Chapter 2

The proto-lexicon: Segmenting word-like units from the speech stream

Caroline Junge [i]

There is growing evidence from experimental studies that infants start acquiring a vocabulary already when they are between six to nine months old (Bergelson & Swingley, 2012; Parise & Csibra, 2012; Tincoff & Jusczyk, 1999). Clearly, infants must have learned what these words mean prior to their lab visit, that is, learned them in their natural environment. However, as infant speech corpora reveal, there are surprisingly few one-word utterances directed to 6-12-month-olds in maternal speech (excluding exclamations, fillers and social expressions that typically do not combine into phrases): ranging from 2% (Morgan, 1996) to 9% (Brent & Siskind, 2001). Even when parents are explicitly instructed to teach their child certain words, they present these words predominantly in multi-word utterances (Aslin et al., 1996; Johnson et al., 2013). It appears that, for most children, most of the words they hear are presented in continuous speech. As Figure 2.1 demonstrates for a typical infant-directed example from Dutch, words are generally glued together without clear pauses between them to signal word onset and offset (unlike written text, in which words are conveniently separated by spaces). While it is debated whether or not isolated words alone can function as a starting point for vocabulary acquisition (Brent & Siskind, 2001; Depaolis, Vihman & Keren-Portnoy, 2014; Lew-Williams, Pelucchi & Saffran, 2011; Yang, 2004), it is undeniably true that in order to learn from the input, infants are required to segment the speech stream into separate word-like units from very early on. From adult research, we know that the speech signal contains several cues that assist successful word extraction from speech, but these cues are probabilistic rather than deterministic: no single cue appears sufficient to signal word boundaries (Cutler, 2012). Fortunately, the past two decades have provided us with two experimental methods that allow us to examine when and under what conditions infants can segment words from speech. In what follows next, we will first discuss the main methods (Section 2.1), before we focus on the cues that infants can use to detect words from speech (Section 2.2). Section 2.3 underscores how vital early speech segmentation skill is for subsequent language development, while Section 2.4 ends with future directions.

<FIGURE 2.1 HERE>

2.1 Methods to test infants' ability to segment words from speech

Infants' ability to segment words from speech has predominantly been tested via one of two methods: behaviorally with the head-turn preference procedure (HPP; Fernald, 1985) and electrophysiologically with an Event-Related Potential (ERP) paradigm. Both methods first familiarize infants with a set of words (familiarization phase), and then compare infants' responses to these familiar words against a control condition which presents infants with novel, unfamiliar words (test phase). Either the familiarization or test phase consists of multi-word utterances: thus a recognition response at test indicates that infants must have segmented the familiar words from speech. Figure 2.2 illustrates the set-up and typical results from both methods, explained in more detail below.

<Figure 2.2.here>

The HPP paradigm exploits infants' tendency to look at interesting visual stimuli (Fernald, 1985; see also Jusczyk, 2000; Bergmann et al., 2013). In a typical HPP set-up, infants sit in a three-sided booth with lamps in front of and to each side of the infant, all of which can be made to blink in order to attract the infant's attention (see top left panel of Figure 2.2 for a schematic representation). When one of the side lamps starts to blink, the typical behavior of a child is to turn her head to fixate this light. As soon as the child has turned her head, an auditory stimulus begins to play and continues until completion or when the child looks away. In this way, we can take visual attention as a proxy for auditory attention– by examining how long the infant looks at the blinking light while listening to a particular audio recording – and then compare whether infants prefer one type of audio recording over another. When infants display a preference for one type of auditory stimuli – by looking longer at the blinking light while listening to this stimulus – it is typically interpreted as their having discriminated between the two types of stimuli. Next we illustrate how this paradigm can be used to examine whether infants segment words from fluent speech.

Using this paradigm, Jusczyk and Aslin (1995) were the first to test whether infants could segment words from speech. They presented infants first with multiple tokens of two monosyllabic words in isolation (either *cup* and *dog* or *bike* and *feet*; around 12 tokens per word). At test, infants listened to four six-sentence passages: half of the passages contained familiarized target words, and the other passages the unfamiliar novel target words. Infants who were 7.5-months old (but not six-months-olds) listened longer to passages containing familiarized words ('familiarity preference'; see Figure 2.2). This familiarity preference holds when infants were familiarized with words embedded in passages and tested with isolated words. Infants further appear to store words in phonetic detail: at test, they did not display any preference for words that differ in only one phoneme from the target words (e.g., familiarized with *tup* and tested with *cup*-passages). To summarize, this study was the first to show that infants can match words presented in isolation with those embedded in a speech stream, thus demonstrating that infants can segment words from speech.

