, Hagoort, & Cutler, 2005). For example, Thierry, Vihman, and Roberts (2003) found that familiar words captured the attention of 11-month-olds in less than 250 msec, a neural response pattern that differed significantly from responses to control words that were matched for phonotactic structure but were unlikely to be familiar to infants at this age.

There is now abundant evidence that infants in the first year attend to the speech they hear around them, making detailed distributional analyses of acoustic-phonetic features of spoken language, and that words experienced frequently in daily interactions begin to emerge as salient acoustic patterns months before these words are perceived as meaningful. Such accomplishments are often cited as evidence of early "word recognition" by infants, and indeed they create a firm foundation for future learning. For example, a study with 17-month-olds showed that the word-forms found by mechanisms of auditory pattern learning are good targets for future mapping to objects (Graf Estes, Evans, Alibali, & Saffran, 2007). This finding, and others reported in the following sections, suggest that exposure to word forms even in the absence of meaningful interpretation can lay the groundwork for future learning of form-meaning correspondences.

Nevertheless, while research on adult word recognition presupposes that familiar words can be meaningfully interpreted as well as recognized, no such assumption can be made about semantic processing by infants in the first year. Identifying sequences of sounds as coherent acoustic patterns is obviously an essential step in lexical processing, but such form-based recognition can occur without any link between a particular sound pattern and a semantic representation. Thus although it is clear that infants by the age by 10 months have some kind of acoustic-phonetic representation for frequently heard sound patterns, this result is best viewed as evidence of pattern detection abilities that are prerequisite for recognizing words in continuous speech, a selective response to familiar words that constitutes word recognition in a somewhat limited sense since it can occur with no evidence of comprehension.

Associating word forms with meanings

At what point do infants begin to link familiar word forms consistently with particular meanings? In a large scale study of early vocabulary growth using the MacArthur Communicative Development Inventory, the median receptive vocabulary for 8–month-olds was around five words, according to parental report (Fenson et al., 1994). But it is hard to know whether this measure accurately reflects infants' appreciation of correct sound-meaning associations, or whether it is an overestimate based on conventionalized responses to frequently heard acoustic patterns. As Tiedemann observed, an infant rewarded for making a stereotyped gesture upon hearing "Swat the fly" does not necessarily understand any of the words. Although parents typically report that by the end of the first year their infants can speak a few words and appear to understand many more, growth in receptive language competence is much harder to observe than growth in productive abilities, because the processes involved in comprehension are only partially and inconsistently apparent through the child's spontaneous behavior. Thus researchers have relied more on experimental than observational methods to study the earliest stages of sound-meaning mapping. For example, Woodward, Markman, and Fitzsimmons (1994) taught 13 and 18 month-olds a new object label by naming an object nine times over the course of a short training session. By using an engaging type of test trial in which the infants chose the named object for use in a game, they found some evidence that 13-month-olds were able to identify the object that was paired with the novel label even after a 24-hour delay.

More recently, studies using procedures that rely on implicit looking-time measures have produced convincing evidence that children around their first birthday begin to make form-meaning mappings, even on the basis of a small amount of laboratory experience. For example, in the "switch" procedure, 14-month-olds are habituated to two different novel word/object pairs through repeated presentation of each object accompanied by its label (Stager & Werker, 1997). Once a habituation criterion is met, infants are shown a word/object pair that violates the learned associations. The outcome measure is infants' response to this switch on test trials, when one novel object is paired with the label previously presented with another novel object. Successful detection of the switch is measured by greater looking time to the incongruent pairing, compared to baseline looking to the congruent pairings. In an experiment with 8-, 12-, and 14-month-olds, the two younger groups of infants were able to learn a single word-object association, detecting the substitution of a novel word or object at test; however, when asked to learn two novel associations, they failed to detect a switch in one of the two novel pairings (Werker, Cohen, Lloyd, Casasola, & Stager, 1998). Werker et al. concluded that initial successes in this paradigm were due to a simple novelty preference: infants dishabituated to a switch because they noticed that some unitary stimulus element had changed. In the two-referent case, however, they had to notice a violated association between a word and its referent, and only 14-month-olds could do this reliably. Thus, although even very young infants may be able to learn some words with sufficient exposure, mapping forms to meanings within the timeframe of a laboratory experiment requires cognitive resources that may not be available until after the first birthday.

Most experimental studies of early word learning have focused on children's abilities to form links between words and object referents in unambiguous situations. But these constrained laboratory situations—in which a single, salient referent is named repeatedly—differ in important ways from the noisier, real-world situations in which children typically find themselves. In the context of daily interactions, an infant will often hear an unfamiliar word spoken in a sentence frame, rather than in isolation, with several objects present as potential referents in the field of view. In this more complex environment, the learner may not be able to infer which object the unfamiliar word relates to from a single exposure. Instead, it may be necessary to aggregate information over multiple uses of a word. This kind of aggregation has been termed "cross-situational learning." For example, if the infant hears the unknown words "ball" and "bat" spoken when both a ball and a bat are present, it will not initially be clear which word goes with which object. But if the words "dog" and "ball" are heard soon afterwards in the presence of a dog and a ball, it might be possible for the young learner to keep track of statistical evidence across these two situations, noting that the ball was the only object present on both occasions when the word "ball" was heard. If so, this could help the infant to map "ball" to the appropriate object. In a simple artificial language study, Yu & Smith (2007) demonstrated that adult participants were able to associate unfamiliar words with the appropriate referent even when up to four objects and four words were presented simultaneously. A follow-up experiment using a preferential looking procedure suggested that 12- and 14-month-old infants could also learn word-object associations from individually ambiguous presentations (Smith & Yu, 2008). Thus, some kind of cross-situational learning mechanism is available to very young word learners in attempting to bridge the gap between form and meaning.

Figure 1 about here

What is the nature of this learning mechanism? One of the most exciting features of recent work on cross-situational word learning is the presence of computational proposals which instantiate ideas about mechanisms and make quantitative predictions about human performance. Early work by Siskind (1996) viewed cross-situational learning as involving a deductive inference about which parts of complex, propositional meanings mapped to words in sentences. While this proposal was ambitious and valuable, it also assumed that learners have access to very good guesses about precisely what speakers are talking about, an unrealistic assumption that limited its applicability to experimental situations. In contrast, more realistic proposals have assumed only that learners have access to their visual environment, but not to the propositional content of the speakers' utterances. For example, both Roy and Pentland (2002) and Yu, Ballard, and Aslin (2005) instantiated models of cross-situational word learning within systems that could parse continuous speech and identify objects via computer vision algorithms. Because they were able to learn word-object pairings using only raw video and audio data as input, these models provided a strong proof of concept that cross-situational mapping can be an effective strategy for acquiring the meanings of object words.

More recent work on this topic has focused primarily on the nature of the mapping mechanisms. Both Yu and Ballard (2007) and Frank, Goodman, and Tenenbaum (in press) have proposed specific mapping mechanisms which could be applied to the schematic output of sensory systems—segmented sets of words and lists of objects. The translation model proposed by Yu and Ballard attempts to estimate direct word-object associations. In contrast, the proposal by Frank et al. posits a specific, though unknown, referential intention – assuming that the speaker most likely intends to talk about only a subset of the objects that are present and uses only a subset of the words she says to refer to them. The idea that speakers' intentions could serve as a mediating variable may prove useful in linking results on early cross-situational word learning to the rich body of research showing that slightly older children soon become expert in using information about speakers' intentions to learn words (e.g., Baldwin, 1993). More broadly, investigations into the learning mechanisms underlying cross-situational learning may provide insights into how to understand the developmental trajectory of early word learning from pure word-form identification through word-object mapping and beyond.

