Statistical Learning in a Natural Language by 8-Month-Old Infants

Bruna Pelucchi,
University of Wisconsin-Madison and University of Ferrara

Jessica F. Hay, and
University of Wisconsin-Madison

Jenny R. Saffran
University of Wisconsin-Madison

Abstract

Numerous studies over the past decade support the claim that infants are equipped with powerful statistical language learning mechanisms. The primary evidence for statistical language learning in word segmentation comes from studies using artificial languages, continuous streams of synthesized syllables that are highly simplified relative to real speech. To what extent can these conclusions be scaled up to natural language learning? In the current experiments, English-learning 8-month-old infants' ability to track transitional probabilities in fluent infant-directed Italian speech was tested ($N = 72$). The results suggest that infants are sensitive to transitional probability cues in unfamiliar natural language stimuli, and support the claim that statistical learning is sufficiently robust to support aspects of real-world language acquisition.

Identifying word boundaries in continuous speech seems like an easy task for native adult listeners. However, the problem of word segmentation is potentially very challenging for infants: words are not consistently delimited by silences (Cole & Jakimik, 1980). Fortunately, there are other forms of information embedded in speech that can be useful markers to word boundaries, and infants can exploit myriad segmentation cues by the time they are 9 months of age, including phonotactic regularities (Friederici & Wessels, 1993; Mattys & Jusczyk, 2001), prosodic patterns (Jusczyk, Cutler, & Redanz, 1993; Jusczyk, Houston, & Newsome, 1999; Morgan, 1996), and allophonic variation (Christophe, Dupoux, Bertoncini, & Mehler, 1994; Jusczyk, Hohne, & Bauman, 1999).

The above-mentioned cues are, however, all language specific. Thus, to make use of them, infants must already know something about the sound patterning of their native language, particularly with respect to correlations between sound patterns and word boundaries. For example, although 5-month-old infants can discriminate between strongly and weakly stressed syllables (Weber, Hahne, Friedrich, & Friederici, 2004), this information is not helpful in word segmentation until infants learn about the distribution of stressed syllables relative to word onset–offsets in their native language. This requires that infants first learn something about which sound patterns constitute words in their native language.

How might infants achieve this initial segmentation? Importantly, it would not need to be complete or fully accurate; infants would merely need to segment a subset of word-like units

Correspondence concerning this article should be addressed to: Bruna Pelucchi, Waisman Center, University of Wisconsin-Madison, 1500 Highland Ave, Madison, WI 53705. Electronic mail may be sent to plb@unife.it.

from which to glean other language-specific regularities. Some of this may be achieved via words presented in isolation (Brent & Siskind, 2001), or by proximity to known highly frequent words (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005). Infants might also exploit statistical regularities to bootstrap initial word forms, from which other regularities can be discovered (Thiessen & Saffran, 2003).

One type of statistical regularity available to infant learners is the transitional probability (TP) between syllable sequences. TP, also termed conditional probability, is the probability of one event (e.g., of one syllable) given the occurrence of another event. This statistic refers to more than the frequency with which one element follows another, as it adjusts for the base rate of the first event or element. The TP of *Y* given *X* is represented by the following equation:

$$\text{TP} = P(Y|X) = \frac{\text{frequency}(XY)}{\text{frequency}(X)}.$$

Corpus analyses suggest that the TPs between syllables are an imperfect but potentially useful cue to word boundaries in natural speech (Swingley, 2005; though see Yang, 2004, for a different view). Infants are sensitive to the probabilities of sound co-occurrences in both speech (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996) and nonspeech domains (Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002; Saffran, Johnson, Aslin, & Newport, 1999). The primary evidence supporting the existence of statistical learning mechanisms in infants comes from studies employing artificial language materials. Infants are typically exposed to a synthetically produced speech stream, without pauses or other acoustic cues to word boundaries. The only available cue is a dip in the TPs (and other related sequential statistics, such as mutual information) between syllables and/or segments at word boundaries. Infants are then tested on their ability to discriminate sequences corresponding to *words* versus *nonwords* (syllables from the language assembled in a novel order), or a more subtle comparison, *words* versus *part-words* (syllable sequences spanning word boundaries, which can be matched for frequency in the speech stream; Aslin et al., 1998).

Although artificial language materials have been invaluable for the initial investigation of infant statistical learning mechanisms, it is obvious that such stimuli lack the complexity of a natural language on virtually every possible dimension. This problem of ecological validity has been acknowledged throughout the literature on infant statistical language learning. For example, the set of phonemes and syllables used in any given experiment is highly circumscribed, there are very few words (typically just four), and words are repeated extremely frequently during exposure (45–90 times). Furthermore, there are typically no other sequential regularities present, such as those engendered by syntactic structure (though see Saffran & Wilson, 2003). The materials are isochronous (devoid of rhythmic patterning), lack pitch changes or any other acoustic variability, and are generally stripped of all other potentially relevant (or distracting) cues found in natural speech.

As a consequence, it is important to ask to what extent results obtained using artificial languages can be applied to real-world language learning. Are infants able to track TPs when faced with input that has the complexity of natural language? One way to address this question is to add progressively more complexity to an artificial segmentation task, to approach a natural language by successive approximations (Curtin, Mintz, & Christiansen, 2005; Sahni, Saffran, & Seidenberg, 2008; Thiessen, Hill, & Saffran, 2005; Thiessen & Saffran, 2003, 2007; Tyler & Johnson, 2006). While research using this technique has been very informative, these artificial languages are necessarily still quite simplistic when

compared with natural language. Another approach, which we adopt in this line of research, is to present infants with an unfamiliar natural language. By doing so, we can maintain virtually all of the complexity of natural language, while still manipulating the statistical patterns of interest in a small subset of the words in the language. Our goal was to test the hypothesis that infants can exploit TPs when faced with natural speech. Indirect evidence suggesting that infants can do so comes from a study by Jusczyk, Houston, et al. (1999), who showed that infants can use distributional information in combination with lexical stress cues in native language natural speech materials. In the current studies, we bring together statistical learning studies and natural language materials to ask whether infants can track statistical patterns in an unfamiliar language.

To this end, we designed a statistical segmentation task similar to the study by Saffran et al. (1996), using naturally produced, grammatically correct, and semantically meaningful Italian in lieu of an artificial language. Although Italian (De Mauro, Mancini, Vedovelli, & Voghera, 1993; Mancini & Voghera, 1994) shares the strong–weak stress patterning characteristic of English speech (Cutler & Carter, 1987), the allophonic and phonotactic regularities found in the two languages are quite different, as are other rhythmic properties. We thus expected Italian materials to be quite novel for our English-learning infants. The Italian corpus maintained virtually all of the complexities found in natural speech— arguably, the only substantial difference between the materials used in this experiment and 'actual' natural language is that the TPs between syllable sequences were expressly manipulated in a subset of the words.