The seminal study of Jusczyk and Aslin (1995) ignited other studies to examine when and under which circumstances infants can detect words embedded in a speech stream (discussed in next section). Note that not all studies reported a preference for familiar words, however. As the top right panel in Figure 2.2 shows, the direction of the preference seems to hinge on the type of the stimuli used. While most studies using natural language stimuli report a familiarity preference, studies using artificial languages typically report a novelty preference (i.e., infants prefer the novel stimuli over the familiar words; see e.g., Saffran, Aslin & Newport, 1996). Artificial languages usually contain no real words from the child's language, and are created in such a way that there is only one cue in continuous speech that signal word onset (see further Section 2.2). It is possible that the direction of the preference is driven by the type of stimuli, or by the perceived difficulty of the task (Pelucchi, Hay & Saffran, 2009a ; Thiessen, Hill & Saffran, 2005). Nevertheless, while the stimulus characteristics that infants base their preference remain unclear (e.g., Houston-Price & Nakai, 2004), any preference – regardless of a novelty or a familiarity preference – signals infants' ability to detect recurrent word-like units in speech. Note that HPP is not the only possible paradigm to measure infants' listening preferences: to investigate infants' speech segmentation ability, one could also use a central-fixation procedure (Altvater-Mackensen & Mani, 2013; Ngon et al., 2013), which allows the possibility for future eye-tracker compatible designs.

Another method to test speech segmentation skill is to measure ERPs while infants are listening to speech. Unlike HPP, this method requires infants only to recognize familiarized words while they are passively listening to input: it does not require any behavioral change or preference. While children listen to speech and their EEG is recorded, they can be simultaneously entertained by showing interesting silent visual stimuli not time-locked to the auditory signal, such as a silent video, a puppet theatre, or toys (See lower left panel in Figure 2.2). Offline, one can derive the ERPs – reflecting the brain responses shared across similar events - by averaging over multiple EEG-epochs, all of which are time-locked to the same kind of event, which is in this case, the onset of familiarized vs. matched control words. Note that ERP studies need to present more test trials than is typical in behavioral research: to obtain meaningful ERPs it is common that infants contribute at least 10 artifact-free epochs (trials) per condition. In fact, infant ERP experiments usually comprise 3-4 times as many trials as are minimally required; the majority of trials will be discarded because they contain movement artifacts, which obscure event-related potentials associated with cognitive events such as recognition.

The ERP-paradigm offers the advantage that it allows for an on-line index of word recognition: by comparing the time course of ERPs time-locked to onset of familiarized words versus novel words one can observe whether and when in time the infant brain recognizes familiar words. The first study to create an electrophysiological analog of the familiarization-and-test HPP was Kooijman, Hagoort and Cutler (2005): they presented Dutch 10-month-old infants with blocks of ten isolated word tokens (familiarization phase) followed by 8 sentences, half containing the familiarized word, and half containing matched novel words. There were up to 20 familiarization-and-test blocks, each contrasting different familiarized and unfamiliar words. In this way, the authors could include more test trials in the experiment, while maintaining the amount of familiarization per word comparable to HPP studies. By comparing ERPs time-locked to familiar words with those time-locked to novel words - both embedded in speech-, we see that they diverge around 350 ms up to 500 ms from word onset, with ERPs to familiar words being more negative in amplitude than novel words (see Figure 2.2). The timing of this word repetition effect, also known as the N200-500, shows that infants are swift in initiating a recognition response even before the word has ended.

However, as with the HPP, we see differences across studies. Compare the two ERP figures in Figure 2.II (note that negative voltage is here plotted upwards): relative to unfamiliar words, familiar words elicited a negativity in 10-month-olds, while it elicited a positivity in 7-m-olds with exactly the same paradigm (Kooijman et al., 2013). Indeed, most studies report a negativity as a word recognition effect in infants of nine months old or older (French infants: Goyet et al., 2010; German infants: Männel & Friederici, 2013), while other studies testing younger infants with a similar set-up report a positivity (i.e., familiar words are more positive than unfamiliar words; German 6-month-olds: Männel & Friederici, 2013). Although it is likely that this polarity change signals age-related changes, brain maturation alone cannot explain this shift: some infants within the same recording session show negative word recognition effects during familiarization of single words, yet showing a less-mature positive word recognition effect when words are embedded in speech (Kooijman et al., 2013). What does such a shift then imply? One possibility is that a shift in polarity from positive to negative amplitude reflects a change from immature acoustic processing to more mature linguistic processing of the speech signal, and is thus indicative of how words are recognized (Kooijman et al., 2013). Similar explanations for shifts in polarity have also been reported infants' phoneme processing (Garcia-Sierra et al., 2011). Another possibility stems from contrasting the ERP-paradigm with the HPP-paradigm. Recall that the word recognition effect with negative amplitude is present in infants from around nine months – which is around the same age that behavioral studies on speech segmentation report familiarity preferences, i.e., that infants prefer listening to familiar words relative to novel control words. It is therefore also possible that the more mature word recognition effect (increased negativity to familiar words) not only reflects infants' ability to recognize words, but simultaneously also their preference for these words.

