Cognition 133 (2014) 420–428

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

From pauses to clauses: Prosody facilitates learning of syntactic constituency

Kara Hawthorne [a,b,*], LouAnn Gerken [c]

[a] University of Arizona, Department of Linguistics, 1100 E University Blvd, Tucson, AZ 85721, United States
[b] University of Alberta, Department of Linguistics, 2-40 Assiniboia Hall, Edmonton, AB T6G 2E7, Canada
[c] University of Arizona, Department of Psychology, 1503 E University Blvd, Tucson, AZ 85721, United States

ARTICLE INFO

Article history:
Received 30 July 2012
Revised 28 July 2014
Accepted 31 July 2014
Available online 23 August 2014

Keywords:
Prosody
Prosodic bootstrapping
Constituency
Syntax acquisition
Language acquisition

ABSTRACT

Learning to parse the speech stream into syntactic constituents is a crucial prerequisite to adult-like sentence comprehension, and prosody is one source of information that could be used for this task. To test the role of prosody in facilitating constituent learning, 19-month-olds were familiarized with non-word sentences with 1-clause (ABCDEF) or 2-clause (ABC, DEF) prosody and were then tested on sentences that represent a grammatical (DEF, ABC) or ungrammatical (EFA, BCD) 'movement' of the clauses from the 2-clause familiarization sentences. If infants in the 2-clause group are able to use prosody to group words into cohesive chunks, they should discriminate between grammatical and ungrammatical movements in the test items, even though the test sentences have a new prosodic contour. The 1-clause, control, group should not discriminate. Results support these predictions and suggest that infants treat prosodically-grouped words as more cohesive and constituent-like than words that straddle a prosodic boundary. A follow-up experiment suggests that these results do not merely reflect recognition of words in boundary positions or acoustic similarity of words across the familiarization and test phases.

1. Introduction

Superficially, language takes the form of a linear string of words, but that linear string is the result of syntactic phenomena operating over hierarchically-organized constituents. Syntactic constituents – minimally defined as groups of words that function as cohesive units in sentences – form the building blocks upon which natural language grammar is organized. Constituency is an important concept even at the early stages of syntax acquisition. A rudimentary appreciation of constituency is necessary to understand the subject-predicate distinction, which underlies comprehension of who did what to whom. For more advanced learners, constituency is important for interpreting phenomena such as proform replacement (e.g., *it* can replace noun phrases) and syntactic movement (e.g., the clauses "when Grandma gets here" and "we'll go to the zoo" can appear in either order). Correctly parsing a sentence into constituents is also crucial for interpreting syntactically-ambiguous strings, such as "old men and women" ([old [men and women]] versus [[old men] and women]).

Given the critical role that constituents play in syntax, it follows that a learner who can parse constituents from the speech stream will be advantaged in her language development. There are many potential sources of information about constituency, including frequently-occurring function morphemes (Gervain, Nespor, Mazuka, Horie, &

* Corresponding author. Address: Department of Linguistics, University of Alberta, 2-40 Assiniboia Hall, Edmonton, AB T6G 2E7, Canada. Tel.: +1 (780) 492 5174; fax: +1 (780) 492 3434.

E-mail addresses: khawthor@ualberta.ca (K. Hawthorne), gerken@email.arizona.edu (L. Gerken).

http://dx.doi.org/10.1016/j.cognition.2014.07.013
0010-0277/

Mehler, 2008; Morgan, Meier, & Newport, 1987), cross-sentential comparisons (Morgan, Meier, & Newport, 1989), transitional probabilities between syllables (Takahashi & Lidz, 2007; Thompson & Newport, 2007), and semantics (Pinker, 1984). However, with the exception of transitional probabilities, these cues rely on prior learning and may not be helpful at the earliest stages of acquisition.

Prosody, the rhythmic and melodic aspects of speech, is one aspect of the input that may serve as a stepping-stone for early constituent learning. The prosodic bootstrapping hypothesis (Gleitmen & Wanner, 1982; Morgan, 1986; Peters, 1983) proposes that infants use prosody to identify word, phrase, and clause boundaries and possibly even to infer constituency and hierarchical syntactic structure. This is a promising hypothesis, since infants have demonstrated sensitivity to a broad variety of prosodic information from the very earliest ages in both perception (e.g., Nazzi, Bertoncini, & Mehler, 1998) and production (Halle, de Boysson-Bardies, & Vihman, 1991; Mampe, Friederici, Christophe, & Wermke, 2009). Prosody is also a powerful cue for word segmentation (Johnson & Seidl, 2009; Jusczyk, Cutler, & Redanz, 1993), and larger prosodic groupings constrain the domain over which infants look for words (Gout, Christophe, & Morgan, 2004; Shukla, White, & Aslin, 2011), suggesting that infants are adept at extracting prosodic regularities from the input.