Phonological Detail in Early Word Representations

Although infants around their first birthday may be able to map words to objects in the lab under the most favorable conditions, more detailed experimental investigations of early word-meaning mapping have produced a striking finding: early on, infants appear to have difficulty representing the detailed phonetic information in newly learned words. Studies using the "switch" procedure show that despite being able both to distinguish minimal differences between phonemes and to learn associations between novel words and objects, 14-month-olds may not succeed in mapping two new words to two new objects when the labels are a minimal phonetic contrast such as *bih* and *dih* or *bin* and *din* (Pater, Stager, & Werker, 2004; Stager & Werker, 1997). However, infants a few months older are more successful; although 14-month-olds fail to dishabituate to violations of the newly learned word/object pairings, 17- and 20-month-olds can correctly recognize this switch (Werker, Fennell, Corcoran, & Stager, 2002). Convergent evidence for developmental differences in the detail of phonetic representations across the second year comes from research using the electrophysiological method of event-related potentials (ERPs). In a study with 14- and 20-month-olds, Mills and colleagues (2004) compared neural responses to known words (*bear*) with responses to non-words that were either phonetically similar (*gare*) or dissimilar (*kobe*) to the familiar words. For 14-month-olds, ERP responses to known words differed from those to dissimilar non-words, but did not differ from responses to similar non-words. In contrast, for the 20-month-olds, ERP responses to the known words differed from responses to both types of distractor, as shown in Figure 2.

Figure 2 about here

The apparent lack of phonetic detail in early word-learners' representations is an interesting puzzle in research on early word learning. Given that much younger infants are able to recognize minimal phonetic differences in the first year of life, why do 14-month-olds fail to make use of this ability in the service of word learning? Recall the results by Jusczyk & Aslin (1995) reviewed earlier: 7.5-month-old infants can distinguish between *tup* and *cup* or *bawg* and *dog*. Yet when tested in the switch procedure at 14 months, children cannot learn that a toy labeled *bih* is different from a toy labeled *dih*. This paradox has motivated dozens of experiments, and a variety of possible explanations have been considered, ranging from the proposal that early phonological abilities are discontinuous with later lexical representations, to accounts based on the sparsity of infants' phonological neighborhoods or the limitations of their attentional and information-processing abilities.

One suggestion was that the precise, language-specific, phonetic representations that develop over the first year of life (Jusczyk, 1997; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984) might be unconnected to the level of detail represented in the lexicon. On this account, lexical representations are not continuous with early acoustic-phonetic knowledge but are built as part of a new, holistic representational system which is specific for word forms and contains only the amount of detail that is necessary (Brown & Matthews, 1997; Metsala & Walley, 1998). If young infants initially store words in a vague or underspecified form until word learning begins in the second year, the increasing need to distinguish similar-sounding neighbors might force attention to subtle phonetic detail. However, this discontinuity view now seems less plausible in light of recent experiments revealing detailed phonological specificity in the representation of highly familiar words by infants early in the second year. Swingley and Aslin (2000; 2002) showed children from 14 to 23 months pairs of pictures of familiar objects and played recorded sentences in which the name of the target object was pronounced either correctly (e.g., *Look at the baby*) or incorrectly (e.g., *Look at the vaby*). If children cannot distinguish *vaby* from *baby*, they should respond identically to both. However, looking to the named picture was significantly above chance across this age range, although it was slower and less reliable in the mispronunciation condition. Based on these findings, Swingley and Aslin concluded that infants as young as 14 months have lexical representations that are encoded in fine detail, even when this detail is not yet functionally necessary for distinguishing phonetically similar words in the child's vocabulary. The finding that 14-month-olds are sensitive to mispronunciations of highly familiar words in meaningful contexts suggests developmental continuity between the abilities involved in early speech perception and in later lexical learning, consistent with the view that infants beginning to build a vocabulary make full use of the impressive perceptual abilities demonstrated in research with younger infants.

Another possible explanation for the apparent lack of phonetic specificity in newly learned words shown by Stager and Werker (1997) is suggested by the work of Charles-Luce and Luce (1990; 1995), who analyzed the density of adults' and children's lexical neighborhoods (i.e., the number of words differing from a target word by only one phoneme). They found that on average, the lexical neighborhoods of adults' words were denser than those of children, lending support to a view that it may not initially be necessary for children to form highly specific lexical representations, given that fewer competitors to words are present to interfere with recognition (Vitevitch, Luce, Pisoni, & Auer, 1999). However, the work cited above and other recent findings also speak against this view. If 14-month-olds are sensitive to mispronunciations within highly familiar words, this accomplishment indicates a level of phonetic detail that should in principle not be necessary if phonetic specificity depends primarily on the size of lexical neighborhoods. Moreover, a series of corpus studies by Coady and Aslin (2003) has reopened the issue of phonological neighborhoods by analyzing transcripts of infants' early speech productions from the CHILDES database. These authors found that when relative vocabulary size is controlled, lexical neighborhoods may actually be *more* rather than less dense earlier in development, presumably because the first words children learn contain highly frequent sound combinations. Thus, the evidence for a large-scale change in the specificity of children's basic lexical representations—whether because of a representational shift or because of sparser phonetic neighborhoods—appears to be limited.

To explain the apparent lack of phonetic specificity in early word learning suggested by the results of Stager and Werker (1997), it may be more productive to focus on other factors that could contribute to children's failure to encode phonetic detail during the period of initial exposure to novel words. One simple explanation for this result is that the novel words taught in brief laboratory experiments differ substantially in frequency of exposure from the familiar words heard over and over during daily life. Recent support for this hypothesis comes from an experiment by Swingley (2007) asking whether previous exposure to a word form influences infant's success in later associating that word form with a meaning. One group of Dutch-learning 18-month-olds was familiarized with the acoustic form of a novel word in the context of a storybook, without linking that word to any particular referent; a control group did not receive additional exposure to the novel target word. Infants were then taught to associate the novel target word with an unfamiliar object. While both groups of infants learned to identify the referent of the novel word, only those with additional familiarization showed the "mispronunciation" effect observed in earlier studies. This result suggests that prior exposure to a particular word form may facilitate early lexical learning by enabling the child to form a phonetic representation that is more detailed and robust when it comes time to associate the word form with a meaning, similar to long term auditory priming effects observed in adults and older children (Fisher, Church, & Chambers, 2004). However, these results still do not explain the ERP findings of Mills et al. (2004) showing that for 14-month-olds, even familiar words like *bear* did not evoke different responses from phonetic neighbors like *gare*. Thus, while a frequency-based explanation is highly parsimonious, there are still other factors that need to be examined.

Werker and colleagues suggest a different kind of explanation for the failures of early word-learners to encode phonological details: an attentional resource conflict that impedes access to the full specificity of phonetic representations by infants in difficult word-learning tasks (Werker & Fennell, 2004). When the 14-month-old infant has to integrate novel visual and auditory information simultaneously, attempting to categorize an unfamiliar object and at the same time link it to a new word form, the cognitive load interferes with phonological processing. According to the PRIMIR model developed by Werker and Curtin (2005), such attentional demands inherent in the task of word-object mapping account for the failure of infants to detect the switch. However, while this proposal could explain the results of Stager and Werker (1997), it does so in terms of the demands of a particular experimental task. In a novel-word learning paradigm that allowed more time for consolidation (e.g., Ballem & Plunkett, 2005), it might still be possible for infants to make use of phonetic detail in learning novel words that are minimal pairs. Support for this view comes from a recent study in which infants tested in a visual preference procedure rather than the switch paradigm were able to identify an object mapped to either *bin* or *din* at greater than chance levels (Yoshida et al., in press). But although the 14-month-old participants' looking to the correct object was significantly greater than chance, their level of preference was very small, suggesting that the difference in results between the visual preference and switch procedures may simply be that one is more cognitively demanding or the other is more precise in its ability to measure fine differences. This result provides further support for the work of Swingley & Aslin (2000; 2002), suggesting that phonetic detail is present even in words learned in the laboratory, however fragile these initial representations may be.

