# Speaker intent influences infants' segmentation of potentially ambiguous utterances

**Elizabeth K. Johnson**
The Johns Hopkins University, Baltimore, MD
E-mail: zab@jhu.edu, elizabeth.johnson@mpi.nl

## ABSTRACT

This study investigated infants' use of subphonemic cues to speaker intent to segment potentially ambiguous utterances (when all other cues to word boundaries are held constant). In Experiment 1, the HPP was used to familiarize English-learning 12-month-olds to one passage containing a S#WS target (e.g. "rue#bequest"), and one passage containing a SW#S target (e.g. "dogma#lines"). Subsequently, infants' were tested on lists of three types of stimuli: intended ("dogma" if familiarized with "dogma#lines"), unintended ("ruby" if familiarized with "rue#bequest"), and unfamiliar ("gumbo" if familiarized with "dogma#lines" and "rue#bequest"). Twelve-month-olds oriented significantly longer to intended words than unintended or unfamiliar words, demonstrating their use of subphonemic cues. Experiment 2 differed from Experiment 1 in only one respect: infants were tested on the weak-strong (WS) portion of the strings ("bequest"from "rue#bequest"/ "ruby#quest"). In this case, 12-month-olds showed weak evidence of segmenting any words, intended or not. Possible explanations for these results are discussed.

## 1. INTRODUCTION

English-learning infants as young as 7.5 months can segment words from fluent speech [1], however, the strategies they employ are far from mature. Current evidence collected using the Headturn Preference Procedure suggests that the earliest stages of word segmentation can be characterized by an over-reliance on stress and statistical cues to word boundaries. Stressed syllables are perceived as word onsets, and syllables that consistently co-occur are perceived as words [2], [3]. Although reliance on these cues allows infants to make a reasonable first pass at segmenting their input, this strategy also leads infants to make a fair number of predictable errors. For example, Jusczyk et al. (1999) report that 7.5-month-olds are not capable of segmenting weak-strong (WS) words ("tar" is segmented from "guitar"). Furthermore, if infants are familiarized with passages in which a target WS word is consistently followed by the same monosyllabic unstressed word ("guitar" followed by "is"), 7.5-month-olds will reliably mistake the reoccurring SW item as a word (e.g. "taris"is segmented from "guitar is"). Fortunately, as infants gain more experience with their native language, their segmentation abilities become increasingly accurate. For example, Jusczyk et al. (1999) reported that 10.5-month-olds reliably segment WS words, even in the face of misleading statistical cues ("guitar"is segmented from reoccurring instances of "guitar is").

An important question which remains to be addressed is how infants overcome their earlier reliance on statistical and prosodic cues to word boundaries. There are at least two classes of explanation for Jusczyk et al.'s finding that 10.5-month-olds can segment WS words in the face of misleading statistical cues. The first class attributes infants' maturing segmentation skills to an increased reliance on lexical knowledge. Although 10.5-month-olds do not have a large lexicon filled with meaningful words, they have started to pick up on the patterning of commonly occurring function words [4], [5]. Therefore, by 10.5 months, infants may possess the necessary prerequisites to use what Christophe et al. have termed the "function-word stripping" strategy [6]. In short, this strategy involves using known function (and sometimes auxiliary) words to break up longer utterances containing one or more unknown words (e.g. if "is" is a word, then "taris" probably isn't a word). This strategy could explain Jusczyk et al.'s results because all of their spurious test items contained embedded high frequency functors or auxiliaries ("taris" from "guitar is", "rayon" from "beret on", "viceto" from "device to", and "risin" from "surprise in"). However, in addition to "function-word stripping", there is another class of explanation for Jusczyk et al.'s finding that seems equally plausible. Infants may learn to use information about the sound patterning of their language to segment WS words. For example, speakers often produce subphonemic cues that can be used to disambiguate the intended segmentation of otherwise ambiguous utterances (e.g. "no#notion" versus "known#ocean") [7]. It is possible that infants could use this sort of subphonemic information to overcome their earlier reliance on statistical and prosodic cues to word boundaries. This hypothesis is supported by evidence that infants use subphonemic (allophonic) cues to syllable boundaries to segment fluent speech [8].