Given the vast difference in complexity between our materials and the artificial languages used in the past, we designed Experiment 1 to determine whether infants could segment anything at all from a brief exposure to Italian. Previous segmentation studies have shown recognition of words in fluent speech after exposure to words in isolation in typologically related languages, such as English and Dutch (Houston, Jusczyk, Kuijpers, Coolen, & Cutler, 2000). However, infants' ability to recognize isolated words after exposure to fluent speech in a foreign language remains unexplored—particularly given language pairs that are typologically distinct. After familiarization with a fluent speech stream in Italian, 8-month-old infants were tested on *familiar words* versus *novel words* (Italian words which never appeared in the speech stream). Successful discrimination would suggest that English-learning infants can detect novel words in an unfamiliar language.

Experiment 1

Following the method of the landmark study by Jusczyk and Aslin (1995), infants were familiarized with a set of Italian sentences and subsequently tested on familiar words (Italian disyllabic words presented during familiarization) versus novel words (actual Italian words that were not presented during familiarization). Successful discrimination would suggest that infants could recognize words previously heard in fluent speech in a novel language.

Method

Participants—Twenty infants (11 male, 9 female) with a mean age of 8.5 months (range = 8.1–9.0) participated in Experiment 1. Although statistical learning has been studied in humans (and nonhumans) of a variety of ages, the original artificial language studies (Saffran et al., 1996) employed 8-month-old infants. Thus, we chose to study infants in the same age range to make our results from natural language speech segmentation comparable with those from artificial language studies. All infants were full-term monolingual English learners with no history of hearing or vision impairments. None of the infants had prior exposure to Italian or Spanish.

Infants were recruited from a local birth announcement database maintained by the Waisman Center at the University of Wisconsin–Madison. Gender, ethnicity, race, and socioeconomic status (SES) of all participants were representative of the demographics of Madison, Wisconsin (male: 50.5%; female: 49.5; Caucasian: 89.5%; more than one group: 5.6%; Hispanic: 4.0%; Asian American: 0.7%; African American: 0.3%).

Participants were assigned to one of two counterbalanced language conditions: Language 1A and Language 1B. Eighteen additional infants were tested and excluded for the following reasons: fussiness (14), experimental error (3), and not paying attention (1). Two additional infants showed looking time preferences > 3 *SD* from the mean (one in each language group, with preferences in opposite directions), and were excluded from the analyses.

Apparatus and stimulus materials—Four Italian words with a strong–weak stress pattern were selected for use in this study: *fuga, melo, pane*, and *tema* (see Table 1). Although these words were phonetically legal in English, the passages in which they were presented contained non-English phonetic features (e.g., a trill, a voiced alveolar affricate, and a palatal nasal).

We created two counterbalanced languages to control for arbitrary listening preferences at test. Language 1A consisted of three identical blocks of 12 grammatically correct and semantically meaningful standard Italian sentences (see the Appendix for sentence lists). These sentences contained the words *fuga* and *melo*, which both occurred six times in each block of 12 sentences. The component syllables of *fuga* and *melo* never appeared without each other (i.e., *fu* never appeared in the absence of *ga*, and vice versa).

Recall that the TP of, for example, *fuga* corresponds to:

$$\mathrm{TP}(ga|fu) = \frac{f(fuga)}{f(fu)}.$$

Because *fu* never appeared without *ga*, the internal TP of *fuga* (and of *melo)* was 1.0. Two other words, *pane* and *tema*, and their component syllables, were never presented in the Language 1A familiarization passages (TP = 0). In the counterbalanced Language 1B, *pane* and *tema* each occurred each six times per block (TP = 1.0), while *fuga* and *melo* (and their component syllables) never occurred (TP = 0). This design is thus exactly analogous to the original Jusczyk and Aslin (1995) study.

To ensure that the sentences from each of the counterbalanced languages were acoustically matched in intensity, length, and speech rate, small acoustic modifications were made to the original recordings. Using the "stretch" algorithm in Adobe® Audition® (Adobe System Corporation, San Jose, CA), the speech rate was adjusted (either up or down) so that the target language was presented at an average rate of six syllables per second. The resulting sentences were then intensity-matched in Adobe® Audition® to be presented at approximately 60 dB$_{SPL}$. Each block of sentences (44 s) was presented three times during familiarization, with 18 presentations of each target word, for a total duration of 2 min 14 s.

The stimuli were recorded by a female native Italian speaker from Florence, who was naïve to the purpose of the experiment. She was asked to read the stimuli in a lively voice, pretending to be in front of a baby. After recording the passages, the speaker read the four words in citation form for use as test items. The isolated words were also digitally edited in Adobe® Audition® to have the same length (750 ms) and amplitude (65 dB$_{SPL}$), while preserving their original pitches.

Procedure—Infants were tested using the Head Turn Preference Procedure as adapted by Saffran et al. (1996). Participants were seated on a caregiver's lap inside a sound-attenuated booth that was equipped with one central light and two laterally placed speakers and sidelights; parents listened to music over headphones. The experimenter observed the infant's head turns over a closed-circuit TV camera outside the test booth. At the beginning of the familiarization phase, a light in the center of the wall facing the infant began to flash, directing the infant's gaze forward. Simultaneously, one of the two languages (Language 1A or Language 1B) began to play from the speakers beneath the two sidelights in the room. The lights flashed contingent on looking behavior (as in the test phase described next), while the familiarization materials played continuously.

Immediately after familiarization, 12 test trials were presented. All infants heard the same test items regardless of familiarization condition. Each of the two familiar word trials and the two novel word trials occurred three times, randomized by block. Trials that contained familiar words for infants in the Language 1A condition were novel for infants in the Language 1B condition, and vice versa. Test trials began with a center blinking light. When the observer signaled the computer that the infant had fixated the center light, one of the sidelights began to flash, and the center light was extinguished. When the infant made a head turn of at least 30° in the direction of the side-light, the experimenter signaled the computer to play a test item from the speaker beneath the flashing light. The test item continued to play until the infant looked away for more than 2 s, or when a maximum looking time of 15 s was reached. The test item then stopped playing, the sidelight was extinguished, and the center light began to blink. This procedure was repeated until the infant had completed all 12 test trials. Trials with total looking time < 1 s were automatically repeated at the end of the test session.

Results and Discussion

We first compared the two counterbalanced familiarization conditions. A t test (all t tests reported are two-tailed) comparing the difference scores from the two counterbalanced languages revealed no significant differences, $t(18) = .66$, $p = .51$, $p_{rep} = .49$ (see Killeen, 2005), suggesting that there were no a priori listening preferences for any of the test words. The two conditions were thus combined in the subsequent analysis. The average looking time was 6.35 s ($SE = .46$) for novel words and 8.21 s ($SE = .58$) for familiar words (see Figure 1); 18 of the 20 infants listened longer to the familiar words. A paired t test revealed a significant difference in looking times between familiar and novel words: $t(19) = 5.63$, $p < .001$, $p_{rep} > .999$.