There is a large body of work examining prosodic cues at the boundaries of major syntactic constituents. Clauses typically correspond to the Intonational Phrase level of the prosodic hierarchy (Nespor & Vogel, 1986; Selkirk, 1984) and, as such, are marked with a pause, pitch resets, and final syllable-lengthening (Beckman & Edwards, 1990; Beckman & Pierrehumbert, 1986; Cooper & Paccia-Cooper, 1980; Cruttenden, 1986). These prosodic features are exaggerated and occur more reliably in infant-directed speech (IDS) (Broen, 1972; Fernald et al., 1989; Garnica, 1977; Morgan, 1986). Infants are able to perceive these cues, preferring to listen to speech that does not violate typical correlations of clause-final prosodic cues (Hirsh-Pasek et al., 1987; Jusczyk, 1989).

To examine the role of prosody on infants' memory for strings of words, Nazzi, Kemler Nelson, Jusczyk, and Jusczyk (2000; c.f. Seidl, 2007; Soderstrom, Seidl, Kemler Nelson, & Juscayk, 2003) familiarized 6-month-olds with strings of words that formed a prosodic constituent (e.g., "Rabbits eat leafy vegetables") or straddled a prosodic boundary (e.g., "...rabbits eat. Leafy vegetables...") and found that words that form a prosodic constituent were better remembered and recognized when embedded in a passage of fluent speech at test. Soderstrom, Kemler Nelson, and Jusczyk (2005) extended these results by placing the familiarization (as well as test items) in a short passage of fluent speech. They found that infants used prosody to pull substrings out of fluent speech and then recognize them during testing. However, there was also an effect of acoustic similarity: infants preferred to listen to a test passage in which the target string had the same prosodic features – either comprising a prosodic constituent or straddling a prosodic boundary – as it did during familiarization.

Together, these studies offer evidence that prosody influences how infants remember linguistic stimuli and even helps with extracting groups of words from continuous speech. However, while parsing speech into substrings is an important prerequisite to syntax acquisition, constituents are crucially not just substrings pulled from the speech stream. Constituents are hierarchically-organized groupings that are cohesive, rule-abiding units of grammar and are independent of a particular acoustic manifestation. Recent work by Langus, Marchetto, Bion, and Nespor (2012) demonstrates that adults can use prosody to segment speech into phrases embedded into sentences, but their work does not address the developmental question of whether infants treat prosodically-grouped words like syntactic constituents – for example, by recognizing the grouping when the non-segmental acoustic features have changed.

To test whether infants can use prosody to chunk sentences into constituent-like units, we examine infants' ability to not only parse the speech stream into substrings, but to recognize those substrings when they behave like cohesive constituent-like chunks and 'move' to a new position in the utterance. Nineteen-month-olds were familiarized with sentences with 1-clause (ABCDEF, where each letter represents a class of two nonsense words) or 2-clause (ABC, DEF) prosody and were then tested on sentences that represent a grammatical (DEF, ABC) or ungrammatical (EFA, BCD) 'movement' of the clauses that the 2-clause group heard during familiarization. If infants in the 2-clause group use prosody to locate groups of words that they treat like cohesive units, they should discriminate between the grammatical and ungrammatical movements in the test items, because the grammatical-movement items maintain the prosodically-cohesive groupings from familiarization, while the ungrammatical-movement items don't. The 1-clause group serves as a control, since their familiarization stimuli contain no internal prosodic boundaries.