The goal of this study is to further understand how 10.5-month-olds know that not all reoccurring trochaic sequences are words (e.g. "taris" from "guitar is" is not a word). The study investigates whether or not infants are able to avoid mistaking spurious words for intended words by using acoustic-phonetic cues to word boundaries. To date, there is no convincing experimental evidence to


ISBN 1-876346-48-5 © 2003 UAB

distinguish between the alternative explanations discussed above. The following two experiments will attempt to shed light on this issue by testing infants'ability to detect the location of intended word boundaries in the absence of disambiguating statistical, lexical stress, phonotactic, and lexical (function word) cues to word boundaries. In short, the aim is to see whether infants are sensitive to subphonemic cues to word boundaries.

## 2. EXPERIMENT 1

Experiment 1 directly tests the hypothesis that infants can use subphonemic cues to speaker intent to segment potentially ambiguous utterances when all other cues to word boundaries are held constant. Infants are familiarized with two passages: one containing a S#WS target (e.g. "rue#bequest"), and the other containing a SW#S target ("dogma#lines"). Each of these targets is potentially ambiguous in the sense that it can be parsed as either S#WS or SW#S (e.g. "rue#bequest" and "ruby#quest). All infants are tested on the same four test items: "toga", "dogma", "gumbo", and "ruby". A crucial aspect of this experiment is the lack of function words in the test phrases. If infants segment intended items ("ruby" from "ruby#quest") more readily than unintended items ("ruby"from "rue#bequest), then this suggests that infants can use subphonemic cues to segment words from fluent speech.

Participants: Forty-eight English-learning 12-month-olds from the Baltimore-Annapolis region were tested (22 females and 26 males). The infants were approximately 12.0 months old (range 11.5-12.5) with a mean age of 364 days. Parental consent was obtained for all participants.

Simuli: Eight passages containing a two-word sequence with a potentially ambiguous parse were recorded in an infant-directed manner by a female speaker naïve to the purpose of the study. The six-sentence passages were constructed so that a target occurred once in each sentence. The position of the target sequence was varied in the sentences. For example, the first three sentences of the "rue#bequest" passage were as follows: *"This rue bequest will surely go down in history. Rue bequest weddings were common last May. The panda was really wild about that rue bequest."* The first syllable in the word "bequest" was produced so that it rhymed with "tea", and the first syllable of "boutique" was pronounced so that it rhymed with "low". After recording the passages, the speaker also recorded four test lists, each consisting of 15 presentations of different tokens of the same word.

Design: Infants were randomly assigned to familiarization with one of the following four pairs of passages: (1) "rue#bequest"and "dogma#lines", (2) "toga#lore" and "gum#boutique", (3) "ruby#quest" and "dog#maligns", or (4) "toe#galore" and "gumbo#teak". Thus, all infants were familiarized with one passage containing reoccuring tokens of a monosyllabic word followed by a bisylalbic word (S#WS) and one passage containing reoccurring tokens of a bisyllabic word followed by a monosyllabic word (SW#S). All infants were tested on the same four test items: "ruby", "dogma", "gumbo", and "toga". Therefore, each infant was tested on three types of test items: intended familiar ("dogma" if familiarized with "dogma#lines"), unintended unfamiliar ("ruby" if familiarized with "rue#bequest"), and unfamiliar ("gumbo" and "toga" if familiarized with "dogma#lines" and "rue#bequest").