To summarize, the literature presents both behavioral and electrophysiological evidence that infants can segment words from speech. Both paradigms require infants to give a differential response to familiar and unfamiliar words. In both paradigms, it is interesting to see that there are two distinct types of responses (i.e. HPP: familiarity vs. novelty preference, ERP: negative vs. positive recognition components around 200-500 ms). Of course, any differential response - regardless of its direction - shows that infants can segment speech in this experiment. Although the interpretation of each switch in one direction is far from clear, comparison across studies show that the direction of the responses is generally consistent across studies when grouped by age or stimulus-type, allowing us to sketch a developmental picture of how speech segmentation ability matures, which is the focus of the next paragraph.

2.2 How can infants extract words from speech?

Although recognition of a word is obviously facilitated when (one of its) boundaries are clearly demarcated with silence, as is the case with isolated words (e.g., Brent & Siskind, 2001; Junge, Kooijman et al., 2012) or with words at phrase boundaries (Seidl & Johnson, 2006), infants still need to segment a large proportion of words from continuous speech. Natural languages generally contain a variety of cues to mark word onset in speech, with multiple cues often (but not always) signalling the same word boundary. However, no single cue appears to be fool proof (Cutler, 2012). For instance, one possible cue would be to use the offset of a known word as an anchor to predict the onset of the next word (Bortfeld et al., 2005) but sometimes this known word is part of a larger word (e.g., 'bed' in 'embedded'), which could lead to this cue misleading infants to wrongly segment the speech stream. Adult listeners can draw from a full array of cues, but they value these including cues hierarchically, favouring top-down (i.e., lexical and semantic-contextual) cues over bottom-up cues like co-articulation, phonotactic and prosodic cues (Mattys, White & Melhorn, 2005).

How do infants learn which cues are useful in extracting word-like units from speech? And how do they derive what word-like units are in the first place? Arguably, words in isolation provide infants with a basis for learning what word-like units are like (Altvater-Mackensen &

Mani, 2013; Lew-Williams et al., 2011). Furthermore, infants can use lexical cues to segment words from speech, e.g., use the offset of a known word as an anchor to predict the onset of the next word (Bortfeld et al., 2005; Shi & Lepage, 2008). However, their lexicon is too small to make this cue very viable. Hence, it appears that, in contrast to adults, infants must predominantly rely on bottom-up cues.

One of the bottom-up cues that infants can use is tracking the probability of one syllable following another in the child's language input (Saffran et al., 1996; Thiessen & Erickson, 2012). If the transitional probability (TP) between syllables is high, this would suggest that they tend to co-occur and form a cluster (i.e., a word). In contrast, when the TP is low, this would suggest that these syllables do not belong together, i.e., are (parts of) different words. Saffran, Aslin and Newport (1996) were the first to demonstrate that eight-month-olds can use this statistical cue to extract word-like units from an artificial two-minute language stream made up of concatenations of four trisyllabic words. Each syllable was distinct and the only cue to signal word boundaries was the TP between syllables: for consecutive syllables within a word the TP was always 1, whereas for consecutive syllables across words it was .33. Following familiarization with this two-minute language stream, infants listened longer to part-words presented in isolation (consisting of novel concatenations of the syllables) relative to words, which suggests that eight-month-olds were able to use the transitional probabilities of the syllables in the stream to distinguish the "words" they had been presented with from the novel concatenations.

Arguably, such an artificial language stream poorly reflects the rich natural input that infants hear in the real world (Johnson & Tyler, 2010), but other studies provide evidence that this learning mechanism is sufficiently robust to support word segmentation in daily life. First, infants readily treat such words from an artificial stream as possible words (Saffran, 2001; Graf Estes et al., 2007). Second, when TP is the only possible cue in passages taken from a natural unfamiliar language (that has more syllable variation than the artificial language stream), infants also extract words from such familiarization (Pelucchi et al., 2009a; Pelucchi, Hay & Saffran, 2009b). Crucially, infants can distinguish between frequent and infrequent syllable combinations in their native language (Ngon et al., 2013). Electrophysiological evidence suggests that even neonates can already recognize words based on their TPs (Kudo et al., 2011). Taken together, these results suggest that infants are able to pull out word-like units from fluent speech by relying on the TPs between syllables in the speech stream. In fact, Thiessen and colleagues note that tracking regularities, such as TPs between syllables, can be useful in two ways: first, it yields extra information about possible words in addition to the isolated words in a child's input ('conditional statistics'); second, from this set of words one can extract language-specific regularities, such as phonotactics and prosodic regularities ('distributional statistics'), which in turn form additional cues to segment words from speech (Thiessen & Erickson, 2012; Thiessen, Kronstein & Hufnagle, 2013).