Such a finding would not necessarily indicate that the infants were computing complex syntactic structures involving moved constituents. However, there are two reasons why movement is a useful test for investigating the underpinnings of constituency learning in an artificial grammar. First, since only words that are in the same constituent can 'move' together in natural languages,[1] a learner who recognizes a substring of a larger sentence when it appears in a transformed version of that sentence is going beyond grouping the speech stream into substrings. She is treating prosodically-marked strings of words as more cohesive than strings containing a prosodic break. Cohesiveness is an identifying property of within-constituent words, as seen through the movement of constituents, pro-form

[1] There are circumstances in which a non-constituent string of words *appears* to move as a single unit. Consider, for example, the sentence "Heidi saw Greg yesterday at the park." The constituents "yesterday" and "at the park" can both 'move' to the beginning of the sentence, leaving the transformed sentence "Yesterday, at the park, Heidi saw Greg." The string 'yesterday at the park' appears to have moved as a single unit, but it does not form a syntactic constituent. Often, it does not form a prosodic constituent either. For the purposes of this paper, we presume that infants are biased to infer that movement is, minimally, evidence for the cohesiveness of certain strings of words, which is a crucial feature of syntactic constituents. Thanks to the reviewer who pointed this out.

replacement of constituents, etc. (Note that we are using the term ‘movement’ here as a way of describing the relationship between constituents in related forms of a given sentence and are not assuming any particular syntactic theory.) Therefore, if a learner recognizes prosodically-grouped words when they appear in a new position in a sentence, we can infer that she is treating the grouping like a cohesive, reorderable chunk, much like a syntactic constituent.

A second and equally important aspect of the movement test lies in the acoustics: a constituent has a different prosodic contour, depending on where it occurs in the sentence. Clause ABC, for example, ends with a fall-rise intonation contour (i.e., continuation rise) in the 2-clause familiarization items, while it has high-falling contour when it has moved to the end of the sentence for the grammatical-movement test items. Therefore, if participants in the 2-clause group abstract away from the input acoustics to recognize prosodically-grouped clauses at test, it will suggest that they have acquired a representation of the experimental grammar that goes beyond what can be directly observed in the familiarization items.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Twenty-four monolingual English-acquiring 19-month-olds were tested (16 male, 9 female). Participants were all of normal birth weight ($\geqslant$5 lb, 7 oz) and term ($\geqslant$37 weeks). Data from an additional 18 infants were not included due to fussiness ($n = 8$), because the participant's mean looking time was more than 2.5 standard deviations above the mean ($n = 1$), and because of equipment or experimenter error ($n = 9$, one experimenter was in training).

2.1.2. Materials

The stimuli consisted of 6-word sentences, with either 1-clause (ABCDEF) or 2-clause (ABC, DEF) prosody. The 1-clause sentences had prosody corresponding to “I ate at home with John,” and the 2-clause sentences had prosody corresponding to “While I ate, John came home.” There were two nonsense CVC words from each class (A = *bup*, *nim*; B = *div*, *pel*; C = *kag*, *rud*; D = *feb*, *toz*, E = *has*, *zaf*; F = *jik*, *vot*).

The sentences were read by a native English speaking female who was naïve to the purpose of the experiment. The speaker produced the stimuli in a natural and happy-sounding way, using IDS. Recordings were made using Praat (Boersma & Weenink, 2008) and an Andrea Anti-Noise USB NC-7100 microphone during a single session in a sound attenuated booth.

The recorded sentences were partially standardized for tempo and pause length, so that they retained a natural amount of variability without any of the sentences sounding distractingly different. Any sentence whose tempo or pause length was more than a standard deviation from the mean was adjusted to the mean value for that variable. The ungrammatical movement test items averaged 80 ms (milliseconds) longer than the grammatical movement test items, a statistically significant difference ($t(31) = 7.20$, $p < .01$). Since the source of this discrepancy is undetermined (it is not accounted for by difficulties measuring the onsets of stop closures or the offsets of unreleased final stops, nor could we find any patterns resulting from combinations of particular consonants in particular prosodic positions, although the speaker did tend to produce shorter pauses (by 40 ms) at the clause boundary if the second clause started with *nim*), the test items were adjusted to have the same mean value, in order to avoid any potentially confounding effects of this durational difference.

After standardization, the stimuli were played to several English-speaking linguists, who verified that the stimuli sounded natural and that it was impossible to tell which sentences had been altered. The sentences were then arranged into lists, with sentences separated by a 1 s (second) pause.