The observed direction of preference, a familiarity preference, differs from some prior statistical learning studies in which a novelty preference was observed (Aslin et al., 1998; Saffran et al., 1996). However, a familiarity preference is consistent with the large body of experiments using natural language segmentation tasks, beginning with Jusczyk and Aslin (1995). The combination of relatively few repetitions of each target word during exposure (especially when compared with artificial language tasks), paired with the rich natural materials, is likely responsible for the observed direction of preference.

The significant preference for familiar words suggests that infants were able to discriminate the items presented during familiarization from novel words, despite the use of foreign language materials. These results parallel Houston et al.'s (2000) findings of cross-linguistic segmentation of prefamiliarized words using English and Dutch, and extend them to languages that are more typologically distinct: English and Italian.

However, there is a potential alternative explanation for these results. Rather than tracking patterns of co-occurrence of syllables, it is possible that infants succeeded on the

discrimination task by detecting the frequencies of individual syllables. The familiar words contained familiar syllables, whereas the syllables from the novel words never occurred in the familiarization corpus. To demonstrate that infants are tracking syllable sequences in a novel language, it is necessary to rule out the possibility that the results of Experiment 1 were driven by syllable-level familiarity alone. To test this alternative hypothesis, we designed Experiment 2. This experiment is a conceptual replication of Experiment 1 using familiar and novel words matched for syllable frequency.

Experiment 2

In this experiment, infants were again familiarized with a set of Italian sentences and then tested on familiar words (Italian disyllabic words presented during familiarization) versus novel words (actual Italian words that were not presented during familiarization). This time, however, the syllables of the familiar and novel words appeared in the familiarization corpus with equal frequency. We expect that if infants are tracking syllable sequences in natural speech, rather than the frequencies of individual syllables, they should still show successful discrimination.

Method

Participants—Twenty infants (9 male, 11 female) with a mean age of 8.4 months (range = 8.1 to 9.0 months) participated in Experiment 2. Infants were recruited from a local birth announcement database maintained by the Waisman Center at the University of Wisconsin–Madison. Gender, ethnicity, race, and SES of all participants were representative of the demographics of Madison, Wisconsin (male: 50.5%; female: 49.5%; Caucasian: 89.5%; more than one group: 5.6%; Hispanic: 4.0%; Asian American: 0.7%; African American: 0.3%).

Participants were assigned to one of two counterbalanced language conditions: Language 2A and Language 2B. Ten additional infants were tested but not included in the analyses for the following reasons: fussiness (6), not paying attention (1), or experimental error (3).

Apparatus and stimulus materials—As in Experiment 1, there were two counterbalanced familiarization corpora (see the Appendix for sentence lists). The four target words were the same used in Experiment 1; *fuga, melo, pane*, and *tema*. In Language 2A, two of the words, *fuga* and *melo*, appeared six times each in the block of 12 sentences (see Table 1). The other two words, *pane* and *tema*, never occurred in the corpus. However, their syllables, *pa*, *ne*, *te*, and *ma*, each appeared six times in the corpus, with *pa* and *te* always occurring in stressed position and ne and ma always occurring in unstressed position (as in the target words). Language 2B had the same structure, but the familiar and the novel words were switched. Items that served as familiar words for infants in the Language 2A condition were novel words for infants in the Language 2B condition, and vice versa. Thus, familiar words always had a TP of 1.0, whereas novel words always had a TP of zero. Stimuli were recorded and edited as in Experiment 1. Target word length, amplitude, and pitch were similar across the two counterbalanced languages (see Table 2).

Procedure—The procedure was identical to Experiment 1.

Result and Discussion

As in Experiment 1, we first compared the two exposure conditions. A *t* test comparing the difference scores from the two counterbalanced languages revealed no significant differences, $t(18) = .42$, $p = .68$, $p_{rep} = .38$. The two conditions were thus combined in the subsequent analysis. The average looking time was 7.94 s ($SE = .44$) for novel words and

9.08 s (*SE* = .66) for familiar words (see Figure 2). Fifteen of the 20 infants looked longer to familiar words. A paired t test revealed a significant difference between the looking times for familiar and novel words: $t(19) = 2.18$, $p < .05$, $p_{rep} = .89$. The significant preference for familiar words suggests that infants were able to discriminate the items presented during familiarization from novel words, even when the individual syllables in the familiar and novel words occurred with equal frequency in the exposure corpus. This preference suggests that infants were not just keeping track of syllable frequency, but were able to discriminate familiar pairings of familiar syllables from novel pairings of familiar syllables.

Given exposure to natural speech, infants do not solely attend to individual syllables. Instead, they track sequences of syllables—a necessary ability if they are to discover word units in fluent speech. However, Experiments 1 and 2 were not designed to determine *which* aspects of the sequences are being tracked. The familiar words had an internal TP = 1.0 and were quite frequent, whereas the novel words had a TP of zero and never occurred in the corpus. It is possible that the results of Experiments 1 and 2 were driven by the frequency of the sequences, rather than their internal probabilities. We thus designed Experiment 3 to assess a more difficult discrimination, in which infants were tested on words occurring with equal frequency during familiarization, differing only in their internal TPs.

Experiment 3

In this study, infants were again familiarized with a set of Italian sentences. However, instead of being tested on familiar versus novel words, we tested them on two different types of familiar words: *high-transitional-probability words* (HTP-words, Italian words from the language with TP = 1.0) versus *low-transitional-probability words* (LTP-words, Italian words with TP = .33). Importantly, both word types occurred equally often during familiarization. Successful discrimination would suggest that infants are sensitive to probability information in natural language stimuli drawn from an unfamiliar language.

Method

Participants—Thirty-two infants (16 male, 16 female) with a mean age of 8.5 months (range = 8.0 to 9.0 months) participated in Experiment 3. Infants were recruited from a local birth announcement database maintained by the Waisman Center at the University of Wisconsin–Madison. Gender, ethnicity, race, and SES of all participants were representative of the demographics of Madison, Wisconsin (male: 50.5%; female: 49.5%; Caucasian: 89.5%; more than one group: 5.6%; Hispanic: 4.0%; Asian American: 0.7%; African American: 0.3%).

Participants were assigned to one of two counterbalanced language conditions: Language 3A and Language 3B. Twenty-two additional infants were tested but not included in the analyses for the following reasons: fussiness (20), not paying attention (1), or experimental error (1).

Apparatus and stimulus materials—As in Experiments 1 and 2, there were two counterbalanced familiarization languages (see the Appendix for sentence lists). Language 3A was identical to Language 1A from Experiment 1. The four target words appeared with the same frequency, each occurring six times in the block of 12 sentences: *fuga, melo, bici*, and *casa* (see Table 1). These target words all followed a strong–weak stress pattern and were phonetically and phonotactically legal in English.