Indeed, infants appear to use language-specific bottom-up cues to word onset at a later age than domain-general cues, with different cues following different trajectories. For (American-English) infants, word stress is the first language-specific cue that infants show sensitivity to (Curtin, Mintz & Christiansen, 2005; Jusczyk, Houston & Newsome, 1999). That is, infants can use the stressed syllable to signal the onset of a word in stress-based languages like English (but not in syllable-based languages like French), where most words receive stress on their first syllable (Cutler & Carter, 1987). When American-English infants are 7.5 months old, they correctly segment strong-weak words like *hamlet* and *doctor* from speech, but only segment the strong syllable (*tar)* of a weak-strong word such as *guitar* (Jusczyk et al., 1999; see also Kooijman et al., 2009, for Dutch). In contrast, infants acquiring a language without lexical stress do not readily use such a strategy (Nazzi et al., 2006, 2014; Polka & Sundara, 2012). Furthermore, cross-linguistic differences even emerge for infants acquiring similar languages: Dutch infants are delayed in extracting strong-weak words from speech compared to American infants, presumably because the difference between stressed and unstressed syllables is less transparent in Dutch (Houston et al., 2000).

Albeit at a slightly older age (from nine months onwards), infants further show sensitivity to the segmental properties of their native language to segment speech: e.g., phonotactics (Gonzalez-Gomez & Nazzi, 2013; Johnson et al., 2003; Mattys & Jusczyk, 2001a); coarticulations (Curtin, Mintz & Byrd, 2001; Johnson & Jusczyk, 2001); and allophones (Mattys & Jusczyk, 2001b). Although this evidence usually comes from studies demonstrating that older infants can use these cues to segment speech (and not that younger infants cannot use these cues), the pattern of results reported suggests that infants learn to use suprasegmental cues to the service of speech segmentation before segmental cues. However, most of what we know comes from American-English infants listening to stimuli in their native language. It is unclear whether similar trajectories also hold for other languages, or whether this is due to suprasegmental markers such as stress being more transparent relative to other segmental cues in English (but not in other languages; for a related discussion see Johnson, 2012). The limited evidence available suggests that languages do, indeed, differ in the developmental timeline of such language-specific speech segmentation cues (e.g., Van Kampen et al., 2007).

Clearly, infants gradually learn which cues are advantageous to segmentation of fluent speech in their language. When these cues work in concord, speech segmentation is facilitated (e.g., Mersad & Nazzi, 2012; Thiessen et al., 2005). What happens, however, when these cues conflict with one another? Do infants weigh each of these cues equally, or do they rely on some cues more than others? The limited studies on conflicting cues in word segmentation suggest that infants, like adults, weigh cues hierarchically, but across development, they vary the cues they increasingly rely on: initially, infants rely more on transitional probability cues (Thiessen & Saffran, 2003), but when infants are eight months old, they weigh language-specific cues more heavily (Hay & Saffran, 2012; Johnson & Jusczyk, 2001; Thiessen & Saffran 2003). Note that this shift emerges around the same time as when infants are said to be tuning into their native language (that is, show decreased sensitivity to non-native speech contrasts; Kuhl, 2004). Thus, it appears that, with increased exposure to their native language, infants become more native-like in their phonetic reorganization as well as in their ability to segment words from speech.

While most (HPP) studies examining infant speech segmentation ability manipulated the availability of speech segmentation cues, other (ERP) studies focused on how frequent infants should hear a word to elicit a recognition response. Because ERPs reflect an on-line manifestation of word recognition, this method allowed researchers to manipulate both the amount and type of familiarization required for word recognition. When both familiarization and test comprise continuous speech, Dutch 10-month-olds overall show a gradual recognition effect within eight tokens, i.e. the N200-500 gradually increases over repetitions (Junge, Cutler & Hagoort, 2014; See also Figure 3). This recognition effect during familiarization developed even faster (within six repetitions) when we examined only those infants (68%) who at test continued to show the mature N200-500 for familiar words over novel words. In another study we severely reduced the familiarization to just one word embedded in an utterance: only infants with relatively large vocabularies show the mature recognition response (Junge, Kooijman et al., 2012). Thus it appears that around 10 months, some infants require less repetitions than others to elicit recognition.

All in all, we can conclude that by the end of their first year, infants become increasingly proficient in recognizing words from running speech (see also White in Chapter 5) because they have multiple cues to speech segmentation coupled with more mature memory skill to match words at their disposal.

2.3 Recognition of proto-words facilitates lexical acquisition

The studies summarized above show that infants gradually become more sophisticated in segmenting word-like units from speech. Jusczyk (2000) suggested that once infants recognize proto-words – word-like units without any meaning – they will start looking for concepts to match these words onto. Do infants treat these proto-words that they have segmented from speech as possible word candidates? Indeed, there is both behavioral and electrophysiological evidence that infants do so.