Procedure: Infants were tested using the same modified version of the Headturn Preference Procedure (HPP) used by Jusczyk & Aslin [1], Jusczyk et al. [2], etc. In this procedure, the infant sits on a caregiver's lap in the center of a 3-sided booth which contains three lights mounted at eye-level: a green light on the front panel and a red light on each of the two side panels. Speakers are located behind the red lights. At the start of a given trial the light on the center panel begins to flash. Once the infant orients to the center light, one of the two side lights immediately begins flashing. When the infant orients to the flashing side light, speech begins to play, and continues to play until the infant turns away for more than two consecutive seconds. The experimenter observes the infant through a small peephole, and relays the infants looking behavior via a button box connected to a computer. The computer controls the selection and presentation of stimuli as well as recording the direction and duration of the infants' head turns (the dependent measure is orientation time towards the flashing side lights). Both caregiver and experimenter wear tight head phones playing loud masking music during the entire experiment. This experiment consisted of two phases: familiarization and test. During familiarization, two passages play alternately until the infant accrues 45 seconds of orientation time towards each passage. Twelve test trials were presented during the test phase (three trials for each of the four test items). Trials were blocked, and presented in random order within those three blocks. Half the trials consisted of repetitions (up to 15 per trial) of two familiar bisyllabic sequences (one spanning a word boundary intended by the speaker, and one no spanning a word boundary); the other half consisted of comparable repetitions of two unfamiliar bisyllabic sequences.

Results: Mean orientation times to each of the three types of test items (intended, unintended, and unfamiliar) were calculated for each of the 48 subjects (see Figure 1). Thirty-three out of 48 subjects had longer average orientation times to familiar intended test items than unfamiliar test items. In contrast, only 22 out of 48 infants had longer orientation times to familiar unintended words than unfamiliar words. Planned comparisons revealed a significant difference in orientation times to intended and unfamiliar test items, $F(1, 47) = 6.12$, $p< .05$. In addition, there was a significant difference in orientation times to intended and unintended familiar items, $F(1, 47) = 4.36$, $p < .05$. However, there was no significant difference in orientation times to unintended and unfamiliar test items, $F(1, 47)$, $= .11$, $p > .1$. As Figure 2 illustrates, these effects were attributable to longer orientation time to intended items (M=9.02 seconds, SD=2.9) than either unintended items (M=8.04 seconds, SD=3.1) or unfamiliar items (M=7.89 seconds, SD=2.7). These results demonstrate that infants segment intended words, but not their unintended counterparts. This suggests that infants do indeed use subphonemic cues to word boundaries, and that an explanation for the results of Jusczyk et al. may lie in the fined-grained acoustic-phonetic aspects of the speech signal rather than (or in addition to) infants' knowledge of function words.

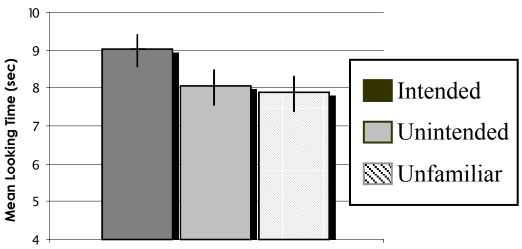


Figure 1: Mean orientation time to intended (e.g. "toga" from "toga#lore"), unintended (e.g. "toga" from "toe#galore"), and unfamiliar test items.

## 3. EXPERIMENT 2

In Experiment 1, infants segmented intended SW words more readily than unintended SW words. This suggests that 12-month-olds are sensitive to subphonemic cues to speaker intent. One way to interpret these findings is to assume that infants parsed the entire SWS phrases into two words in accordance with the speaker's intent. In short, a phrase such as "rue#bequest" was properly parsed into two word-sized constituents: "rue" and "bequest". At the same time, a phrase such as "ruby#quest" was properly parsed into "ruby"and "quest". Although this conclusion is consistent with the findings of Experiment 1, this interpretation needs to be tested. In short, the goal of Experiment 2 is to test the hypothesis that the infants in Experiment 1 segmented the entire potentially ambiguous phrases properly. In Experiment 2, infants are familiarized with the same passages used in Experiment 1, but tested on the iambic (WS) portion rather than the trochaic (SW) portion of the target phrases (e.g. "bequest" rather than "ruby). If infants segment WS items such as "bequest" more readily in an intended ("rue#bequest") than an unintended ("ruby#quest") context, then this lends more strength to the argument that infants use subphonemic cues. to segment the whole phrase properly, and not just part of it.