Although the two pairs of words, *fuga–melo* and *casa–bici*, were equally frequent, they contained different internal TPs. As in Experiments 1 and 2, the syllables *fu, ga, me*, and lo appeared only in the context of the word *fuga* and *melo*. Consequently, the TP of these two

words was 1.00 (HTP-words) in Language 3A. However, for *casa* and *bici* there were 12 additional occurrences of the syllables *ca* and *bi* in each block of sentences. *Ca* and *bi* thus occurred a total of 18 times in each block, all in strong (stressed) position. As a consequence, the TPs of *casa* and *bici* were .33 (LTP-words), relative to the Language 3A familiarization sentences. The counterbalanced familiarization sentences, Language 3B, had the same structure: the four target words were equally frequent but contained different TPs: *casa* and *bici* had a TP of 1.00 (HTP-words) while *fuga* and *melo* had a TP of .33 (LTP-words). As in Experiments 1 and 2, each block of 12 sentences was presented three times during familiarization. The resulting language corpus thus included a total of 18 presentations of each of the four target words, and 36 additional occurrences of the first syllables in the two LTP-words.

The test words in Experiment 3 were *fuga*, *melo*, *casa*, and *bici*. Items that served as *HTP-words* for infants in the Language 3A condition were *LTP-words* for infants in the Language 3B condition, and vice versa. Stimuli were recorded and edited as in Experiments 1 and 2. Target word length, amplitude, and pitch were similar across the two counterbalanced languages (see Table 2).

Procedure—The procedure was identical to Experiments 1 and 2.

Result and Discussion

We first compared the two exposure conditions. A *t* test comparing the difference scores from the two counterbalanced languages revealed no significant differences, $t(30) = .64$, $p = .53$, $p_{rep} = .48$, suggesting that there were no a priori listening preferences for any of the words. The two conditions were thus combined in the subsequent analysis. The average looking time was 7.71 s (*SE* = .31) for LTP-words and 8.75 s (*SE* = .36) for HTP-words (see Figure 3). Twenty-six of the 32 infants looked longer to HTP-words. A paired *t* test revealed a significant difference between the looking times for HTP-words and LTP-words: $t(31) = 3.94$, $p < .001$, $p_{rep} > .99$. This preference suggests that infants are sensitive to distributional statistics in natural speech. The HTP- and LTP-words occurred with the same frequency during familiarization. Moreover, these items shared the same (strong–weak) stress pattern. Infants thus appear to have discriminated between the test items based on the sequential statistics of the sounds in the speech stream.

Infants may also be tracking segment-level TPs (Newport, Weiss, Wonnacott, & Aslin, 2004). In Language 3A, the segment-level TPs for the LTP-words were comparable with the syllable-level TPs. In Language 3B, the segment-level TPs for the LTP-words were slightly higher than the syllable-level TPs. Task difficulty should be comparable, or slightly harder, given segment-level TPs relative to syllable-level TPs.

As with Experiment 1, we considered the alternative hypothesis that the results of Experiment 3 reflect infants' attention to syllable frequency, rather than the statistics of syllable sequences. The first syllable of each LTP-word was 3 times more frequent than the first syllable of the HTP-words. This asymmetry was necessary to create word-level probability differences while matching word frequencies. Thus, rather than reflecting a familiarity preference for the familiar HTP-words, it is possible that the results from Experiment 3 reflect a 'novelty' preference for the infrequent syllables in HTP-words. However, there is no reason to expect a change in direction of preference across these three experiments, given that they used the same methods and materials. Moreover, Experiment 2 demonstrates that the familiarity preference from Experiment 1 also emerges when syllable frequency is controlled, and is influenced by word-level familiarity. It thus seems most likely that the results of Experiment 3 reflect a familiarity preference based on discrimination of sequence-level probability cues.

This discussion reflects a broader issue that pervades laboratory experiments focused on language learning. In any artificial language, and by extension any natural language that is manipulated to conform to specific statistical patterns, controlling for one aspect of linguistic structure necessarily introduces correlated statistical cues that may support subsequent learning (see Seidenberg, MacDonald, & Saffran, 2002). For example, the original infant statistical learning studies (Saffran et al., 1996) were criticized because the tested words and part-words occurred with different absolute frequencies. To address this concern, follow-up work in this area used test words and part-words that were frequency matched (Aslin et al., 1998). As a result, individual syllables were no longer equated for frequency; syllables in the part-words were twice as frequent as those in the words. It is mathematically impossible to frequency-match both the test words and their component syllables; any manipulation of sentence-level TPs will necessarily alter syllable and word frequencies. For this reason, the current studies followed the strategy of using nearly identical corpora while manipulating one variable at a time. A hybrid approach, one in which both test words and syllables are closer in frequency but neither is exactly matched, would be difficult to interpret; it would be unclear whether successful discrimination is due to differences in syllable frequency, word frequency, TP, or some combination thereof.

Given that the current study establishes a role for TP computation over natural language corpora, it will be possible to test subtler TP contrasts in future studies. That is, instead of 1.0 versus .33, we can compare infants' discrimination of smaller differences between TPs. As the TPs between target words move closer together, it will be possible to decrease the syllable-level frequency differences between items. The current work thus lays the necessary groundwork for a range of future studies designed to probe the limits of infant statistical learning in natural speech.

General Discussion

Artificial languages have been invaluable for identifying potential mechanisms involved in language acquisition. In studies of word segmentation, these materials typically contain a single cue to word boundaries, namely, differences in within- and between-word TPs, and are devoid of any additional acoustic information (unless other segmentation cues are also being artificially manipulated; Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003, 2007). Infants appear to rapidly track distributional regularities in simplified artificial materials.

In this study, we challenged statistical learning accounts by using materials that were markedly more similar to those that infants might actually encounter in their native language environments: an unfamiliar natural language (Italian). Experiment 1 demonstrated that following familiarization with Italian speech, infants were able to discriminate words from the familiarization stream from novel Italian words. Experiment 2 demonstrated that infants were sensitive to the familiarity of syllable *sequences*, rather than the familiarity of individual syllables. In Experiment 3, we manipulated TPs to determine whether infants tracked statistical regularities when listening to the Italian passages. The HTP- and LTP-words occurred equally often during familiarization, and shared the trochaic stress pattern typical of both English and Italian. Infants nevertheless successfully discriminated the HTP-words from the LTP-words.

Natural languages represent a noisy stimulus, in which the words of interest are interspersed amidst myriad other words, word repetitions are necessarily limited, and TPs are just one of many regularities in the input (e.g., prosodic patterns, morphological agreement, word order, etc.). These results thus provide a striking demonstration of statistical learning, in which infants detected sequential probabilities despite the richness of the experimental materials. In fact, despite the potential drawbacks of natural materials, their complexities may prove to

be advantageous for infant learners. Natural languages have the benefit of providing infants with multiple redundant cues to word boundaries, and are inherently more engaging than artificial languages. Indeed, prior research suggests that making artificial languages just a little more natural, by using infant-directed speech intonation contours, can facilitate statistical learning (Thiessen et al., 2005). Similarly, infants learn more about word order when a word sequence is sung rather than spoken, providing engaging redundant cues to learning (Thiessen & Saffran, in press).