First, several behavioral studies demonstrate that infants learn novel mappings better when they first hear them as proto-words in fluent speech (i.e., without the matching object present; Graf Estes et al., 2007; Hay et al., 2011; Lany & Saffran, 2010; Swingley, 2007). Even when infants are confronted with ambiguous word-object pairings, they rely more on the word-form than on the object to anticipate upcoming events (Zamuner, Fais & Werker, 2014).

Second, we note that two electrophysiological studies reveal similar word recognition effects, regardless of whether infants were presented with accompanying visual referents or not (Junge, Cutler & Hagoort, 2012 presented infants with pictures of typical early words and their matching words, while Junge, Cutler & Hagoort, 2014 presented infants with utterances in which low-frequency words were repeated, with no matching visual stimuli; cf. Figure 2.3): in both cases, the N200-500 becomes larger with repetition, suggesting that the words presented without visual context were treated like meaningful words. One might argue that this comparison merely shows that N200-500 indexes word-form repetition, but not necessarily the mapping of words to meaning. Yet, our finding that – after the word-object familiarization phase the same infants notice at test whether or not specific pictures and words belonged together (as indexed by an adult-like N400; Junge, Cutler et al., 2012) suggests that infants treated the words as meaningful during the familiarization phase (See also Friedrich in Chapter 6). In short, both behavioral and electrophysiological studies highlight the likelihood that infants treat proto-words as possible words.

<FIGURE 2.3HERE>

Other studies underscore the importance of infants' ability to find words in continuous speech by linking infant performance in speech segmentation studies to their future vocabulary development. Although infant research generally focuses on group reports, there is sufficient individual variation that reflects the infant's skill in segmenting words from speech. Several longitudinal analyses report (either via correlations or subgroup comparisons) positive linear relationships between infants' ability to detect words from fluent speech, and the subsequent size of their lexicons (behavioral evidence: Newman et al., 2006, 2016; Singh et al., 2012; electrophysiological evidence: Junge, Kooijman et al., 2012; Kooijman et al., 2014; Junge & Cutler, 2014). A recent meta-analysis estimates this effect to be moderate (r = +.33; 95% CIs [+0.17; +0.48]; Junge & Cutler, 2014). This further underscores how vital it is that infants can recognize word-forms from continuous speech: infants who show more advanced speech segmentation skill continue to develop larger vocabularies.

2.4 Future directions

Infants predominantly hear multi-word utterances. By the end of their first year, infants show great mastery in pulling out word-like units from speech. Most studies cited tested American-English hearing infants; only a few other (European) languages have been considered. It is therefore difficult to ascertain whether the trajectory observed in American-English infants also holds for infants in general (Fernald, 2010). Although it is likely that this development broadly generalizes to all infants, languages might differ in what they consider word-like units (i.e., compare a language like Thai, which has mainly monosyllabic words: Johnson, 2012, to a morphologically complex language like Hungarian: Gervain & Mehler, 2010). One direction for future research would be to see what kind of word-forms infants from more linguistically diverse backgrounds extract from speech, and whether their development also reflects increased reliance on language-specific cues. Notice further that even in the well-studied case of English, computational models do not agree on the units extracted from natural IDS (e.g. trigram phones; syllable or stressed syllable; e.g., Aslin et al., 1996; Bergmann et al., 2013; Curtin et al., 2005; Lignos, 2011; Yang, 2004). Future research should therefore consider the possible units into which infants decompose running speech, and whether this remains the same across development. Indeed, it is even plausible that infants show a transition in the units they segment from speech (i.e., from any type of syllable to only stressed syllable as a possible marker or word-onset; Thiessen & Saffran 2003).

By being able to segment proto-words from speech, infants are well on their way to building a lexicon. Note that not only does a proto-lexicon enhance lexical acquisition, but it can, in turn, also affect phonological development (Martin, Peperkamp & Dupoux, 2013), which has also been linked with later vocabulary development (see also Chapter 1). Hence, it appears that when infants are learning their first words, there are bi-directional links between different levels of spoken language acquisition (i.e., between phones, proto-words, words). This is in line with current popular models of language acquisition (e.g., PRIMIR, Werker & Curtin, 2005). Future research should examine the links between these levels more carefully to better understand differences across infants in early word learning. How do these levels of representation interact? How can we explain why some infants show more advanced speech segmentation skill than others? For instance, do infants with more advanced speech segmentation skill show earlier reliance on language-specific speech segmentation cues than others? If so, does this mirror their phonological development, or can we relate it to the number of words they already understand? The previous two decades has brought us crucial insights in how infants learn to break a speech stream into proto-words, but it remains yet unclear how this development relates to their concurrent phonological or lexical development.