Acoustic analyses confirmed that the familiarization items had the expected prosodic features. The clause boundary in the 2-clause items was marked by an average of a 173 ms pause and with final-syllable lengthening of the pre-boundary word (word C is significantly longer in the 2-clause condition ($t(4) = 9.70$, $p < .001$)). Clause-final words (C and F in the 2-clause items and F in the 1-cause items) were significantly longer than non-clause final words across both familiarization conditions ($F(1, 57) = 38.06$, $p < .001$). Both 1-clause and 2-clause items had intonational contours consistent with the intended prosody. In terms of the prosodic hierarchy proposed by Nespor and Vogel (1986), the 1-clause items had 1-IP (Intonational Phrase) prosody, while the 2-clause items had 2-IP prosody (see Fig. 1).

2.1.3. Design

During familiarization, half of the infants listened to sentences with 1-clause prosody and half to sentences with 2-clause prosody. Both groups heard the familiarization sentences in the same order, divided across two familiarization phases. The initial familiarization phase was conducted in a small room, where the infant, a parent, and an experimenter played quietly while stimuli sentences played in the background for 10 min. During the second familiarization phase, participants listened to an additional 2 min of familiarization sentences in the testing room. Since the 2-clause sentences were longer than the 1-clause sentences (due to their prosodic features), the 2-clause group heard a total of 230 familiarization sentences and the 1-clause group heard a total of 261 sentences. The total duration of familiarization was identical in the two conditions. Since the 1-clause infants are predicted to not discriminate between the grammatical and ungrammatical movement test items, exposing them to additional familiarization sentences can only work against our hypothesis.

While familiarization prosody was a between-participants variable, test item type was a within-participants variable. All participants were tested on items with 2-clause prosody of the forms DEF, ABC and EFA, BCD. The grammatical-movement items (DEF, ABC) had the same clause-like constituents that the 2-clause infants

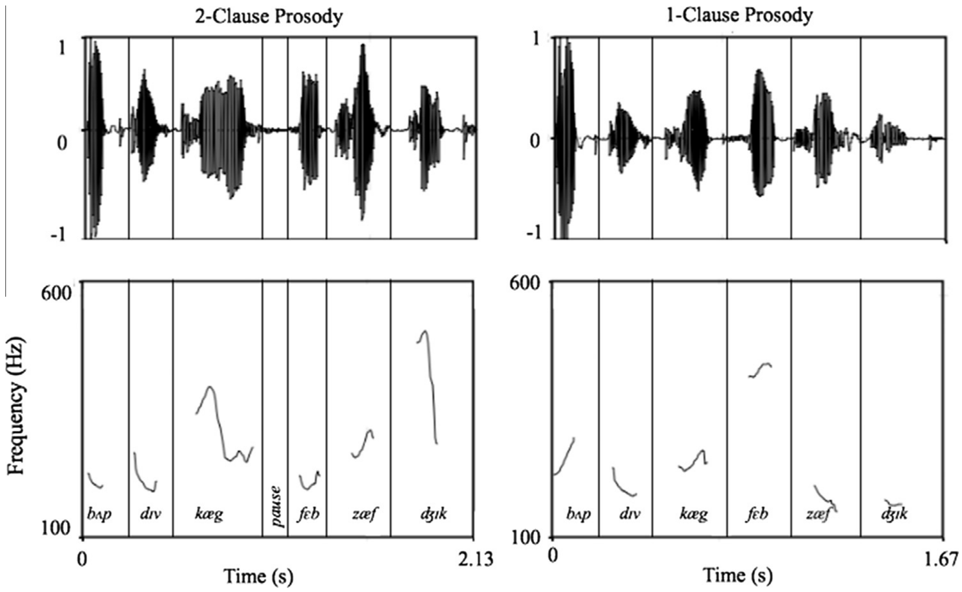

Fig. 1. Example waveforms and pitch contours for a 2-clause and 1-clause familiarization sentence.

heard during familiarization, although the clauses appeared in the opposite order and had new prosodic contours. Conversely, the ungrammatical-movement items (EFA, BCD) chunked the sentence into different, non-constituent units.

2.1.4. Task

Participants were tested using a modified version of the Head-Turn Preference Procedure (Kemler Nelson et al., 1995). The infant was seated in a high chair in a sound-proof booth, with the parent seated behind and to the side. The parent wore headphones, so that she would not bias the infant's responses. There was an amber light in front of the infant, and two red lights directly over speakers, one on each side.