Although our goal was to assess statistical learning in a more natural environment, all laboratory experiments necessarily simplify the learning experience. In particular, the range of TPs used in this experiment was obviously extreme relative to real Italian. Nevertheless, the fact that TP-based discriminations emerged after just a few minutes of exposure to these materials, which are complex relative to prior artificial language studies, is encouraging. Future studies will further explore infants' natural-language statistical learning abilities in the lab. For example, the current data do not allow us to determine whether infants treated the LTP-words as possible, yet weaker, word candidates than the HTP-words, or whether they did not treat them as word candidates at all. Further experiments contrasting LTP-words with novel words will be necessary to address this issue. It will also be of great interest to assess infants' abilities to track statistics amidst linguistic stimuli containing stress and phonotactic patterns that do not conform to the infants' native language.

While the current results take an important step in the direction of ecological validity, the degree to which infants actually *use* statistical cues for word segmentation remains unknown. One issue is that the potential usefulness of TPs or other related statistics (e.g., mutual information), as assessed via corpus analyses, remains in dispute (Frank, Gold-water, Mansinghka, Griffiths, & Tenenbaum, 2007; Swingley, 2005; Yang, 2004). Importantly, the initial impetus for this line of work was not to claim that TPs could explain all of infant word segmentation, but rather that TPs might serve a bootstrapping function. Learners must somehow break into the system, finding some initial candidate words. Based on that nascent corpus, learners can then begin to discover native-language regularities that require some knowledge of word forms (e.g., correlations between syllable stress and position within a word, as in lexical stress cues). TP—an imperfect cue—is thus hypothesized to work in concert with other available but imperfect segmentation cues in natural languages (Christiansen, Allen, & Seidenberg, 1998; Swingley, 2005; Thiessen & Saffran, 2003, 2007).

A somewhat different issue confronting studies of infant word segmentation, including those using natural languages, is that it is unclear what successful test discrimination actually tells us about *segmentation* per se. While these familiarization–discrimination experiments are typically described as 'segmentation tasks,' neither the current study, nor the many other studies in this literature, explicitly test the segmentation of individual words from fluent speech. The results from such studies do provide evidence concerning the types of cues that influence discrimination between target words and nontarget foils. However, whether or not infants have actually segmented words during familiarization is not directly tested. One method that comes closer to testing segmentation involves a hybrid task combining word segmentation and novel word learning (Graf Estes, Evans, Alibali, & Saffran, 2007). The idea behind this method is that segmentation should render candidate words available for linking to meaning. Results from these studies suggest that exposure to fluent speech containing TP cues to word boundaries facilitates the acquisition of label–object pairings consistent with the TP cues. Future studies will employ the natural language materials used here with subtler measures designed to more directly tap segmentation itself.

In sum, it appears that sequential statistical learning is not just an artifact of the artificial materials used in previous studies. Infants were able to make use of these cues given rich natural language input, differentially tracking sequences based on their internal probabilistic structure. While it is clear that many other types of information are necessarily integrated during the word segmentation process, the current results represent an important step in showing that statistical learning is robust to the complexities of naturally produced language. Future research will continue to explore the hypothesis that statistical learning plays a fundamental (but not exclusive) role in infants' everyday language acquisition.

Acknowledgments

References

Aslin RN, Saffran JR, Newport EL. Computation of conditional probability statistics by 8-month-old infants. Psychological Science. 1998; 9:321–324.

Bortfeld H, Morgan JL, Golinkoff RM, Rathbun K. Mommy and me: Familiar names help launch babies into speech-stream segmentation. Psychological Science. 2005; 16:298–304. [PubMed: 15828977]

Brent MR, Siskind JM. The role of exposure to isolated words in early vocabulary development. Cognition. 2001; 81:B33–B44. [PubMed: 11376642]

Christiansen MH, Allen J, Seidenberg MS. Learning to segment speech using multiple cues: A connectionist model. Language and Cognitive Processes. 1998; 13:221–268.

Christophe A, Dupoux E, Bertoncini J, Mehler J. Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. Journal of the Acoustical Society of America. 1994; 95:1570–1580. [PubMed: 8176060]

Cole, R.; Jakimik, J. A model of speech perception. In: Cole, R., editor. Perception and production of fluent speech. Hillsdale, NJ: Erlbaum; 1980. p. 133-163.

Curtin S, Mintz TH, Christiansen MH. Stress changes the representational landscape: Evidence from word segmentation. Cognition. 2005; 96:233–262. [PubMed: 15996560]

Cutler A, Carter DM. The predominance of strong initial syllables in the English vocabulary. Computer Speech and Language. 1987; 2:133–142.

De Mauro, T.; Mancini, F.; Vedovelli, M.; Voghera, M. Lessico di frequenza dell'italiano parlato. Milano: Etaslibri; 1993.

Fiser J, Aslin RN. Statistical learning of new visual feature combinations by infants. Proceedings of the National Academy of Sciences of the United States of America. 2002; 99:15822–15826. [PubMed: 12429858]

Frank, MC.; Goldwater, S.; Mansinghka, V.; Griffiths, T.; Tenenbaum, J. Proceedings of the 29th Annual Meeting of the Cognitive Science Society. Austin, TX: Cognitive Science Society; 2007. Modeling human performance in statistical word segmentation; p. 281-286.

Friederici AD, Wessels JM. Phonotactic knowledge and its use in infant speech perception. Perception & Psychophysics. 1993; 54:287–295. [PubMed: 8414887]

Graf Estes K, Evans JL, Alibali MW, Saffran JR. Can infants map meaning to newly segmented words? Statistical segmentation and word learning. Psychological Science. 2007; 18:254–260. [PubMed: 17444923]

Houston DM, Jusczyk PW, Kuijpers C, Coolen R, Cutler A. Cross-language word segmentation by 9-month-olds. Psychonomic Bulletin & Review. 2000; 7:504–509. [PubMed: 11082857]

Johnson EK, Jusczyk PW. Word segmentation by 8-month-olds: When speech cues count more than statistics. Journal of Memory and Language. 2001; 44:548–567.

Jusczyk PW, Aslin RN. Infants' detection of the sound patterns of words in fluent speech. Cognitive Psychology. 1995; 29:1–23. [PubMed: 7641524]

Jusczyk PW, Cutler A, Redanz NJ. Infants' preference for the predominant stress patterns of English words. Child Development. 1993; 64:675–687. [PubMed: 8339688]

Jusczyk PW, Hohne EA, Bauman A. Infant's sensitivity to allophonic cues for word segmentation. Perception & Psychophysics. 1999; 61:1465–1476. [PubMed: 10598463]

Jusczyk PW, Houston DM, Newsome M. The beginnings of word segmentation in English-learning infants. Cognitive Psychology. 1999; 39:159–207. [PubMed: 10631011]

Killeen PR. An alternative to null-hypothesis significance tests. Psychological Science. 2005; 16:345–353. [PubMed: 15869691]

Kirkham NZ, Slemmer JA, Johnson SP. Visual statistical learning in infancy: Evidence for a domain general learning mechanism. Cognition. 2002; 83:B35–B42. [PubMed: 11869728]

Mancini F, Voghera M. Lunghezza, tipi di sillabe e accento in taliano. Archivio Glottologico Italiano. 1994; 79(1):51–77.