The phonological specificity of the words children learn first has been a source of persistent interest because this question bears directly on the relationship between the sophisticated auditory pattern-recognition skills of infants in their first year and their later lexical development. Although the failure in the switch task at 14 months found by Stager and Werker (1997) remains puzzling, knowledge of the lexical representations of early word-learners has increased tremendously. It now appears that young word-learners do have access to detailed phonetic information in words, just as they did a few months earlier as infants, contrary to hypotheses positing reorganization or discontinuity. But learning to interpret a new word involves much more than learning to recognize an auditory pattern: appreciating phonetic and semantic detail in a new word initially requires more attention (Werker & Curtin, 2005) and more exposure (Swingley, 2007) than previously thought. As Naigles (2002) has aptly characterized the situation: "Form is easy, meaning is hard".

Developing skill in interpreting familiar words

Over the second year children typically learn to speak a few hundred words, and beginning around 18 months many show evidence of an increase in the rate of lexical learning known as the "vocabulary spurt". Although it is more difficult to track receptive language growth through spontaneous behavior, children reveal progress in understanding through increasingly differentiated verbal and behavioral responses to speech. Until recently, research on early lexical comprehension has focused almost exclusively on questions about what words children "know" and how this knowledge is acquired, i.e. the principles that guide their decisions in mapping new words to potential meanings (see Bloom, 2002; Woodward & Markman, 1997). Here we focus on a different aspect of early receptive language competence, reviewing research that explores how children put their lexical knowledge to use. These studies ask how young language learners develop efficiency over the second year in interpreting familiar words in fluent speech.

The majority of research on early comprehension has used methodologies that rely on offline measures (see Bloom, 2002), i.e. measures based on children's responses to a spoken word or sentence *after* it is complete rather than as it is heard and processed. For example, with diary observations and parental-report checklists of vocabulary, the judgment that a child does or does not "understand" a word such as *cup* or *eat* is made informally by adults interacting with the child in everyday life. In the case of offline experimental measures, judgments are based on the child's behavior in a more controlled situation, with a clearly operationalized response measure such as pointing to an object given two or more alternatives. While offline procedures may enable researchers to assess whether or not a child responds systematically in a way that indicates some understanding, such measures are inherently limited in their potential for illuminating underlying processes in the development of skill in interpreting familiar and novel words. Because offline measures do not tap into the real-time properties of spoken language, they can only capture the endpoint of the process of interpretation. As a consequence, data from offline procedures are often interpreted as all-or-none measures of competence, e.g. as evidence that a child either "knows" or "does not know" the word *cup* at 15 months, revealing little about the child's developing efficiency in identifying and interpreting this and other familiar words in continuous speech.

Of course it is possible that an observable behavior indicating understanding of a particular word at 15 months reflects a sudden epiphany. However, at this early stage of word learning it is more likely that the child's spontaneous demonstration of knowledge was preceded by a gradual strengthening of the association between this word and the class of objects this word is used to refer to, a process not as easily observed in spontaneous behavior. Haith (1988) has stressed the importance of tracking such incremental progress in cognitive development, urging researchers to think more in terms of graded concepts for the acquisition of "partial knowledge" instead of dichotomous terms that infants at a certain age either have or do not have a mature ability. Maratsos (1998) makes a similar point in discussing methodological difficulties in tracking the acquisition of grammar. Much of language acquisition goes on "underground", he argues, since intermediate stages of developing knowledge systems rarely show up directly in the child's behavior. To understand how the system progresses to mature competence in lexical as well as in grammatical development, it is necessary to gain access to these underground stages of partial knowledge. Researchers using online measures of spoken language processing with infants and young children are now making progress in that direction. Here we review recent findings based on online measures from two different procedures now used to investigate spoken language understanding by very young children: measures of neural responses in an ERP paradigm, and measures of gaze patterns in an eye-tracking paradigm. Both procedures yield gradient measures of the young listener's responses to the speech signal as it unfolds, illuminating different aspects of children's developing skill in making sense of spoken language over the second year.

Using ERP measures to investigate early comprehension

The event-related-potential reflects brain activity in response to a specific stimulus event. In studies with adults, particular components of the ERP response to linguistic stimuli have been linked to phonological, semantic, and syntactic processing (Osterhout, McLaughlin, & Bersick, 1997), and recent developmental studies have begun to ask whether and when children show comparable responses in each of these three aspects of language processing (Friederici, 2005). As mentioned earlier, researchers interested in early phonological development have found that infants by the age of 10 months show distinctive ERP responses to words presumed to be familiar versus unfamiliar, and that these differential neural responses are evident within 250 ms. of word onset (Kooijman, et al., 2005; Thierry, et al., 2003). Only a few ERP studies so far have focused on lexical-semantic processing of individual words. Mills, Coffrey, and Neville (1993) compared the responses of 13- and 20-month-olds to words reported by parents to be understood or not understood by the child. They found that patterns of brain activity associated with infants' responses to familiar words became more lateralized over this age range and similar in other ways to those of adults in comprehension tasks. While the younger infants showed bilateral differences in brain activity to known versus unknown words, the older infants showed unilateral differences in activity in the left hemisphere. Moreover, those 13-month-olds who were relatively more advanced in lexical development also showed a somewhat more mature ERP response, i.e. more similar to that of the 20-month-olds. However, it was not clear from these results whether it was overall vocabulary size or experience with individual word–object associations that was responsible for the dynamic shifts with age in the pattern of ERP activity. Moreover, because the auditory stimuli in this study were presented with no visual referents, it was also not clear that semantic processing was involved. To address these issues, a follow-up study measured ERPs from 20-month-old infants with high and low expressive vocabulary scores, for novel words they had just learned either paired with an object, or not paired with an object (Mills, Plunkett, Prat, & Schafer, 2005). The authors concluded that infants' individual experience with words in a meaningful context is a crucial factor in determining the patterns of brain activity that occur during lexical processing. As the child's experience with particular words increases, this leads to increased hemispheric specialization in the brain activity associated with these words.

While the most striking finding in the Mills et al. (1993) study was the extent of change in the *topography* of the ERP response to known and novel words over the second year, other studies have focused on developmental changes in the *time course* of responses to familiar and unfamiliar words. A question of particular interest is when young language learners begin to show neural responses known to be strongly associated with lexical-semantic processing in adults, such as the ERP component known as the N400, a negative waveform peaking at around 400 ms. The N400 effect is manifested in a larger amplitude for a word that is semantically incongruous in a particular context as compared to a word that is expected in that context (e.g., *He wore a hat on his foot/head*). Moreover, the amplitude of the N400, which is thought to indicate semantic integration difficulties, is sensitive to the immediate context of a particular word regardless of whether that context is a single word, a sentence, or a longer discourse (see Kutas & Federmeier, 2000). In a series of studies with German-learning children, Friedrich and Friederici (2005) used an ERP paradigm to investigate the development of semantic knowledge as well as phonological knowledge about possible word forms between the ages of 12 and 19 months. On each trial they presented a picture of a familiar object accompanied by one of four types of auditory stimulus: a congruent familiar word that matched the name of the object, an incongruent familiar word that did not match the object, or a "pseudoword" that was either phonotactically legal or illegal in German. Although 12-month-olds responded differently to congruous and incongruous words paired with familiar pictures, they did not show an N400 effect, which only began to emerge around 14 months. However, by 19 months an N400 was more clearly evident in response to both semantically incongruous words and phonotactically legal pseudowords, although not in response to pseudowords that were phonotactically illegal.

