23 Language Comprehension in Monolingual and Bilingual Children

KRISTA BYERS-HEINLEIN and CASEY LEW-WILLIAMS

1 Concordia University
2 Princeton University

Introduction

Long before they speak their first words, children begin to understand the language that they hear around them. Indeed, language comprehension—extracting meaning from speech—outpaces language production throughout development (Fenson et al., 2007). Nevertheless, understanding speech is a challenging and multi-faceted task. At minimum, children must identify and perceive speech sounds, parse the speech stream into its constituent words, identify the meaning of these words, consider their order in the context of a language’s grammar, and link the entire message to the speakers’ intended meaning (see Figure 23.1). All of this happens quickly and in real time: even “slow” infant-directed speech occurs at a rate of several syllables per second (Fernald & Simons, 1984). When and how do children come to understand what is spoken to them?

Recent research has shown that infants extract meaning from speech much earlier than previously thought. By age six to nine months, infants understand the meanings of many common words like feet, juice, and spoon (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 2011). However, the path to language comprehension neither begins nor ends there. Throughout development, successful language comprehension intertwines with developing linguistic, cognitive, and social abilities.

The development of language comprehension depends on the specific nature of children’s language environments. Monolingual children hear one of the world’s many languages; bidialectal children hear two varieties of the same language; and bilingual and multilingual children hear two or more languages. Diversity in language experiences can take many other forms as well. Children growing up in poverty often have fewer opportunities to hear words and sentences relative to children growing up with more resources (Hart & Risley, 1995; Weisleder & Fernald, 2013). Some children are exposed to signed rather than to spoken languages. Children with cochlear implants hear speech that is degraded relative to children with acoustic hearing (see Pisoni, this volume; Grice-Calab, Saffran, & Litovsky, 2009). And relative to typically developing children, children with developmental language disorders experience a complex interaction between cognition and language input (Rice, Warren, & Beiz, 2005). All children must adapt to the specific challenges presented by their environments. Children whose experience reduces the quantity and quality of language exposure are often slower in language acquisition. Other children, such as bilinguals or sign language learners, acquire language differently but are not delayed (see Kegl, this volume; Peña, Gillam, Bedore, & Bohman, 2011; Petitto et al., 2003; Werker & Byers-Heinlein, 2008).

In this chapter, we focus on cross-linguistic research with monolingual and bilingual infants and toddlers to explore how they navigate the path from hearing to understanding. While most research to date has investigated monolingual children, there is growing interest in understanding language acquisition in the many children around the world who encounter multiple languages early in life. Note that while we use the blanket term “bilingual” to refer to children acquiring two or more languages, this is anything but a one-size-fits-all category. There are vast differences in the quality and quantity of language experiences across different households and populations, the timing of exposure to different languages, and...
abilities to visually discriminate unfamiliar languages (Sebastián-Galés, Albreda-Castelló, Weikum, & Werker, 2012). Such sensitivities could help bilinguals extract meaning in their complex linguistic environments. Thus, infants use a range of auditory and visual cues to break into multiple languages, laying the foundation for discovering the sounds and sequences of sounds that comprise each language (Byers-Heinlein, Morin-Lessard, & Lew-Williams, in press).

**Phonetic development**

Words are built from sounds, and languages vary in terms of which sound differences are meaningful. These meaningful differences group speech sounds into phonetic categories. For example, the phonetic difference between /r/ and /l/ is meaningful in English, as in the words *red* and *lile*. This difference is not meaningful in Japanese, and so Japanese speakers tend to ignore it and group /r/ and /l/ into the same phonetic category. Infants cannot know at birth whether they will be growing up in an English, Japanese, or bilingual English-Japanese environment.

As such, newborn infants are sensitive to most sound differences that are meaningful across the worlds' languages. Important development occurs within the first year of life, when monolinguals lose sensitivity to non-native sound distinctions (Werker & Tees, 1984), but gain sensitivity to native language distinctions (Kuhl et al., 2007). This developmental pattern is often referred to as perceptual narrowing, and is thought to be driven in part by infants’ innate sensitivity to distributional regularities available in the input (Mey耶, Werker, & Gerken, 2002; Thiessen & Pavlik, 2013). Developing phonetic categories, together with growing knowledge of native language words, help children interpret whether a speech sound difference is meaningful or not (Dietrich, Swingley, & Werker, 2007).