Mattys SL, Jusczyk PW. Phonotactic cues for segmentation of fluent speech by infants. Cognition. 2001; 78:91–121. [PubMed: 11074247]

Morgan JL. A rhythmic bias in preverbal speech segmentation. Journal of Memory and Language. 1996; 35:666–688.

Newport, E.; Weiss, DJ.; Wonnacott, L.; Aslin, RN. Statistical learning in speech: Syllables or segments?. Boston, MA: Presented at the annual Boston University Conference on Language Development; 2004.

Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. Science. 1996; 274:1926–1928. [PubMed: 8943209]

Saffran JR, Johnson EK, Aslin RN, Newport EL. Statistical learning of tone sequences by human infants and adults. Cognition. 1999; 70:27–52. [PubMed: 10193055]

Seidenberg MS, MacDonald MC, Saffran JR. Does grammer start where statistic stop? Science. 2002; 298:553–554. [PubMed: 12386323]

Saffran JR, Wilson DP. From syllables to syntax: Multilevel statistical learning by 12-month-old infants. Infancy. 2003; 4:273–284.

Swingley D. Statistical clustering and the contents of the infant vocabulary. Cognitive Psychology. 2005; 50:86–132. [PubMed: 15556130]

Thiessen ED, Hill EA, Saffran JR. Infant-directed speech facilitates word segmentation. Infancy. 2005; 7:53–71.

Thiessen ED, Saffran JR. When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. Developmental Psychology. 2003; 39:706–716. [PubMed: 12859124]

Thiessen ED, Saffran JR. Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. Language Learning and Development. 2007; 3:73–100.

Thiessen ED, Saffran JR. How the melody facilitates the message, and vice versa, in infant learning and memory. Proceedings of the New York Academy of Sciences. in press

Tyler, M.; Johnson, E. Testing the limits of statistical language learning; Kyoto, Japan. Paper presented at the annual meeting of the Biennial International Conference on Infant Studies; 2006 Jun.

Weber C, Hahne A, Friedrich M, Friederici AD. Discrimination of word stress in early infant perception: Electrophysiological evidence. Cognitive Brain Research. 2004; 18:149–161. [PubMed: 14736574]

Yang C. Universal grammar, statistics, or both. Trends in Cognitive Sciences. 2004; 8:451–456. [PubMed: 15450509]

Appendix

Language 1A/3A

Torno a casa con le bici cariche di frutta in bilico sulla sella.

La zia Carola si è esibita in una fuga colla bici verde.

Se porti il melo sulla bici forse cali un po' di chili.

La bici ha subito un danno dentro la casa del capo di Lara.

La cavia Bida è in fuga da casa per aver giocato con le bilie blu.

La biscia in lenta fuga dal giardino capita in casa mia.

Il tuo melo arcano fuga l'afa che debilita la folla.

Arriviamo in bici fino al bivio del grande melo con un caro amico.

Il picchio si abitua a fare la sua casa in ogni melo cavo e alto.

Gusto i bigoli dentro casa o coricata all' ombra del melo verde.

Di rado una bici in rapida fuga rincorre la moto bigia e rossa.

Per ascoltare la fuga quasi cadi sul melo e inciampi sulla biro sull'erba.

Language 1B

Torno a casa con le bici cariche di frutta in bilico sulla sella.

La zia Carola si è esibita in una tema colla bici verde.

Se porti il pane sulla bici forse cali un po' di chili.

La bici ha subito un danno dentro la casa del capo di Lara.

La cavia Bida è in tema da casa per aver giocato con le bilie blu.

La biscia in lenta tema dal giardino capita in casa mia.

Il tuo pane arcano tema l'afa che debilita la folla.

Arriviamo in bici fino al bivio del grande pane con un caro amico.

Il picchio si abitua a fare la sua casa in ogni pane cavo e alto.

Gusto i bigoli dentro casa o coricata all'ombra del pane verde.

Di rado una bici in rapida tema rincorre la moto bigia e rossa.

Per ascoltare la tema quasi cadi sul pane e inciampi sulla biro sull'erba.

Language 2A

Il giovane figlio di Marisa ha tagliato il melo per fare pali da lavoro.

Quella pazza di Tèrri si è esibita in una fuga avventurosa dal negozio.

Di solito cerco l'ombra del melo verde presso la casa di Paco Rossi.

La verde biscia in fuga viene da te per trovare riparo fra le macerie.

Il cane di Matilde gusta le tagliatelle sotto al melo ombroso.

La zia passa le sue ferie in montagna dove fuga l'afa.

Le sirene pedalano in bici fino al bivio del grande melo con un caro amico.

Ho visto una bici in rapida fuga sulla strada nevosa.

Lilla la maghetta si nasconde sempre dietro al melo antico.

Per ascoltare la fuga quasi ho macchiato il tappeto di the verde.

Se porti la terra per il m i cali un po' di chili.

L'altro Lunedì Maddalena era in fuga da casa tra la paglia gialla.

Language 2B

Ogni un mese compro il pane e i rigatoni dal fornaio della Futa.

La zia Méda ha scritto un futile tema sul gazebo artistico di Locarno.

Gabriella ha messo il pane sulla bici per calare un po' di chili.

Ieri ho portato in officina la Tema colorata che fu della nonna Carolina.

Tua sorella Carla ha preso il pane dalla gavetta usando il mestolo nuovo.

Il gattino Refuso è il simpatico protagonista del tema che ho svolto.

Purtroppo per comprare il pane sono andati in fumo tutti i soldi.

Luigi vuole assolutamente rispettare l'orario stabilito per la consegna del tema di storia.

Dentro la busta del pane è anche un regalino che ti ho portato per Natale.

Mescolo il tema musicale ad altra musica scritta dal fu Lorenzo Bianchi.

I tuoi bimbi corrono velocissimi a comprare il pane fresco.

Questo mese ho scritto un Tema sulla cometa dell'anno scorso.

Language 3B

Non è da me scendere dal melo in una futile fuga dalle api.

Torno a casa dalla futa con la bici piena di mele mature.

Il melo e diverse bici furono portate presso la mescita di vino.

Zio Luigi Medo è in fuga colla bici verde.

Vi fu l'età dei tentativi di fuga in bici verso il rifugio del melo antico.

Il fu Romero Rossi temeva di andare in gita colla bici nuova.