The Friedrich and Friederici (2005) findings showed that when infants are tested in a semantic processing task, the ERP response characteristic of adult lexical-semantic processing begins to emerge during the second year. However, when compared to the mature form of the N400 response, the effect observed in 14- and 19-month-olds began somewhat later and was longer in duration, suggesting that lexical-semantic processing is slower overall in children than in adults. Another important finding was that infants did not respond to psuedowords as adults do. In studies with adults, the N400 amplitude is typically larger for phonotactically legal pseudowords than for real words (e.g. *dottle* vs. *bottle*), presumably reflecting the listener's difficulty in accessing a lexical representation for the pseudoword in the mental lexicon. However, infants did not respond differentially to pseudowords and incongruous real words, as long as the nonwords were phonotactically plausible. This result is hardly surprising, given that an infant just beginning to build a vocabulary may be able to distinguish between phonotactically legal and illegal word forms, but cannot possibly know whether a phonotactically legal word form is "pseudo" or not. For a one-year-old who can correctly interpret only 500 words, all but a handful of the tens of thousands of words in the adult vocabulary are not yet represented in the mental lexicon. While an adult native speaker of English can immediately decide that *dottle* is indeed a *possible* word, while rejecting it as an *actual* word in the language, an infant in the second year can only make the first call. An unfamiliar word leads reliably to an N400 response for adult listeners because it is unexpected, although infants encounter unfamiliar words all the time. One interpretation of the finding that 19-month-olds responded similarly to known words and pseudowords is that they consider all unfamiliar but phonotactically plausible word forms as potential lexical items for which they will eventually discover meanings, an approach that would seem to be adaptive for the young language learner.

Recent research using measures of neural responses during lexical processing is providing new insights into the development of infants' skill in spoken language understanding, extending previous research based on offline measures in several important ways: ERP studies reveal gradual change over the second year in the brain areas involved in semantic processing, with greater hemispheric lateralization as the child increases in age and experience with language. Moreover, by providing precise measures of the time course of young children's responses to speech, these studies also reveal the gradual emergence of adult-like patterns of response to semantic anomalies in the temporal as well as the spatial domain. These promising recent findings suggest that research using ERP and other more sensitive brain measures will lead to greater understanding of the neural mechanisms that underlie the early development of receptive language skill.

Using eye-tracking measures to investigate early comprehension

To understand spoken language, children must learn to make sense of rapidly spoken strings of words, interpreting fluent speech incrementally using different sources of linguistic and nonlinguistic information. Fluent understanding by adults occurs automatically without time for reflection, so it is important to study the listener's online interpretation *during* speech processing using online measures that monitor the time course of the listener's response in relation to key points in the speech signal. Some of the classic online techniques used in psycholinguistic research with adults (e.g. phoneme monitoring, gating, and cross-modal priming) were adapted for use with school-aged children, but the task demands were problematic for younger children. However, refinements in eye-tracking procedures for use with children in recent years have provided a powerful new methodology for exploring the early emergence of skill in language comprehension. Automated eye-tracking procedures have been used very productively with adults, (e.g., Tanenhaus, Magnuson, Dahan, & Chambers, 2000) as well as with preschool children (e.g., Snedeker & Trueswell, 2004), although not to the same extent with younger children. In studies with infants, some developmental researchers have used a simpler two-alternative "preferential looking" procedure that relies on summary measures of looking time to a named picture (e.g., Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Meints, Plunkett, & Harris, 1999). However, such preferential-looking measures do not monitor the time course of the infant's response to the unfolding speech signal. While looking preference can reveal greater attention to one picture over another, it provides no more information than a traditional offline measure of object choice. Another version of the two-alternative procedure is similar in presentation format but differs importantly in measurement techniques and methods of data analysis (e.g., Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998; Swingley & Aslin, 2002). This "looking-while-listening" procedure incorporates the same online temporal measures of speed and accuracy used in eye-tracking studies with adults and older children, enabling more sensitive assessment of children's developing efficiency in interpreting familiar words in different contexts (Fernald, Zangl, Portillo, & Marchman, 2008).

In a cross-sectional study using the looking-while-listening paradigm, Fernald et al. (1998) found that English-learning infants made dramatic gains in the efficiency of word recognition over the second year. Infants looked at pictures of familiar objects while listening to speech naming one of the objects, and their eye movements were coded frame-by-frame by highly trained coders blind to trial type, yielding a detailed record of gaze patterns time-locked to key points in the auditory stimulus. The developmental changes in the speed and accuracy of understanding were striking: While 15-month-olds responded inconsistently and shifted their gaze to the appropriate picture only after the offset of the target word, 24-month-olds were faster and more reliable, initiating a shift in gaze before the target word had been completely spoken (see Figure 3). A similar pattern of developmental change in speech processing efficiency across the second year was found in a parallel study with Spanish-learning infants from Latino families living in the U.S. (Hurtado, Marchman, & Fernald, 2007). Extending these findings in a longitudinal design, Fernald et al. (2006) followed 60 English-learning infants at multiple time points from 12 to 25 months, asking whether measures of speed and accuracy in word recognition were stable over time and to what extent individual differences in processing efficiency were related to traditional offline measures of lexical and grammatical development. Analyses of growth curves showed that those children who were faster and more accurate in online comprehension at 25 months were also those who showed faster and more accelerated growth in expressive vocabulary and grammatical complexity across the second year. Infants' success at interpreting familiar words in degraded speech is also correlated with vocabulary size (Zangl, Klarman, Thal, Fernald, & Bates, 2005), further evidence that early development in speech processing efficiency is related to growth in central domains of language competence.

Figure 3 about here

This impressive increase in speed and accuracy over the second year shows that young children are increasingly able to identify words incrementally, responding based on partial phonetic information rather than waiting to hear the whole word. One consequence of incremental processing is that the young language learner is increasingly confronted with problems of temporary ambiguity. When Allopenna, Magnusen, & Tanenhaus (1998) presented adults with objects that included candy and a candle and asked them to *Pick up the can-*, participants waited to hear the next speech sound before orienting to the appropriate object, postponing their response until the final syllable of the target word made it clear which object was the intended referent. The child who hears *Where's the dog?* in the presence of a dog and a doll is also faced with temporary ambiguity, given that *dog* and *doll* overlap phonetically and are thus indistinguishable for the first 300 ms or so. Swingley, Pinto, and Fernald (1999) found that 24-month-olds in this situation also delayed their response by about 300 ms until disambiguating information became available. Even when they heard *only* the initial phonemes in a familiar word (e.g. the isolated first syllable of *baby* or *kitty*), 18-month-olds were able to use this limited information to identify the appropriate referent (Fernald, Swingley, & Pinto, 2001). Further evidence that infants interpret phonetic information in a probabilistic fashion during lexical processing comes from the studies by Swingley and Aslin (2000; 2002) mentioned earlier, showing that even younger infants can identify familiar words when they are mispronounced (e.g., *baby* vs. *vaby*), although they respond more strongly to the correct than to the incorrect version.