To summarize, this chapter has highlighted the link between early word-form recognition (from fluent speech) and early word learning. Clearly, early word learning greatly progresses when infants are able to pull out possible words from the available input. Future research should aim to explain what drives individual differences in speech segmentation skill.

References

Altvater-Mackensen, N, & Mani, N (2013). Word-form familiarity bootstraps infant speech segmentation. *Developmental science*, *16*(6), 980-990.

Aslin, RN, Woodward, JZ, LaMendola, NP & Bever, TG. (1996). Models of word segmentation in fluent maternal speech to infants. In: J. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition (117-134)*. Mahwah, NJ: Lawrence Erlbaum.

Bergelson, E, & Swingley, D (2012). At 6–9 months, human infants know the meanings of many common nouns. *PNAS*, *109*(9), 3253-3258.

Bergmann, C, Ten Bosch, L, Fikkert, P, & Boves, L (2013). A computational model to investigate assumptions in the headturn preference procedure. *Frontiers in psychology*, *4*.

Bortfeld, H, Morgan, JL, Golinkoff, RM, & Rathbun, K (2005). Mommy and me familiar names help launch babies into speech-stream segmentation. *Psychological Science*, *16*(4), 298-304.

Brent, MR & Siskind, JM (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), B33-B44.

Curtin, S, Mintz, TH, & Byrd, D (2001). Coarticulatory cues enhance infants' recognition of syllable sequences in speech. In *Proceedings of the 25th Annual Boston University Conference on Language Development (Vol. 1*, p. 190). Cascadilla Press.

Curtin, S, Mintz, TH, & Christiansen, MH. (2005). Stress changes the representational landscape: Evidence from word segmentation. *Cognition*, *96*(3), 233-262.

Cutler, A (2012). *Native listening: Language experience and the recognition of spoken words*. MIT Press.

Cutler, A, & Carter, DM (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, *2*(3), 133-142.

Depaolis, RA, Vihman, MM & Keren-Portnoy, T. (2014) When do infants begin recognizing familiar words in sentences? *Journal of child language*, *41(01)*, 226-239.

Fernald, A (1985). Four-month-old infants prefer to listen to motherese. *Infant behavior and development*, *8*(2), 181-195.

Fernald, A (2010). Getting beyond the "convenience sample" in research on early cognitive development. *Behavioral and Brain Sciences, 33,* 91-92

Garcia-Sierra, A, Rivera-Gaxiola, M, Percaccio, CR, Conboy, BT, Romo, H, Klarman, L, Ortiz, S. & Kuhl, PK (2011). Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. *Journal of Phonetics*, *39*(4), 546-557.

Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual review of psychology*, *61*, 191-218.

Gonzalez-Gomez, N, & Nazzi, T (2013). Effects of prior phonotactic knowledge on infant word segmentation: the case of nonadjacent dependencies. *Journal of Speech, Language, and Hearing Research, 56(3),* 840-849.

Goyet, L, de Schonen, S, & Nazzi, T (2010). Words and syllables in fluent speech segmentation by French-learning infants: An ERP study. *Brain Research*, *1332*, 75-89.

Graf Estes, K, Evans, JL, Alibali, MW, & Saffran, JR (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, *18*(3), 254-260.

Hay, JF., & Saffran, JR (2012). Rhythmic grouping biases constrain infant statistical learning. *Infancy, 17(6),* 610-641.

Hay, JF, Pelucchi, B, Estes, KG., & Saffran, JR (2011). Linking sounds to meanings: infant statistical learning in a natural language. *Cognitive psychology*, *63*(2), 93-106.

Met opmaak: Nederlands (standaard)

Met opmaak: Nederlands (standaard)