Evidence from bilingual infants suggests that early language experience can affect phonetic development in unexpected ways (Byers-Heinlein & Fennell, 2014). For example, Spanish-Catalan bilingual infants show a U-shaped developmental pattern for their perception of vowels that exist only in Catalan (/e/-/e/); they readily discriminate this phonetic difference at 4 and 12 months of age, but sometimes fail to do so at eight months of age (Bosch & Sebastian-Galles, 2003). Monolingual Catalan-learning infants discriminate the same phonetic difference throughout the first year of life. While there are numerous different explanations for this finding (Byers-Heinlein & Fennell, 2014), one explanation focuses on how Spanish and Catalan link sound to meaning. Spanish and Catalan are both Romance languages with many cognates, which have similar meanings and differ on only a few sounds, which are often vowels (for example Catalan pilota and Spanish pilota, both meaning “ball”). Bilinguals acquiring these close languages may learn to ignore some vowel variability, and to focus on the invariant consonants (Sebastián-Galés & Bosch, 2009). Studies with populations of bilinguals learning languages that are not closely related, such as French-English and Spanish-English bilinguals, have found patterns of phonetic development that are similar to those of monolinguals (Burns, Yoshida, Hill, & Werker, 2007; Sundara & Saffran, 2010; Sundara, Polka, & Molnar, 2009). More research with bilinguals is needed to investigate a wider variety of phonetic contrasts and language pairs.

Looking and listening

Language acquisition begins with looking at and listening to native speakers of the ambient language(s). From very early in life, infants attend to speech over other types of sounds (Vouloumanos & Werker, 2007), and quickly target their attention to the native language or languages (Byers-Heinlein, Burns, & Werker, 2010; Moon, Cooper, & Fifer, 1993). Infants' preference for language is not limited to the spoken modality; six-month-old hearing infants with no exposure to sign language look more at linguistic signs than non-linguistic gestures (Christ & Corina, 2008). Young infants may also be sensitive to the notion that language can convey information between speakers. For example, 12 month olds understand that speech, but not other types of vocalizations such as coughing, can communicate information to a listener (Martin, Onishi, & Vouloumanos, 2012). However, non-speech signals, such as tones, can quickly approximate the same status if infants witness them being used to communicate in a natural dialogue (Ferguson & Lew-Williams, 2016).

For infants growing up bilingual or multilingual, it is not enough to simply attend to their languages in an undifferentiated fashion. Instead, they must acquire each as an independent communicative system, which hinges on an ability to detect the differences between languages (Byers-Heinlein, 2014b). While fully disentangling their two languages might be a somewhat gradual process (Byers-Heinlein, 2014b), there is evidence that the ability to differentiate two languages emerges early in life. At birth, monolingual and bilingual infants can discriminate between languages that differ in rhythm, such as English and French (Byers-Heinlein, Burns, & Werker, 2010; Mehler et al., 1988; NazzI, Bertocci, & Mehler, 1998). By age four to five months, monolinguals and bilinguals can also discriminate between rhythmically similar languages that belong to the same category as their own native language (Bosch & Sebastian-Galles, 2001; Molnar, Garvan, & Carreiras, 2014; NazzI, Josczyk, & Johnson, 2000). Infants can also distinguish languages using visual cues available on the lips and face of their interlocutors. Both English-monolingual and French-English bilingual four- and six-month-old infants can tell apart visual English and French when they see talking faces with the sound turned off. However, only bilingual infants retain this sensitivity at eight months (Weikum et al., 2007). Eight-month-old bilinguals also show enhanced
Mature language-specific phonetic categories can help infants interpret meaning in speech, but there is also evidence that consistent links between sound and meaning can actually help infants interpret speech sounds. In laboratory studies, infants who hear two sounds consistently paired with two different objects are more likely to discriminate these sounds than infants who hear the sounds paired randomly with the objects (Yeung & Werker, 2009; Yeung, Chen, & Werker, 2013).