Such results from eye-tracking studies that use detailed time course measures show that infants make rapid improvement in both speed and accuracy of spoken word recognition across the second year. Although 24-month-olds are somewhat slower than adults to interpret a familiar object name in a simple sentence frame, consistent with findings of research on ERP responses in lexical processing tasks, two-year-olds can identify the named referent within a fraction of a second of hearing the onset of the noun. Moreover, measures of individual differences in speed of response to familiar words at 25 months are related not only to trajectories of vocabulary growth across the second year (Fernald, et al., 2006) but also to traditional offline measures of language and cognitive abilities in middle childhood

(Marchman & Fernald, 2008). These findings indicate that developmental gains in the efficiency of online processing in infancy are linked both to concurrent and long-term growth in lexical and grammatical competence.

Interpreting words in combination

Although infants may start making sense of language one word at a time, by the second year they are beginning to combine words in speech production and to understand increasingly complex multiword utterances. The young child learning to interpret fluent speech is confronted constantly with ambiguity on multiple levels, facing the challenges of identifying which unfamiliar sound sequences in the speech stream are possible words, determining what those novel words might be used to refer to, and figuring out how those words can be combined and recombined to convey many different kinds of meaning. As children begin to appreciate regularities at higher levels of linguistic organization in the second and third year, they can increasingly make use of their emerging lexical and morphosyntactic knowledge to identify and interpret new lexical items and find meaning in the strings of words they hear. A review of developmental research on syntactic processing is beyond the scope of a chapter focusing on early word recognition (see Gleitman & Trueswell, this volume), but we describe here a few recent studies exploring children's early use of morphosyntactic information in the noun phrase, as they attend to (or ignore) determiners in the process of interpreting object names and establishing reference.

Although the article *the* is one of the words English-learning children hear most frequently, it is never among the first words they speak. The observation that children produce dozens of nouns before they begin to use determiners led researchers to wonder whether spontaneous speech really captures what children "know" about language. In an early experimental study, Petretic and Tweney (1977) found that 2-year-old children noticed syntactic anomalies in the speech they heard (e.g. *Throw ronta ball* vs. *Throw me the ball*), revealing an awareness of linguistic regularities not yet evident in their own productions. Gerken and McIntosh (1993) pursued this question further by more systematically manipulating the functor words preceding familiar object names. In an offline picture book task, children responded more accurately to requests in which the article was grammatical than when the article was replaced by a nonce syllable, confirming that a violation of the determiner/noun pattern could disrupt the process of sentence interpretation. Extending these earlier findings, Zangl and Fernald (2007) used online measures to examine the timing of the disruption in response to an anomalous article before a familiar noun (*Where's po ball?* vs. *Where's the ball?)* in children from 18 to 36 months. They found that younger and linguistically less advanced children who did not yet use articles in their own speech were slower and less accurate in recognizing a familiar noun preceded by a nonce syllable than by a grammatical article, with disruption occurring within milliseconds of hearing the target noun. However, older and linguistically more advanced children showed no disruption; they were able to "listen through" a nonce article preceding a familiar word.

The apparent indifference of linguistically more advanced children to violations of the familiar article/noun sequence in this experiment could at first seem puzzling. One might expect children who produce articles in their own speech to be *more* vulnerable to disruption when encountering an uninterpretable functor-like syllable in the speech stream, as compared to children who have not yet begun to use determiners as grammatical elements in multiword utterances. However, the negative findings with linguistically more advanced children can also be interpreted as a sign of more advanced competence in speech processing, rather than as a paradoxical "failure to notice" the ungrammatical word. Because the sentences used as stimuli were highly predictable, all ending in familiar nouns in prosodically similar carrier frames, the words preceding the noun were redundant and uninformative. Efficient processing in this case could take the form of ignoring an ambiguous but irrelevant nonce syllable in the process of rapidly identifying the object name that is the focus of the sentence. The younger children may not have been able to take advantage of this redundancy because the target words were less well known to them, and because they did not yet appreciate articles as grammatical elements separable from the noun that follows.

Zangl and Fernald (2007) explored this possibility in a second experiment using a less predictable processing task, in which 34-month-olds were taught two novel words and then tested on sentences in which the newly-learned words were preceded by nonce syllables or grammatical articles. Although more linguistically advanced children at this age could ignore an anomalous article in conjunction with a familiar noun in a highly redundant context, they were significantly slower and less accurate in identifying a newly learned target word when it followed a nonce syllable than a grammatical article. That is, while children could ignore an anomalous functor-like word when interpreting a sentence in a highly redundant context, the same anomalous element was disruptive when they were listening for a less familiar object name and uncertainty was higher, even for linguistically more experienced children. Highly frequent function words are typically much less salient than content words as acoustic elements in the speech stream, and speakers reduce the salience of function words even further when the linguistic context is more predictable (Bell, et al., 2003). Skilled listeners can compensate for this reduction in phonological specificity by relying on top-down linguistic knowledge to make sense of the utterance, as long as other features of the sentence context are sufficiently predictable. The Zangl and Fernald findings provide new evidence for the gradual emergence of this kind of flexibility in speech processing over the first three years of life. The young language learner who is just beginning to use words in combination relies on surface regularities and lexical familiarity in the speech input, and so finds it more difficult to interpret a familiar object name when it co-occurs with an unexpected functor-like nonce word. The older and linguistically more advanced child has had more extensive experience with determiners as a class of words that occur in highly constrained contexts with varying pronunciation. Increasingly, acoustic variability within this class of words can be ignored as irrelevant as long as the discourse context is predictable. More advanced language learners are also more efficient in anticipating the focus of the spoken sentence, in this case the upcoming object label, and thus can ignore a "noisy" syllable in place of the article when it is unlikely to modulate the meaning of the focused word.

Although children learning English can afford to ignore a redundant article in some contexts, articles in languages with grammatical gender are potentially more informative. In Spanish, for example, all nouns have grammatical gender, with obligatory gender-marking on preceding articles (e.g., *la*$_{[f]}$, *el*$_{[m]}$, 'the'), and adult native speakers of languages with grammatical gender exploit this cue in online sentence interpretation (Dahan, Swingley, Tanenhaus, & Magnuson, 2000). Although the article *the* reveals little about the English noun that follows, hearing *la* or *el* in Spanish can inform the listener about the gender and number of the upcoming noun and in some contexts can facilitate identification of the referent before the noun is spoken. To investigate the early development of this ability, Lew-Williams and Fernald (2007) tested Spanish-learning children in the looking-while-listening procedure. Children saw pairs of pictures with names of either the same (e.g., *la pelota*, 'ball$_{[f]}$', *la galleta*, 'cookie$_{[f]}$') or different grammatical gender (e.g., *la pelota*, *el zapato*, 'shoe$_{[m]}$'), as they heard sentences referring to one of the pictures (e.g., *Encuentra la pelota*, 'Find the ball'*)*. On same-gender trials, the article could not be used to identify the referent before the noun was spoken; on different-gender trials, the gender-marked article was potentially useful in predicting the referent of the subsequent noun. If young Spanish-learning children are able to take advantage of gender agreement in interpreting speech, they should orient to the correct referent more quickly on different-gender trials than on same-gender trials. Indeed, children were significantly faster to orient to the referent on different-gender trials than on same-gender trials, as were native Spanish-speaking adults tested in the same procedure (see Figure 4). Moreover, children's ability to take advantage of grammatical gender cues was correlated with productive measures of lexical and grammatical competence. Although they were slower overall than adults in interpreting spoken language, young Latino children learning Spanish as their first language already demonstrated a significant processing advantage that is characteristic of adult native speakers but not of second-language learners (Guillelmon & Grosjean, 2001). With only a few hundred words in their productive lexicon, 2- to 3-year-old Spanish-learning children are able to identify familiar nouns 90 ms faster when a gender-marked article gives them an edge. This ability to exploit morphosyntactic information in the process of establishing reference reveals how the young child learning a richly inflected language makes progress in "becoming a native listener" (Werker, 1989).