Houston, DM, Jusczyk, PW, Kuijpers, C, Coolen, R, & Cutler, A (2000). Cross-language word segmentation by 9-month-olds. *Psychonomic Bulletin & Review*, *7*(3), 504-509.
Houston-Price, C, & Nakai, S (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development*, *13*(4), 341-348.
Johnson, EK (2012). Bootstrapping language: Are infant statisticians up to the job? In P. Rebuschat & J. Williams (Eds.). *Statistical learning and language acquisition* (pp.55-89). Mouton de Gruyter.
Johnson, EK, & Jusczyk, PW (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of memory and language*, *44*(4), 548-567.
Johnson, EK, Jusczyk, PW, Cutler, A, & Norris, D (2003). Lexical viability constraints on speech segmentation by infants. *Cognitive Psychology, 46(1),* 65-97.
Johnson, EK, Lahey, M, Ernestus, M., & Cutler, A (2013). A multimodal corpus of speech to infant and adult listeners. *The Journal of the Acoustical Society of America*, *134*(6), EL534-EL540.
Johnson, EK, & Tyler, MD (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science, 13(2),* 339-345.
Junge, C, & Cutler, A (2014). Early word recognition and later language skills. *Brain sciences*, *4*(4), 532-559.
Junge, C, Cutler, A, & Hagoort, P (2012). Electrophysiological evidence of early word learning. *Neuropsychologia*, *50*(14), 3702-3712.
Junge, C, Cutler, A, & Hagoort, P (2014). Successful Word Recognition by 10-Month-Olds Given Continuous Speech Both at Initial Exposure and Test. *Infancy*, *19*(2), 179-193.
Junge, C, Kooijman, V, Hagoort, P, & Cutler, A (2012). Rapid recognition at 10 months as a predictor of language development. *Developmental science*, *15*(4), 463-473.
Jusczyk, PW (2000). *The discovery of spoken language*. MIT press.
Jusczyk, PW & Aslin, RN (1995). Infants′ detection of the sound patterns of words in fluent speech. *Cognitive psychology*, *29*(1), 1-23.
Jusczyk, PW, Houston, DM, & Newsome, M (1999). The beginnings of word segmentation in English-learning infants. *Cognitive psychology*, *39*(3), 159-207.
Kooijman, V, Hagoort, P, & Cutler, A (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cognitive Brain Research*, *24*(1), 109-116.
Kooijman, V, Hagoort, P, & Cutler, A (2009). Prosodic structure in early word segmentation: ERP evidence from Dutch ten-month-olds. *Infancy, 14(6),* 591-612.
Kooijman, V, Junge, C, Johnson, EK, Hagoort, P, & Cutler, A (2013). Predictive brain signals of linguistic development. *Frontiers in psychology*, *4*.
Kudo, N., Nonaka, Y, Mizuno, N, Mizuno, K, Okanoya, K (2011). On-line statistical segmentation of a non-speech auditory stream in neonates as demonstrated by event-related brain potentials. *Developmental Science 14*, 1100–1106
Kuhl, PK (2004). Early language acquisition: cracking the speech code. *Nature reviews neuroscience, 5(11),* 831-843.
Lany, J, & Saffran, JR (2010). From statistics to meaning: infants' acquisition of lexical categories. *Psychological Science*, *21*(2), 284-291.
Lew-Williams, C, Pelucchi, B, & Saffran, JR (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science, 14(6),* 1323-1329.
Lignos, C (2011). Modeling infant word segmentation. In *Proceedings of the fifteenth conference on computational natural language learning* (pp. 29-38). Association for Computational Linguistics.
Männel, C, & Friederici, AD (2013). Accentuate or repeat? Brain signatures of developmental periods in infant word recognition. *Cortex, 49,* 2788-2798.

Martin, A, Peperkamp, S, & Dupoux, E (2013). Learning phonemes with a proto-lexicon. *Cognitive Science*, *37*(1), 103-124.
Mattys, SL, & Jusczyk, PW (2001a). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78(2),* 91-121.
Mattys, SL & Jusczyk, PW (2001b). Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance, 27(3),* 644.
Mattys, S L, White, L, & Melhorn, JF (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *Journal of Experimental Psychology: General*, *134*(4), 477.
Mersad, K, & Nazzi, T (2012). When Mommy comes to the rescue of statistics: Infants combine top-down and bottom-up cues to segment speech. *Language Learning and Development*, *8*(3), 303-315.
Morgan, JL (1996). Prosody and the roots of parsing. *Language and cognitive processes*, *11*(1-2), 69-106.
Nazzi, T, Iakimova, G, Bertoncini, J, Frédonie, S, & Alcantara, C (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, *54*(3), 283-299.
Nazzi, T, Mersad, K, Sundara, M, Iakimova, G, & Polka, L (2014). Early word segmentation in infants acquiring Parisian French: Task-dependent and dialect-specific aspects. *Journal of child language*, *41*(03), 600-633.
Newman, RS, Ratner, NB, Jusczyk, AM, Jusczyk, PW, & Dow, KA (2006). Infants' early ability to segment the conversational speech signal predicts later language development: a retrospective analysis. *Developmental Psychology*, *42*(4), 643-655.
Newman, RS, Rowe, ML & Ratner, NB (2016). Input and uptake at 7 months predicts toddler vocabulary: the role of child-directed speech and infant processing skills in language development. *Journal of child language, 43*(05), 1158-1173.
Ngon, C, Martin, A, Dupoux, E, Cabrol, D, Dutat, M, & Peperkamp, S (2013). (Non) words,(non) words,(non) words: evidence for a protolexicon during the first year of life. *Developmental Science, 16(1), 24-34.*
Parise, E, & Csibra, G (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychological Science 23*, 728-733.
Pelucchi, B, Hay, JF, & Saffran, JR (2009a). Statistical Learning in a Natural Language by 8-Month-Old Infants. *Child Development*, *80*, 674-685
Pelucchi, B, Hay, JF, & Saffran, JR (2009b). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition, 113*(2), 244-247.
Polka, L, & Sundara, M (2012). Word segmentation in monolingual infants acquiring Canadian English and Canadian French: native language, cross-dialect, and cross-language comparisons. *Infancy, 17(2),* 198-232.
Saffran, JR (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition, 81(2),* 149-169.
Saffran, JR, Aslin, RN, & Newport, EL (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928.
Seidl, A, & Johnson, EK (2006). Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental science, 9(6),* 565-573.
Shi, R, & Lepage, M (2008). The effect of functional morphemes on word segmentation in preverbal infants. *Developmental Science, 11(3),* 407-413.
Singh, L, Reznick, SJ, & Xuehua, L (2012). Infant word segmentation and childhood vocabulary development: a longitudinal analysis. *Developmental science, 15(4),* 482-495.