Finding words in the speech stream

While infants are learning about the sounds of their native language(s), they also begin learning which sounds go together to form words (see Levine, Strother-Garcia, Hirsh-Pasek, & Michnick Golinkoff, this volume). Spaces signal word boundaries in written language, but silent pauses are not reliable cues to word boundaries in spoken language as they often occur in the middle of words. Children do sometimes hear words in isolation or at the edge of an utterance, and these words are relatively easy for them to pick out of the speech stream (Brent & Siskind, 2001; Johnson, Seidl, & Tyler, 2014; Lew-Williams, Pelucchi, & Saffran, 2011; Shukla, Nespor, & Mehler, 2007). However, most words occur in the middle of utterances, and infants must locate these words in order to eventually learn word meanings and interpret word combinations.

Infants can recognize familiar sound combinations in running speech—word forms—during the middle of their first year (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005). In a typical study, infants hear a list of familiar words, and later hear passages that either do or do not contain those words. Infants prefer listening to passages with the familiar list of words (Houston & Jusczyk, 2003), and are not fooled by similar-sounding words (Jusczyk & Aslin, 1995). This demonstrates their ability to segment the speech stream. Infants are especially skilled at segmenting words from familiar talkers and languages, particularly when words adhere to patterns typical of the native language (Brent & Cartwright, 1996; Houston & Jusczyk, 2000; Jusczyk, Houston, & Newsome, 1999; Polka & Sundara, 2012). Bilingual infants can flexibly and efficiently recognize word forms in each of their two native languages (Polka & Sundara, 2003; Vihman, Thierry, Lien, Keren-Portnoy, & Martin, 2007).

How do infants locate word forms in the speech stream? A learning mechanism known as statistical learning allows infants to track sounds and syllables that occur together in a way that is consistent (see Bomberg & Saffran, 2010). The central idea is as follows: sounds that occur together are more likely to be words than sounds that occur together in a way that is not related to the speech stream. After even brief opportunities to learn, eight-month-olds can detect words hidden in artificially constructed languages (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996) and in carefully controlled passages of natural but unfamiliar languages (Lew-Williams et al., 2011; Pelucchi, Hay, & Saffran, 2009). Some sources of variation, such as varying word lengths, can make statistical learning more difficult (Johnson & Tyler, 2010).

Over time statistical learning begins to interact with children's growing native language experience (Graf Estes & Bowen, 2013; Lew-Williams & Saffran, 2012). For example, infants sometimes use language-specific cues (e.g., in English, paying attention to the stressed syllables that often occur at the beginning of words) rather than statistical cues when segmenting speech (Johnson & Jusczyk, 2001; Johnson & Seidl, 2009; Thiessen & Saffran, 2003). Moreover, there is a coupling between language input and the dynamic nature of caregiver-child interaction: tactile cues from adults (Seidl, Tincoff, Baker, & Cristià, 2014), and highly familiar word forms such as the child's own name also aid in segmenting the speech stream (Bortfeld et al., 2005; Mensad & Nazzi, 2012). To date, most research on speech segmentation has studied monolingual infants, although recent research suggests that bilingual infants outperform monolingual infants in tracking regularities embedded in two interleaved artificial speech streams (Antov & Graf Estes, in press). Future research will need to investigate the complexities of segmentation in bilingual contexts.

Word learning

Once children locate a word in the speech stream, how do they figure out its intended meaning? Despite the potential difficulty of this task, children are powerful word learners, deploying a myriad of cognitive, linguistic, and social resources (Hollich, Hirsh-Pasek, & Golinkoff, 2010). As they gradually gain familiarity with common sequences of sounds, they begin to link those sequences to meaning (Graf Estes, Evans, Alibali, & Saffran, 2007). Children can sometimes infer the basic meaning of a word from a single example, a process called fast mapping (Carey & Bartlett, 1978). These processes set the stage for more protracted learning of a word's full meaning (Horst & Samelson, 2008; Swingley, 2010). The rest of this section will discuss some of the many contributions to successful word learning, as well as the ultimate outcome of this learning: a child's vocabulary.