During the 2-min familiarization phase, the entire 2-min familiarization list (1-clause or 2-clause, depending on condition) was played simultaneously from both speakers, while the lights flashed according to the infant's direction of looking. The infant's looking behavior was monitored by an experimenter in an adjacent room, who could not hear the stimuli. During testing, the infants heard 12 test trials, divided into three blocks, each consisting of the same four trials (two grammatical-movement and two ungrammatical-movement test lists). The order was randomized within each block. A test trial began when the infant oriented to the center flashing light. The test list was then played from a single speaker (left or right), while the corresponding light flashed. The trial ended when the infant looked away from the flashing light for 2 s.

2.2. Results

Each participant's looking time was recorded for each trial, and looking times shorter than 2 s were excluded from analysis. The mean looking time per trial was 6.75 s. Nine of the 12 infants in the 2-clause condition and five of the 12 in the 1-clause condition listened longer to grammatical-movement test items.

The data were analyzed using linear mixed effects modeling because it is better able to deal with missing data points and to account for random effects corresponding to both overall looking time (intercept) and experimental manipulations (slope). Familiarization condition (1-clause versus 2-clause) was a between-participants variable, and test condition (grammatical- versus ungrammatical-movement) was a within-participants variable. Since pilot testing revealed effects of test order, test block (1st, 2nd, 3rd) was included, along with test condition, as a random effect grouped by participants. Looking time was the dependent variable, and looking times were log transformed in order to improve the normality of the residuals. Likelihood ratio tests comparing a full model to a minimally reduced model were used to assess the significance of the fixed effects, so results are reported using the chi-squared (χ^2) statistic.

There was a significant main effect of test block ($\chi^2(24) = 37.74$, $p = .037$), reflecting an expected overall decrease in looking times during the course of the experiment, but no significant interactions involving test block. There was a significant interaction between familiarization and test condition ($\chi^2(1) = 4.21$, $p = .040$). The effect of test condition was significant for the 2-clause condition ($\chi^2(1) = 5.94$, $p = .015$), with infants showing a preference for the grammatical-movement test items. The effect of test condition was not significant for the 1-clause condition ($\chi^2(1) = .23$, $p = .63$). Results are depicted in Fig. 2.

2.3. Discussion

Participants in the 2-clause group used prosody to extract constituent-like chunks from the speech stream

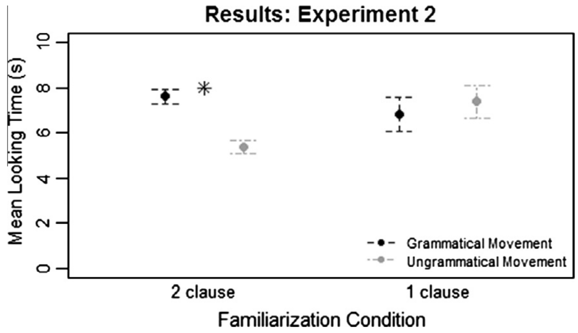

Fig. 2. Error bars represent standard error of the mean difference. Asterisk indicates a significant difference between the adjacent means.

and then recognized those groupings when they had moved to a new part of the sentence and occurred with a new prosodic contour. This provides evidence that infants treat prosodically-grouped words as cohesive, reorder-able units that are independent of a particular acoustic manifestation. Though it is possible that participants were not linking the prosodic groupings with a concept of syntactic constituency, these properties are hallmarks of syntactic constituents, and this behavior is what we would expect from a listener who has the ability to acquire the syntax of a natural language.

There are caveats that influence interpretation of these results. First, the learning shown by the 2-clause group may have been due to the infants preferentially attending to clause-initial, post-pausal words; indeed, children are biased to attend to the beginnings of sentences (Newport, Gleitman, & Gleitman, 1977). The clause-initial words in the 2-clause familiarization items and grammatical-movement test items are the same (words A and D), while the ungrammatical-movement test items have different clause-initial words (E and B). (The same argument could be made regarding clause-final/pre-pausal words, which are also in perceptually-prominent positions). The 1-clause condition was included as a partial control for this possibility, since the 1-clause familiarization items and the grammatical-movement test items share two words (A and F) in prominent positions. If infants are merely attending to pre- or post-pausal words, participants in the 1-clause group should show a similar, though perhaps smaller, discrimination effect. There is no hint that this is the case. While participants in the 2-clause conditioned listened an average of 2.24 s longer to the grammatical-movement test items, those in the 1-clause condition had no trend towards a preference for either type of test item.