Figure 4 about here

Conclusions

While early traces of connections among word forms and meanings are evident in the first year of life, "learning a word" is a gradual process. Much of this learning process occurs underground: representations of the forms and meanings of words are built up gradually before they begin to surface in children's productive vocabularies. Perhaps the best evidence for this graded, online view of early language understanding comes from measures like eye-tracking and event-related potentials. Unlike traditional looking-time or offline measures, these online methods allow for the detailed characterization of individual infants, allowing researchers to track the development of "partial knowledge." In addition, going forward, these methods will provide the best chance for researchers to test quantitative as well as qualitative models of development.

This gradualist view contrasts dramatically with the binary construal of word learning as a process of "fast mapping" between word forms and their meanings, a metaphor that is often used along with statistics about the exponential growth of vocabulary. As demonstrated by the original work on "fast mapping," only a few exposures may be necessary for a veteran word-learner to form a partial representation of a word (Carey & Bartlett, 1978). However, much more practice is necessary before the same learner can successfully interpret and produce that word appropriately across a range of contexts. In this review, we have attempted to give an overview both of the developmental progression in skills involved in word recognition by the young language learner, and in the historical progression of research on early spoken word recognition. In many ways, these two trajectories mirror one another, proceeding from the early groundwork laid by studies of auditory language processing and speech perception to a more complete understanding of the complexities involved in learning to communicate using words.

References

Aslin, R. N., Pisoni, D. B., & Jusczyk, P. W. (1983). Auditory development and speech perception in infancy. In M. M. Haith & J. J. Campos (Eds.), Infancy and the biology of development (Vol. 2, pp. 573-687). P. Mussen (Ed.). *Handbook of child psychology* (4th ed.). New York: Wiley.

Ballem, K. D., & Plunkett, K. I. M. (2005). Phonological specificity in children at 1; 2. *Journal of Child Language*, 32(01), 159-173.

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

Bell, A., Jurafsky D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America, 113,* 1001-1024.

Bloom, P. (2000). *How children learn the meanings of words: Learning, development, and conceptual change.* Cambridge, MA: MIT Press.

Brown, C., & Matthews, J. (1997). The role of feature geometry in the development of phonemic contrasts. In S. J. Hannahs & M. Young-Scholten (Eds.), *Focus on Phonological Acquisition* (pp. 67–112). Amsterdam: John Benjamins.

Brown, R. (1957). Linguistic determinism and the part of speech. *Journal of Abnormal and Social Psychology, 55*, 1 - 5.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development, 15*, 17-29.

Charles-Luce, J., & Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language*, 17(1), 205-215.

Charles-Luce, J., & Luce, P. A. (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies *Journal of Child Language*, 22(3), 727-735.

Coady, J. A., & Aslin, R. N. (2003). Phonological neighbourhoods in the developing lexicon. *Journal of Child Language*, 30(02), 441-469.

Curtin, S., Mintz, T.H., & Christiansen, M.H. (2005). Stress Changes the Representational Landscape: Evidence from Word Segmentation. *Cognition, 96,* 233-262.

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory & Language, 42*, 465-480.

Eimas, P., Siqueland, E., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. Science, 171: 303 – 306.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., et al. (1994). Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development*, 59(5).

Fernald, A. & Hurtado, N. (2006). Names in frames: Infants interpret words in sentence frames faster than words in isolation. *Developmental Science*, 9: F33 – F40.

Fernald, A., Pinto, J.P., Swingley, D., Weinberg, A., & McRoberts, G.W. (1998). Rapid gains in speed of verbal processing by infants in the second year. *Psychological Science*, 9, 72-75.

Fernald, A. , Perfors, A., & Marchman, V. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the second year. *Developmental Psychology*, 42: 98-116.

Fernald, A., Swingley, D., & Pinto, J. P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development*, 72, 1003-1015.

Fisher, C., Church, B. A., & Chambers, K. E. (2004). Learning to identify spoken words. In: D. G. Hall & S. R. Waxman (eds.), *Weaving a Lexicon* (pp. 3 – 40). Cambridge MA: MIT Press.

Frank, M. C., Goodman, N D., & Tenenbaum, J. B. (in press). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science.*

Friederici, A. D. (2005). Neurophysiological markers of early language acquisition: From syllables to sentences. *Trends in Cognitive Sciences*, 9: 481 – 488.

Friedrich, M. & Friederici, A. D. (2005). Semantic sentence processing reflected in the event-related potentials of one- and two-year-old children*.* *Neuroreport*, 16: 1801 – 1804.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M. and Gordon, L. 1987. The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language* 14*:* 23-45.

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.

Guillelmon, D., & Grosjean, F. (2001). The gender marking effect in spoken word recognition: The case of bilinguals. *Memory & Cognition, 29*, 503-511.

Hirsh-Pasek, K. & Golinkoff, R. M. (2006). *Action meets word: How children learn verbs*. Oxford University Press: Oxford UK.

Hurtado, N., Marchman, V.A., & Fernald, A. (2007). Spoken word recognition by Latino children learning Spanish as their first language. *Journal of Child Language*, 37: 227 – 249.

Jusczyk, P. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Kooijman, V., Hagoort, P., & Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cognitive Brain Research*, 24: 109 – 116.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews: Neuroscience*, 5: 831 – 843.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044), 606-608.

Lew-Williams, C. & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, 18: 193 – 198.

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.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, 134, 477-500.

Meints, K., Plunkett, K. and Harris, P. 2002. What is "on" and "under" for 15-, 18-and 24-month-olds? Typicality effects in early comprehension of spatial prepositions. *British Journal of Developmental Psychology* 20: 113-130.

Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). New York: Earlbaum.

Miller, G.A., Heise, G.A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, 41: 329 – 335.

Mills, D. L., Prat, C., Zangl, R., Stager, C. L., Neville, H. J., & Werker, J. F. (2004). Language Experience and the Organization of Brain Activity to Phonetically Similar Words: ERP Evidence from 14-and 20-Month-Olds. *Journal of Cognitive Neuroscience*, 16, 1452-1464.

Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1993). Language acquisition and cerebral specialization in 20-month-old infants. *Journal of Cognitive Neuroscience*, *5*, 317–334.

Mills, D. L., Plunkett, K., Prat, C., Schafer, G. (2005). Watching the infant brain learn words: effects of vocabulary size and experience. *Cognitive Development*, 20:19–31

Mills, J.H. (1975). Noise and children: A review of the literature. *Journal of the Acoustical Society of America*, 58: 767 – 779.

Naigles, L. R. (2002). Form is easy, meaning is hard: Resolving a paradox in early child language. *Cognition*, 86: 157 – 199.

Naigles, L. R., Bavin, E. L., & Smith, M. A. (2005). Toddlers recognize verbs in novel situations and sentences. *Developmental Science*, 8: 424 – 431.

Osterhout, L., McLaughlin, J., & Bersick, M. (1997). Event-related potentials and human language. *Trends in Cognitive Science*, 1: 203 – 209.

Pater, J., Stager, C., & Werker, J. (2004). The perceptual acquisition of phonological contrasts. *Language,* 80, 384-402.

Roy, D., & Pentland, A. (2002). Learning words from sights and sounds: A computational model. *Cognitive Science, 26*, 113-146.