Associative learning mechanisms

The ability to form associations between words and their referents is foundational to mature word learning. One year olds can successfully associate a picture of an object with a repeated word (Mackenzie, Curtin, & Graham, 2012b; 2012a; Werker, Cohen, Lloyd, Cassoula, & Stager, 1998). Even six-month-olds can do so if given appropriate prosodic information (Shuksa, Whith, & Aslin, 2011). Associative word learning abilities are robust regardless of whether children are growing up monolingual or bilingual (Byers-Heinlein, Fennell, & Werker, 2012). Young infants can also associate words and objects in more challenging conditions. Even when the same word is paired with several pictures, or when the same picture is paired with several words, infants are able to track which words and pictures co-occur most reliably (Smith & Yu, 2008; Vouloumanos & Werker, 2009).

A number of different perceptual and attentional cues can support the formation of these word-object links. For example, English learners tend to learn concrete
words such as nouns before other types of words (Bergelson & Swingley, 2013; Gentner, 1982), suggesting that some types of words are easier to learn than others. 

Low-level information can also affect how easily infants learn a new word. Infants are better at forming associations if an object is labeled synchronously with its motion (Gogate & Bahrkric, 2001; Matayaho-Bullaro, Gogate, Mason, Cadavid, & Abdel-Mottaleb, 2014), or if the labeled object is dominant in the infant's field of view (Yu & Smith, 2012).

**Word learning biases**

Children do not associate words with just any meaning. Instead, they expect new words to refer to whole objects, rather than to their parts, and expect newly learned words to refer to categories of objects of the same shape or kind (see Hollich, Golinkoff, & Hirsh-Pasek, 2007; Marchman, 1991). The origin of such expectations continues to be an important area of inquiry. Researchers have proposed diverse explanations: that these biases are built into the word-learning system (Markman, 1991), that they arise from children's social understanding (Bloom & Markson, 1998), or that they are learned from regularities in the environment (Smith, Jones, Landau, Cerreta-Kowale, & Samuels, 2002). Regardless of their origin, evidence from bilingual and multilingual infants suggests that early word learning environments can change infants' use of these biases. Using a word learning bias known as mutual exclusivity, children reject two labels for the same object, expecting each object to have only one basic-level label (Markman & Wachtel, 1988). This has been demonstrated in monolinguals as young as 16 to 18 months (Halberda, 2003; Markman, Wasow, & Hansen, 2005). However, children growing up in bilingual and multilingual environments do not show mutual exclusivity from the same age (Byers-Heinlein & Werker, 2009, 2013; Houston-Price, Caloghiris, & Ravignione, 2010). This difference is likely because bilingual and multilingual children, unlike monolinguals, hear multiple labels for the same object—one in each language. Thus, while monolinguals’ experience supports the notion of one-to-one mappings between words and objects, bilinguals’ experience could lead to more flexible word learning.

**Social information**

Infants also exploit rich social cues available in the environment, such as pointing and eye gaze, to help determine a word's meaning (Holllich et al., 2009). For example, 18-month-old infants are more likely to link a speaker's utterance with an object when the speaker is attending to that object (Baldwin et al., 1996). Similarly, when several different objects are present, children use their interlocutor's eye gaze and pointing to figure out what she is referring to (Baldwin & Moses, 2001). Beyond simply providing cues to a word's meaning, infants' understanding of a speaker's referential intentions is foundational to learning new words (Frank, Goodman, & Tenenbaum, 2009; Waxman & Gelman, 2009).

There is also evidence that infants' language background can affect their sensitivity to different types of social information. For example, because different speakers use different languages, bilingual children might be particularly sensitive to communicative information provided by a speaker. Consistent with this possibility, three-year-old bilinguals are better than monolinguals in using a speaker’s gaze to find where a toy is hidden (Yow & Markman, 2011).