Dario fu l'ingenuo che portò una bici a casa il mese scorso.

Una fuga da casa è il sogno della topina Mela verso la libertà.

Il ratto Meco tentò la fuga da casa quando vi fu la tempesta.

Il micio Refuso medita in casa o dimena la coda sotto al melo ombroso.

Sui rami del melo che sembrano fusi c'è la casa del fuco solitario.

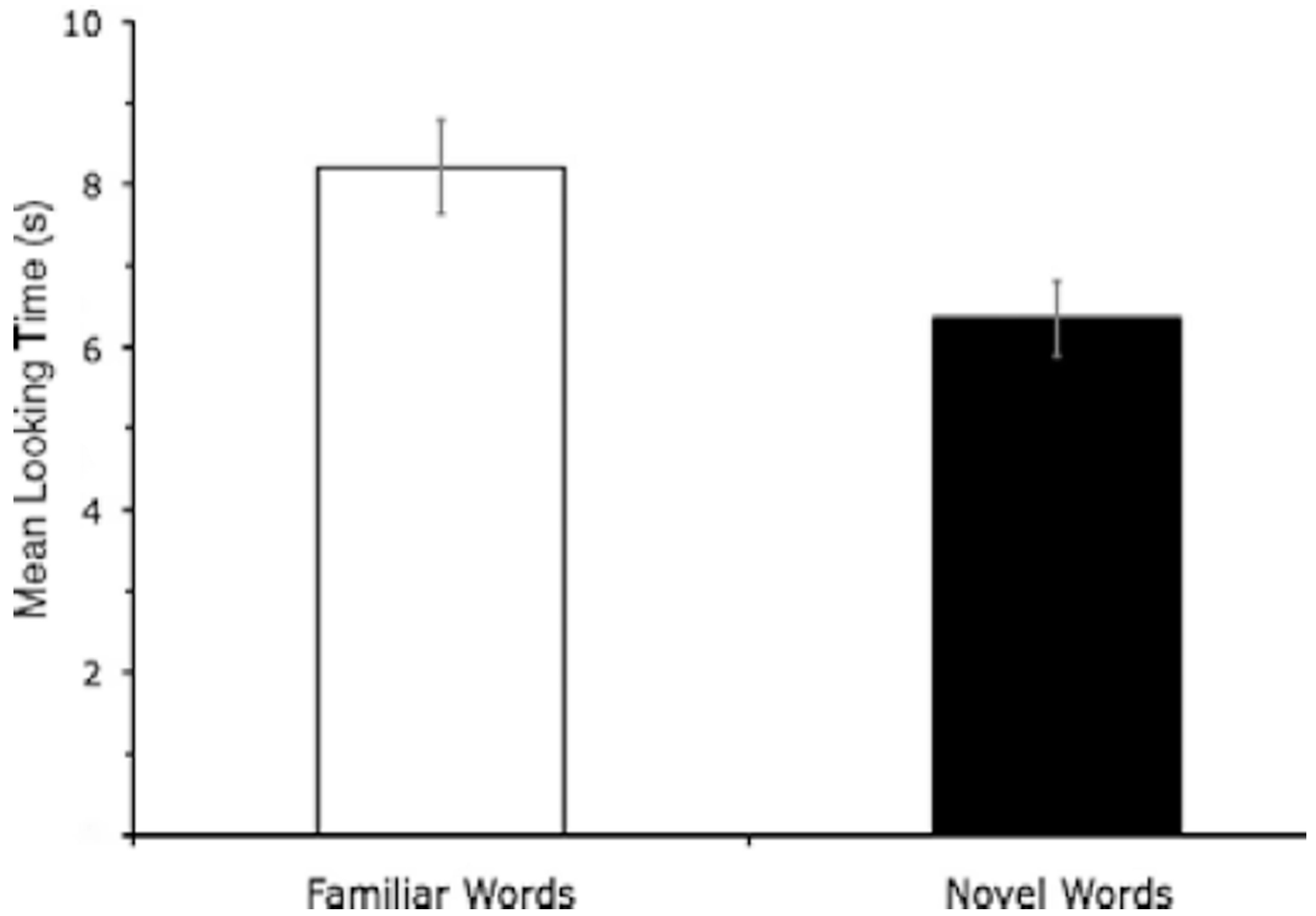La fuga della stella cometa si è fermata sul melo che fu della zia.

Figure 1.
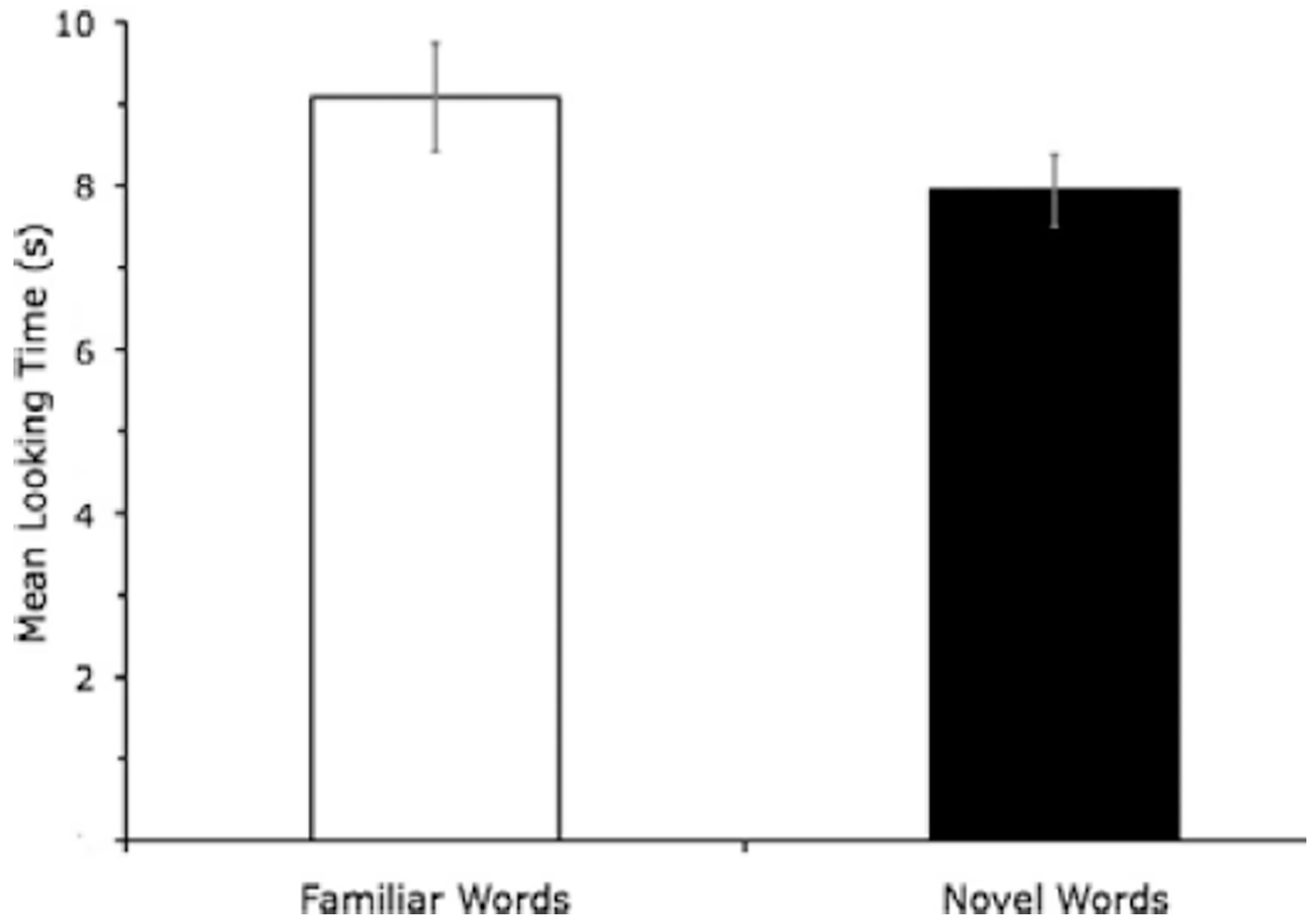Results of Experiment 1: Mean looking times (±1 *SE*) to familiar words and novel words.