Participants: Forty-eight English-learning 12-month-olds from the Baltimore-Annapolis region were tested (25 females and 23 males). The infants were approximately 12.0 months old (range 11.5-12.5) with a mean age of 367 days. Parental consent was obtained for all participants.

Simuli, Design, and Procedure: The exact same familiarization passages were used as in Experiment 1. In addition, the exact same familiarization and test procedure were used. Essentially the same design was used as in the first experiment. The crucial difference was the use of weak-strong test items in the current experiment: "galore", "boutique", "maligns", and "bequest". Therefore, each infant was tested on three types of test items: intended familiar ("bequest" if familiarized with "rue#bequest"), unintended unfamiliar ("maligns" if familiarized with "dogma#lines"), and unfamiliar ("boutique" and "galore" if familiarized with "dogma#lines" and "rue#bequest").

Results: Mean orientation times to each of the three types of test items (intended, unintended, and unfamiliar) were calculated for each of the 48 subjects (see Figure 2). Thirty-three out of 48 subjects had longer average orientation times to familiar intended test items than unfamiliar test items. In contrast, only 27 out of 48 infants had longer orientation times to intended words than unintended words. And 30 out of 48 had longer average orientation times towards unintended items than unfamiliar items. Planned comparisons revealed no significant difference in orientation times to intended and unintended familiar items, $F(1, 47) = .09$, $p > .1$. However, there was a marginal difference between orientation times to intended and unfamiliar words, $F(1, 47) = 3.59$, $p < .1$. Similarly, there was a marginal difference between orientation times to unintended and unfamiliar words, $F(1, 47) = 3.57$, $p < .1$. Mean orientation time to intended items (M=9.12 seconds, SD=4.1) was roughly equal to mean orientation to unintended items (M=9.33 seconds, SD=4.01). Orientation time to unfamiliar items (M=8.1, SD=4.0) was shorter than that to intended and unintended items, however the data were too variable for this difference to be significant. As Figure 2 illustrates, it appears that infants segmented the intended and unintended items similarly. Since there was no significant difference between intended and unintended test items, orientation times for these test items were collapsed into general familiar test item category for further analysis. A 2 X 4 mixed design ANOVA with test item (familiar versus unfamiliar) and familiarization condition indicated a significant effect of test item, $F(1, 44) = 4.9$, $p < .05$. There was no significant effect of familiarization condition, $F(3, 44) = .7$, $p > .1$, or interaction between test item and familiarization condition, $F(3, 44) = 1.6$, $p > .1$.

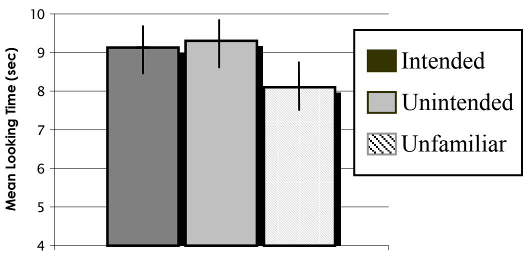


Figure 2: Mean orientation time to intended (e.g. "galore" from "toe#galore"), unintended (e.g. "galore" from "toga#lore"), and unfamiliar test items.


ISBN 1-876346-48-5 © 2003 UAB

These results suggest that infants are not segmenting the entire SWS phrases correctly. In fact, there is very weak evidence that they recognize either the intended or unintended familiar iambic word. Infants' orientation time to the intended test words is only marginally greater than that of their orientation to the unfamiliar test words. Infants'orientation time to familiar test words is only significantly greater than their orientation to unfamiliar test items when mean orientation times to intended and unintended familiar items are collapsed.

## 4. GENERAL DISCUSSION AND CONCLUSIONS

Experiment 1 demonstrated that English-learning 12-month-olds can use subphonemic cues to segment fluent speech (infants segment "ruby" from "ruby#quest", but not from "rue#bequest"). This suggests that infants do indeed use subphonemic cues to word boundaries, and that an explanation for the results of Jusczyk et al.'s (1999) results may lie in the fine-grained aspects of the speech signal rather than (or in addition to) infants' knowledge of function words. This finding suggests that it may be necessary to revise our current model of word segmentation.