**Vocabulary**

Children's vocabulary size provides a key index of their language development. To measure their receptive vocabulary—the words they can understand—children as young as 2 1/2 years are typically asked to point at which picture corresponds to a particular word (Dunn & Dunn, 2007). For younger children, parents check off different words that their child understands from a predetermined list (Tensen et al., 2007). In both cases, the number of words understood is compared to age-referenced norms to understand how a particular child compares to her peers. Studies that extrapolate from such measures suggest that the average monolingual 12 month old can understand about 100 words, which jumps to around 550 words for the average monolingual 18 month old (Mayor & Plunkett, 2011).

Children's receptive vocabulary is almost invariably larger than their productive vocabulary, as typically they understand all the words they can say, but do not say all the words they can understand. This appears to be especially true for bilingual children, who may have particularly disproportionate receptive vocabularies compared to their productive vocabularies (Gibson, Oller, Jarmulowicz, & Ethington, 2011).

Typically, bilingual children understand fewer words in either of their languages than monolingual children understand in their single language (Bialystok, 1997, 2007; Peets, & Yang, 2010; Poulin-Dubois, Bialystok, Blaye, Polenka, & Yott, 2012). This is thought to arise because bilingual children's language input is inherently split between two languages (Byers-Heinlein & Lew-Williams, 2014). Despite knowing fewer words in each language, bilingual children usually learn words at the same rate as monolinguals, and importantly, they understand a similar number of total words when both languages are considered (De Houwer, Bornstein, & Putnick, 2013; Marchman, Fernald, & Hurtado, 2010; Poon, Fernández, & Oller, 1995; Thordardottir, 2011). Bilingual children also understand translation equivalents—cross-language synonyms like English cat and Spanish gato—from an early age (De Houwer, Bornstein, & De Coster, 2006; Poon et al., 1995; Umbel, Pearson, Fernández, & Oller, 1992).

While there can be imbalances in vocabulary across a bilingual’s two languages, there is no consistent evidence that bilingual children are more likely than monolingual children to experience delays or deficiencies in language learning. Bilingualism is not considered a risk factor for language learning, and bilingualism does not impose an additional burden on children diagnosed with impairments such as specific language impairment and autism spectrum disorders (Paradis, Crago, Gémesse, & Rice, 2003; Petersen, Martino-Todd, & Mitre, 2012).
Understanding language in real time

Listeners usually encounter the words they know, as well as those they have yet to learn, in the context of running speech. Imagine if it took minutes or hours to determine the meaning of each incoming sentence—conversation would be impossible. Instead, communication occurs in real time, and young children show a developing ability to process speech as it unfolds. Fernald and colleagues (1998) presented young children with simple sentences (Where's the baby?), and found that 15-month-olds take approximately one second to move their eyes toward a picture of a baby, while 24-month-olds do so considerably faster. Similar developmental gains in real-time processing have been observed in Spanish-learning children from low-income households (Hurtado, Marchman, & Fernald, 2007). These findings are also echoed in studies of children's neural responses to familiar words (Friedrich & Friererici, 2005; Mills, Plunkett, Pratt, & Schafer, 2005). Young children even begin to recognize words after hearing partial phonetic information, such as the onset /t/-of a word (Swingley, Pinto, & Fernald, 1999).

Counterintuitively, words in sentences can be easier for children to understand than words in isolation (Fernald & Hurtado, 2006). One reason is that children can leverage information across different parts of an utterance. For example, Spanish-learning children can use gender-marked articles like el and de ("the") to predict whether a speaker will name an object with a masculine or feminine grammatical gender (Low-Williams & Fernald, 2007). Other studies show how young children exploit color and size adjectives (Fernald, Thorpe, & Marchman, 2010). Young monolingual children can even use familiar verbs and visual scenes to learn novel nouns (Ferguson, Graf, & Waxman, 2014; Waxman, Lidz, Braun, & Levin, 2009), and use sentence structure to learn the meanings of novel verbs (Naigles, 1990).

Experimental studies with young children are beginning to elucidate how these words are organized in the developing mind. Priming studies investigate whether hearing one word (e.g., cat) helps children access words that are related in meaning (dog) or sound (mat). Research shows that from around their second birthday, both monolingual and bilingual children indeed make links between words with related meanings (Anis-Trejo & Plunkett, 2009; Singh, 2013) or with overlap in their sounds (Holzen & Mani, 2012; Mani, Durrant, & Flocchia, 2012).