Saffran, J.R. (2003) Statistical language learning: Mechanisms and constraints. *Trends in Cognitive Science*, 12: 110 – 114.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature,* 388(6640), 381-382.

Swingley, D. (2007). Lexical exposure and word-form encoding in 1.5-year-olds. *Developmental Psychology*, 43(2), 454-464.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76(2), 147–166.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13, 480-484.

Thierry, G., Vihman, M., & Roberts, M. (2003). Faniliar words capture the attention of 11-month-olds in less than 250 ms. *Neuroreport*, 14: 2307 – 2310.

Thorpe, K. & Fernald, A. (2006). Knowing what a novel word is not: Two-year-olds "listen through" ambiguous adjectives in fluent speech. *Cognition*, 100: 389–433.

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science*, 275(5306), 1599-1603.

Silva-Pereyra, J., Rivera-Gaxiola, M., & Kuhl, P. K. (2005). An event-related brain potential study of sentence comprehension in preschoolers: semantic and morphosyntactic processing. *Cognitive Brain Research*, 23: 247 – 258.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition, 61*, 39–91.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106*, 1558-1568.

Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science*, 8: 432 – 443.

Swingley, D. (2007). Lexical exposure and word-form encoding in 1.5-year-olds. *Developmental Psychology*.

Swingley, D., Pinto, J.P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, 71, 73-108.

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.

Thiessen, E. D., & Saffran, J. R. (2004). Spectral tilt as a cue to word segmentation in infancy and adulthood. *Perception & Psychophysics*, 66, 779-791.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language*, 68, 306-311.

Vouloumanos, A. & Werker, J.F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, 10: 159-164.

Werker, J. F. (1989). Becoming a native listener: A developmental perspective on human speech perception. *American Scientist, 77*, 54-59.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*: 197 – 234.

Werker, J.F. & Yeung, H.H. (2005). Infant speech perception bootstraps word learning. *Trends in Cognitive Sciences*, 9: 519-527.

Woodward, A. L., & Markman, E. M. (1997). Early word learning. In W. Damon, D. Kuhn & R. Siegler (eds*.), Handbook of child psychology, (Vol. 2), Cognition, perception, and language*. New York: Wiley.

Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (in press). Fourteen-month-olds learn similar sounding words. *Developmental Science*.

Yu, C., & Ballard, D. (2007). A unified model of word learning: Integrating statistical and social cues. *Neurocomputing, 70*, 2149-2165.

Yu, C., Ballard, D. H., & Aslin, R. N. (2005). The role of embodied intention in early lexical acquisition. *Cognitive Science, 29*, 961-1005.

Yu, C., & Smith, L. (2007). Rapid Word Learning Under Uncertainty via Cross-Situational Statistics. *Psychological Science, 18*, 414-420.

Zangl, R., Klarman, L., Thal, D. J., Fernald, A. & Bates, E., (2005). Dynamics of word comprehension in infancy: Development in timing, accuracy, and resistance to acoustic degradation. *Journal of Cognition and Development,* 6, 179-208.

Zangl, R. & Fernald, A. (2007). Increasing flexibility of children's online processing of grammatical and nonce determiners in fluent speech.. *Language Learning and Development.*

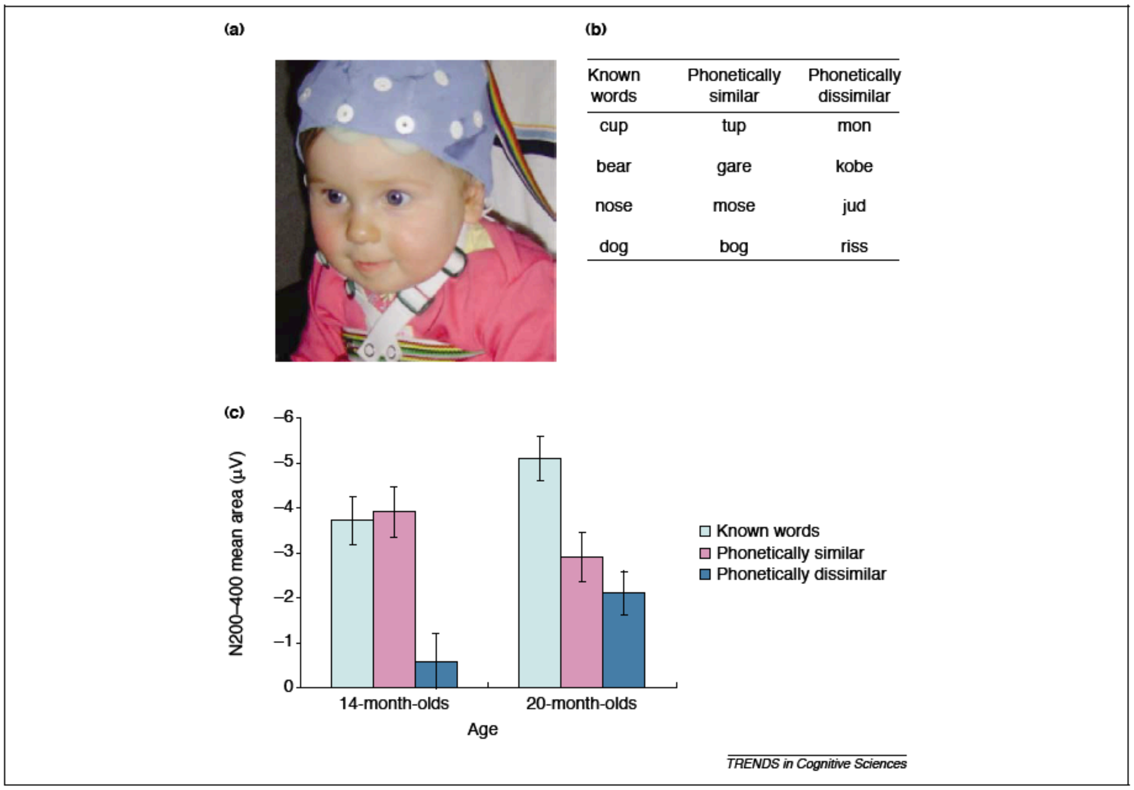
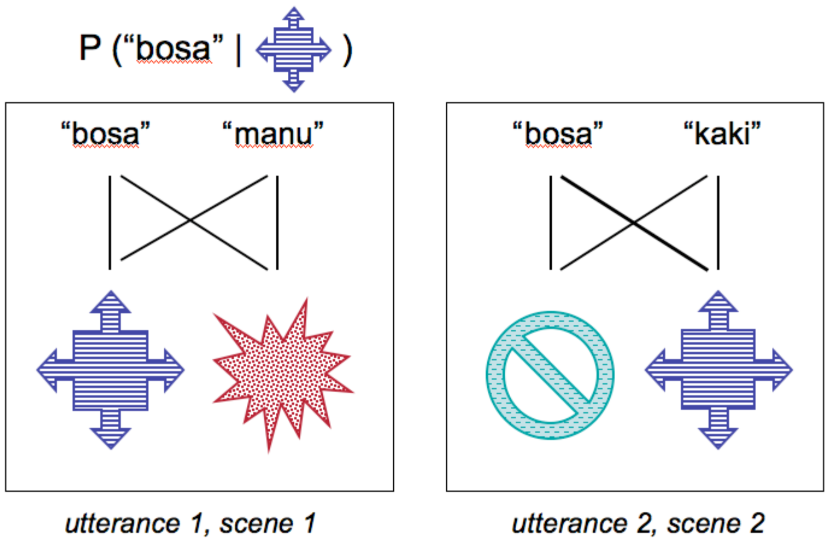


Figure 1. Infants in an ERP paradigm (a) heard known words and nonsense words that were either phonetically similar or dissimilar to those known words, as shown in the abbreviated list of stimulus words (b). The mean amplitude of the N200–400 word recognition component is shown in (c) in response to known words, phonetically dissimilar nonsense words, and phonetically similar nonsense words. At 20, but not 14 months, the neural representations accessed for known words are phonetically detailed enough to distinguish similar-sounding foils. Data are from Mills et al. (2004); Figure from Werker & Yeung (2005).