The results of Experiment 2 are not as easy to interpret. When tested on the iambic portion of the same potentially ambiguous phrases used in Experiment 1, infants no longer segmented intended words ("bequest" from "rue#bequest") more readily than they segmented unintended words ("bequest" from "ruby#quest). In fact, in Experiment 2, there was weak evidence that infants segmented any words at all. It seems that the subphonemic cues defining the WS words were somehow weaker than the subphonemic cues defining the SW words.

One possible explanation for the differences observed in Experiment 1 and 2 could have to do with the way infants weigh and integrate different cues to word boundaries. The SW words carried word initial stress, and were therefore cued by a prosodic cue. In addition, the SW items were located at the onset of the potentially ambiguous utterances. Therefore, the first syllable of the SW targets was always preceded by a different syllable ("this ruby#quest… that ruby#quest.. singing ruby#quest…."). In short, the first syllables of the SW targets were cued by three types of information: prosodic information (lexical stress), statistical information (transitional probabilities between syllables), and subphonemic information. The onsets of the WS items, on the other hand, were only cued by subphonemic cues. In addition, there were prosodic and statistical information indicating that the first syllable of the WS word belonged to the same word as the stressed syllable that consistently preceded it (e.g. in "rue#bequest", the onset of "bequest" was always preceded by the syllable "rue"). Therefore, if we take multiple cues into account, the WS targets may have been more difficult to segment than the SW items.

Future work will need to explore how infants integrate and weigh various cues to word boundaries at different ages, and in different environments (e.g. a noisy preschool versus a sound-proof booth). Although research in this area has just begun, the results have revealed a great deal about how word segmentation develops [9], [10].

## 5. Acknowledgements

I thank Joane Miller and Stefanie Shatuck-Hufnagel for research advice leading to the completion of this project, and Ann Marie Jusczyk for help in running and recruiting subjects. Amanda Seidl, Sarah Shomstein, and Andy Leber deserve thanks for helpful comments on a previous draft of this document.

## REFERENCES

[1] P.W. Jusczyk, and R.N. Aslin. "Infants' detection of sound patterns of words in fluent speech", *Cognitive Psychology*, 29, pp. 1-23, 1995.

[2] P.W. Jusczyk, D. Houston, M. Newsome (1999). "The beginnings of word segmentation in English-learning infants." *Cognitive Psychology*, 39, 159-207.

[3] J.R. Saffran, R.N. Aslin, E.L. Newport (1996). "Statistical learning by 8-month-old infants". *Science*, 274, pp. 1926-1928.

[4] V.L. Shafer, J.L. Shucard, L.A. Gerken, D.W. Shucard (1998). "The'and the brain: An electrophysiological examination of infants' responses to familiary and unfamiliar function morphemes". *J. Speech, Language and Hearing Research*, 41, 874-886.

[5] M. Shady, P.W. Jusczyk, L.A. Gerken (1998). "Infants' sensitivity to function morphemes". *Proceedings of the 23rd Annual BUCLD*. Somerville, MA: Cascadilla Press.

[6] A. Christophe, M.T. Guasti, M. Nespor, E. Dupoux, B. Van Ooyen (1997). "Reflections on prosodic bootstrapping: its role for lexical and syntactic bootstrapping". *Lang. & Cog. Processes*, 12, 585-612.

[7] L.H. Nakatani and K.D. Dukes (1977). Locus of segmental cues for word juncture. JASA, 63, 234-245.

[8] S. Mattys and P. Jusczyk (2001). "Do infants segment words or recurring contiguous patterns?" Journal of Experimental Psychology: HPP, 27, 644-655.

[9] S.L. Mattys, P.W. Jusczyk, P.A. Luce, J.L. Morgan (1999). Phonotactic and prosodic effects on word segmentation in infants. Cognitive Psychology, 38, 465-494.

[10] E.K. Johnson and P.W. Jusczyk (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. JML, 44, 1-20.