Children's ability to process language in real-time matters for later development. Children who respond faster to familiar words at age two have better language and cognitive outcomes in third grade, even when matched on overall vocabulary size (Marchman & Fernald, 2008). Similar longitudinal patterns have been observed in children with autism spectrum disorders (Vesker, Emslie, Saffran, & Weismier, 2013). Moreover, the speed of children's processing predicts which 18-month-old "late talkers" will make gains in language learning over the subsequent year (Fernald & Marchman, 2012).

Children's developing language expertise is built on a foundation of exposure to high-quality, high-quantity child-directed speech (Hart & Risley, 1995; Hultencker, Haight, Bryk, Seltzer, & Lyons, 1991). On average, children from high-income families hear three to four times as much language as children from lower-income families on welfare. Even within low-income families, there is striking variability in the use of language in the household. Latino children from low-income families who hear more child-directed speech at home are faster in real-time language processing, and less likely to fall behind in language learning (Weisleder & Fernald, 2013). For bilingual children, relative exposure to each language shapes language processing and word learning in each language (Hurtado, Grueter, Marchman, & Fernald, 2014). Proficiency in one language does not carry over to the other language, as vocabulary size in one language is not related to the other, and processing efficiency in one language is not related to the other (Marchman et al., 2010).

Challenges to language comprehension

While most research has tested infants' and toddlers' understanding of clearly articulated speech, real listening conditions are far from perfect. Everyday speech is replete with mispronunciations, accents, disabilities, and background noise. Children have more difficulty understanding degraded speech than normal speech, but their ability to cope improves with age and vocabulary size (Zangl, Klaaren, Thal, Fernald, & Bates, 2005).

Children's processing of mispronunciations provides a particularly interesting window into how they handle challenges to language comprehension. In typical laboratory studies of mispronunciations, children are shown pairs of objects on a screen (e.g., a dog and a baby), and then hear a label either correctly pronounced (Look at the baby), or mispronounced (Look at the bab). As early as age 12 months, monolinguals detect the mispronunciation, by looking less often and/or more slowly at the labeled object (Bailey & Plunkett, 2002; Swingley, 2005; White & Morgan, 2008). However, they still successfully identify the target object, demonstrating considerable flexibility in language comprehension. Experience improves children's word recognition: infants notice small sound changes more easily in familiar words than in newly learned words (Fernald, 2013; Stager & Werker, 1997).

Interestingly, there is evidence that Spanish-Catalan bilingual infants show a different pattern of processing mispronunciations than monolingual infants. As discussed previously, Spanish and Catalan share a high proportion of cognate words, such that they could be considered variant pronunciations rather than mispronunciations. Bilingual toddlers do not respond differently to correctly pronounced versus mispronounced cognates (Ramón-Casan, Swingley, Sebastián-Gallés, & Bosch, 2009). Likely because they have learned to ignore small sound variations in cognates, which do not change a word's meaning across the languages. Indeed, when non-cognate words are mispronounced, bilinguals respond like monolinguals, showing less robust recognition than when words are correctly pronounced (Ramón-Casan & Bosch, 2010). Similarly, infants exposed to two dialects of the same language show less sensitivity to variant pronunciations than those exposed to a single dialect (Durrant, delle Luche, Cattani, & Flocchia, 2014), perhaps mirroring infants' ability to ignore surface variation across speakers and instead attend to underlying structure (see Pardo, this volume; Matsui, this volume; Graf Estes & Law-Williams, 2015).
Acknowledgments

We thank the members of the Concordia Infant Research Lab / Laboratoire de Recherche sur l’Enfance de Concordia, and the Princeton Baby Lab. We also thank the National Institute of Child Health and Human Development for funding our recent bilingualism research (R01HD079779).

REFERENCES


infants’ memory for arbitrary syllable-object relations. *Infancy,* 2(2), 219–231. doi:10.1077/0891-1198.2.2


Acquisition

Language and Thought: Interrelations in Development. Cambridge Univ. Press.

Language Comprehension in Monolingual and Bilingual Children

Acquisition