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

Thiessen, ED & Erickson, LC (2012). Discovering words in fluent speech: the contribution of two kinds of statistical information. *Frontiers in psychology, 3*.

Thiessen, ED, Hill, EA., & Saffran, JR (2005). Infant-directed speech facilitates word segmentation. *Infancy*, 7(1), 53-71.

Thiessen, ED, Kronstein, AT & Hufnagle, DG (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological bulletin, 139(4)*, 792.

Thiessen, ED & Saffran, JR (2003). When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental psychology*, *39*(4), 706.

Tincoff, R, & Jusczyk, PW (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, *10*(2), 172-175.

Van Kampen, A, Parmaksiz, G, van de Vijver, R, & Höhle, B (2008). Metrical and Statistical Cues for Word Segmentation: The Use of Vowel Harmony and Word Stress as Cues to Word Boundaries by 6-and 9-Month-old Turkish Learners. In *Language Acquisition and Development: Proceedings of GALA (Vol. 2007*, pp. 313-324).

Werker, JF, & Curtin, S (2005). PRIMIR: A developmental framework of infant speech processing. *Language learning and development*, *1*(2), 197-234.

Yang, CD (2004). Universal Grammar, statistics or both? *Trends in cognitive sciences*, *8*(10), 451-456.

Zamuner, TS, Fais, L, & Werker, JF (2014). Infants track word-forms in early word–object associations. *Developmental science*, *17*(4), 481-491.

[i] This work was supported by a VENI grant (016.154.051) from the Dutch Organisation for Scientific Research (NWO), awarded to Caroline Junge.

Figure 1: A spectrogram of an utterance that Dutch infants will frequently hear *waar is je flesje nou?* ‘where is your bottle then?’. This utterance contains no silences (otherwise there would have been moments where there was no signal): there is speech from beginning to end. The dotted lines correspond to word boundaries – demonstrating that words are not demarcated by silences.

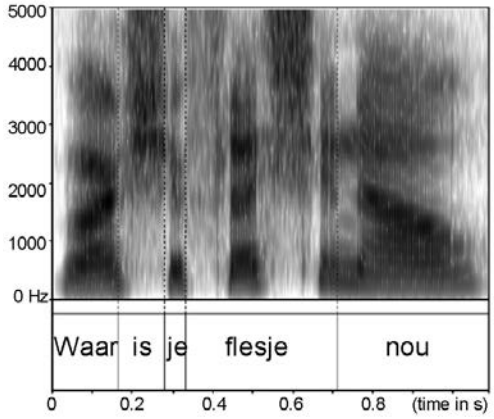

Figure 2: The current set-ups and main findings of the two methods to test infant segmentation skill. I – the Headturn Preference procedure (HPP): *left,* a schematic representation of how testing in lab setting occurs (adapted from Bergmann et al., 2013); *right,* the two dominant infant responses at the test phase: mean listening times to familiar (black) vs. novel words (grey) are typically larger for studies with native language stimuli (here, Jusczyk & Aslin, 1995) but shorter for studies with artificial language stimuli (here, Saffran et al. 1996). II, the event-related potential (ERP) procedure: *left,* a typical set-up in which a 10-month-old is presented with auditory stimuli while she is watching a screensaver, *right,* the two dominant infant responses at the test phase: ERP results for familiar words (solid line) vs. novel words (dotted line) are more negative-going for 10-month-olds (Kooijman et al., 2005) whereas they are positive-going for 7-month-olds (Kooijman et al., 2013). Note that negativity is plotted upwards, and that time is 0 denotes critical word onset.

Figure 3: Results from two studies on word familiarization in 9-10-month-olds, differing in the context in which words are repeated: *left,* repetition of low-frequency words embedded in eight different sentences (cf. Junge et al., 2014); *right,* repetition of typical early words paired with six matching images (e.g., word 'dog' paired with picture of a dog, cf. Junge et al., 2012). The plot shows that in both studies, the negativity corresponding to the N200-500 becomes larger for each subsequent pair of repetitions, relative to the first two tokens.