A.



B.

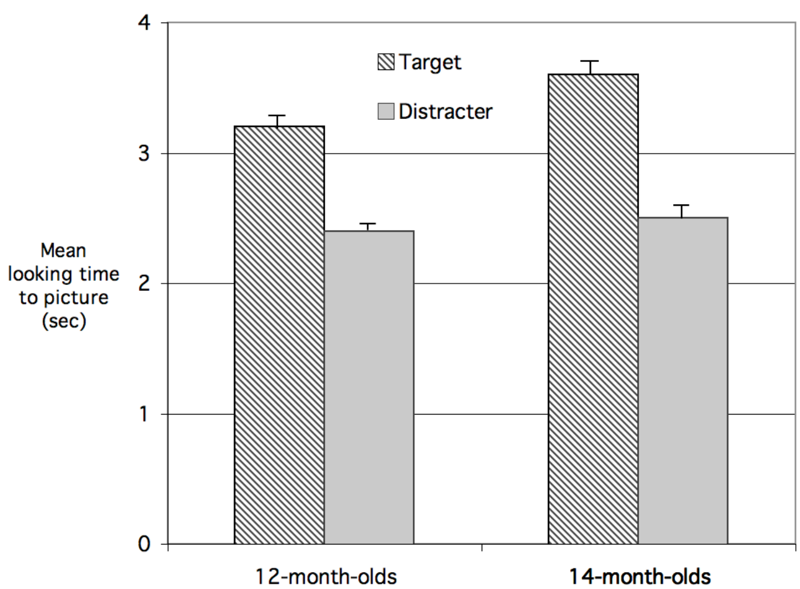


Figure 2. **A**. Schematic of associations among words and referents across two individually ambiguous scenes on sequential trials. If the infant calculates co-occurrence frequencies *across* these two trials, s/he can discover the mapping of the word "bosa" to the appropriate referent. **B**. Mean looking times to target and distracter pictures for younger and older infants. Adapted from Smith and Yu (2008)

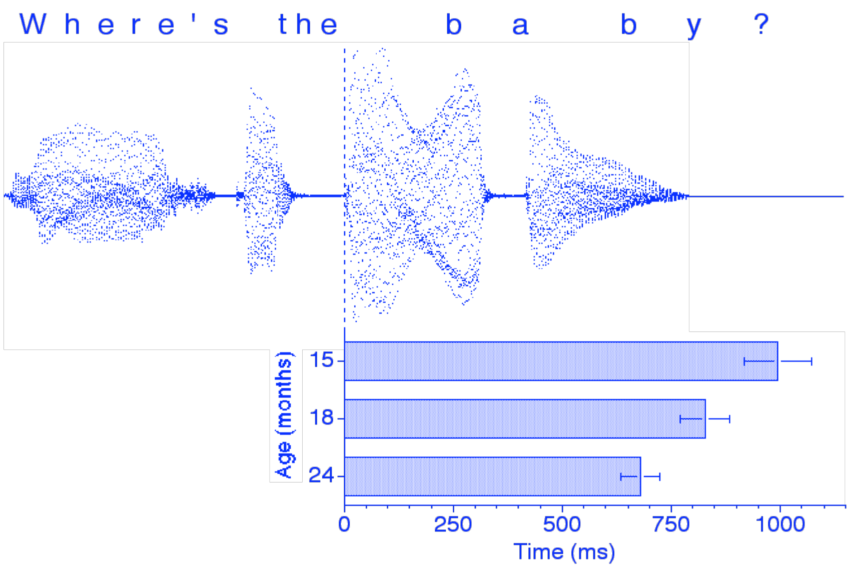
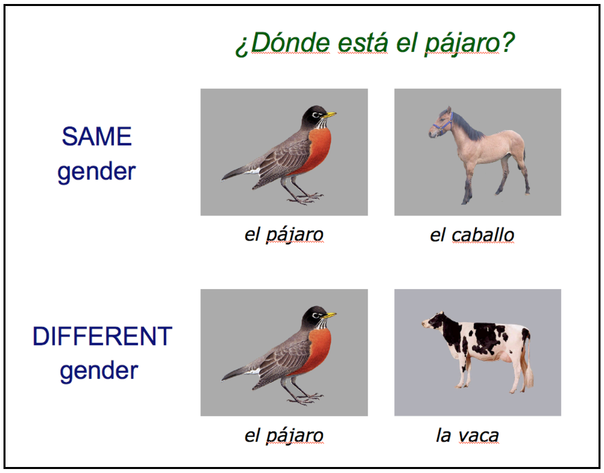


Figure 3. Mean latencies to initiate a shift in gaze from the distracter picture to the target picture, measured from the beginning of the spoken target word, for 15-, 18-, and 24-month old infants. This analysis included only those trials on which the infant was initially looking at the incorrect picture and then shifted to the correct picture when the target word was spoken. The graph is aligned with an amplitude waveform of one of the stimulus sentences. Figure from Fernald et al. (1998).

A.



B.

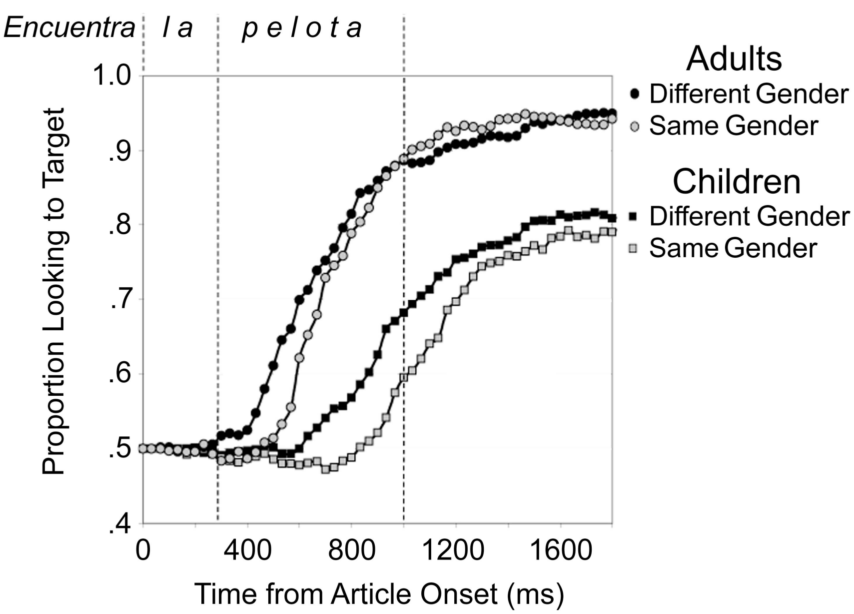


Figure 4. **A**. Examples of stimuli on Same-Gender and Different-Gender trials in Spanish. **B**. Curves depict changes in the proportion shifts from distractor to target picture by 3-year-olds and adults as the article and noun unfold, measured from article onset (in ms). Filled squares show responses on Different-Gender trials, when the article was potentially informative; open squares show responses on Same-Gender trials when the article was not informative. Vertical dashed lines indicate offsets of article and target word. Adapted from Lew-Williams and Fernald (2007).