Figure 2.
Results of Experiment 2: Mean looking times (±1 *SE*) to familiar words and novel words.
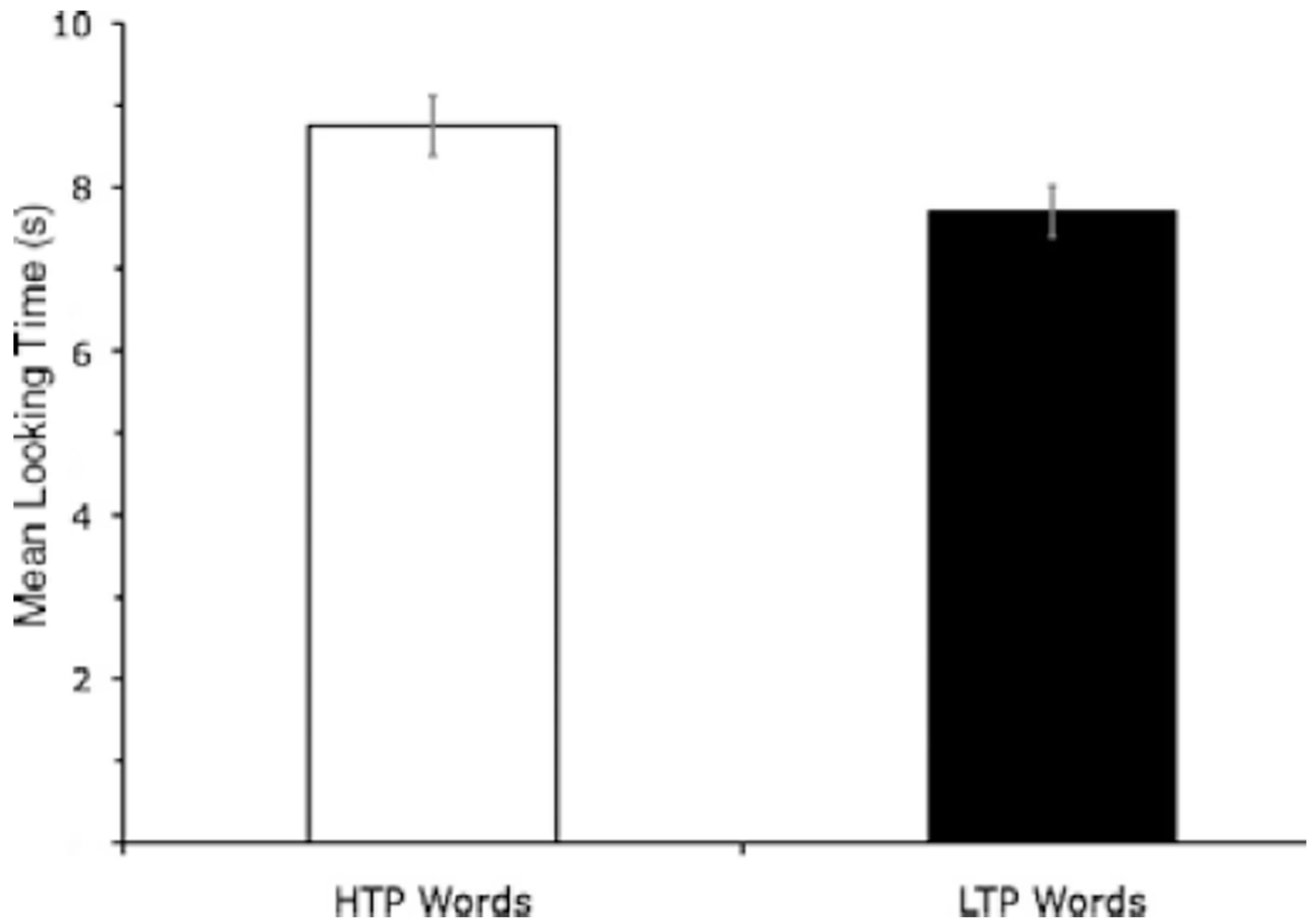
Figure 3.
Results of Experiment 3: Mean looking times (±1 *SE*) to HTP- and LTP-words.

Table 1

Familiar Words, Novel Words, HTP-Words, and LTP-Words Used in the Testing Phase of Experiments 1, 2, and 3

Exp.	Language	Familiar words	Novel words	HTP-words	LTP-words
1	Language 1A	*fuga, melo*	*pane, tema*	—	—
	Language 1B	*pane, tema*	*fuga, melo*	—	—
2	Language 2A	*fuga, melo*	*pane, tema*	—	—
	Language 2B	*pane, tema*	*fuga, melo*	—	—
3	Language 3A	—	—	*fuga, melo*	*bici, casa*
	Language 3B	—	—	*bici, casa*	*fuga, melo*

Note. HTP = high transitional probability; LTP = low transitional probability.

Table 2

Average Duration, Intensity and Pitch (±SE) of Target Words in the Familiarization Corpus of Experiment 3

Language	Words	
	fuga–melo	*bici–casa*
3A		
Duration (ms)	**334 (29.8)**	338 (22.6)
Intensity (dB$_{SPL}$)	**62.4 (0.79)**	61.1 (0.38)
Pitch (Hz)	**266 (11.65)**	276 (13.1)
3B		
Duration (ms)	320 (15.3)	**322 (15.9)**
Intensity (dB$_{SPL}$)	62.0 (0.67)	**60.2 (1.18)**
Pitch (Hz)	278 (13.7)	**248 (14.2)**

Note. Items for which infants displayed a listening preference during the testing phase are indicated in bold.