Second, the 2-clause familiarization items are acoustically more similar to the grammatical test items on a word-by-word basis.[2] For an infant in the 2-clause group, all of the lengthened words (C and F) remained lengthened in the grammatical test items, while for an infant in the 1-clause group, a new word (C) was lengthened. Performance, therefore, could be due to the similarity of individual words rather than to recognition of clause-like sequences. Again, however, we would expect to see a similar, though probably smaller, discrimination effect from the 1-clause group, since word F was lengthened for both the 1-clause familiarization and the grammatical test items.

Third, it is possible that the 1-clause group was distracted by the change from 1-clause to 2-clause prosody at test. The 2-clause familiarization items, grammatical-movement test items, and ungrammatical-movement test items were all prosodically modeled on "While I ate, John came home," so the 2-clause group did not experience a shift in prosody from familiarization to test, while the 1-clause group did.

In Experiment 2, we build on our results from Experiment 1 by explicitly addressing these three possibilities. In Experiment 2 we also correct an error we made when constructing the familiarization audio files for Experiment 1. In Experiment 1 the particular combinations of ABC and DEF used at test were withheld during the 2-min familiarization period that immediately preceded testing, but they were not withheld during the initial long familiarization period. Each of the 64 unique word combinations, including those used at test, occurred three times in the initial familiarization audio file. In Experiment 2 we completely withheld particular word combinations until the test period. This allows us to address the question of whether infants are generalizing at the level of word class categories.

3. Experiment 2

In Experiment 2, infants were familiarized with the same ABC, DEF 2-clause familiarization items used in Experiment 1. During the test period, infants heard sentences of the forms DEF'ABC (original clauses) and DBF'AEC (scrambled clauses), where the single quote indicates a Phonological Phrase (PhP, see Nespor & Vogel, 1986) boundary. PhPs are a level of the prosodic hierarchy immediately below the IP, and are marked more weakly. IP boundaries typically have final lengthening, pitch resets, and pauses; PhP boundaries are marked with final lengthening and pitch. The 2-IP familiarization items were prosodically modeled on "While I ate, John came home," and the 1-IP test items were modeled on "John came home while I ate." This switch from 2-IP to 1-IP prosody at test more accurately reflects the prosody that would occur naturally if the clauses were swapped. It also allows us to determine whether the 1-clause group in Experiment 1 failed to discriminate solely because they experienced a change in prosody between the familiarization and test phases.

Though they have different prosody, the original-clauses test items are structurally the same as the grammatical-movement test items from Experiment 1. For the scrambled-clauses items (DBF'AEC), the clause-initial and clause-final words appear in the same position as they do in the original-clauses items and the dependency between the clause-initial and clause-final word classes has been maintained. However, the clause-internal words remain in their original locations instead of moving with the rest of the clause. If participants reliably discriminate between

[2] Thank you to the action editor for raising this possibility.

the original-clauses and scrambled-clauses items, it will suggest that they are indeed learning the underlying structure of the artificial grammar. Additionally, it will suggest that infants in Experiment 1 were not preferentially attending to boundary words (since the same words occur in pre- and post-boundary positions in both types of test items) and that they were not discriminating based on acoustic similarity of words across familiarization and test (since the same words are lengthened in both types of test items).

3.1. Methods

3.1.1. Participants

Twelve monolingual English-acquiring 19-month-olds were tested (5 male, 7 female). Participants were all of normal birth weight ($\geqslant$5 lb, 7 oz) and term ($\geqslant$37 weeks). Data from an additional five infants were not included due to fussiness.

3.1.2. Materials, design, and task

Infants were familiarized with the same 6-word sentences of the form ABC, DEF that were used for the 2-clause familiarization in Experiment 1. Unlike in Experiment 1, however, particular combinations of ABC and DEF were withheld during both the initial long and the final 2-min familiarization periods.

Test stimuli were read by the native English speaking female who produced the stimuli for Experiment 1. Items were standardized for tempo and pause length as described in Section 2.1.2. The items were then examined for the presence of prosodic boundary cues between the third and fourth words. The pre-boundary words C and F were significantly lengthened relative to the other words ($F(1, 27) = 22.06$, $p < .001$), even though word C preceded a PhP boundary rather than an IP boundary. Unlike the 2-IP test items from Experiment 1, there were no pauses inside the 1-IP sentences. The pitch contour of the 1-IP items showed a pre-boundary rise in pitch, followed by a notable drop in pitch across the clause boundary. This is consistent the presence of a single IP and two PhPs (see Fig. 3). Note that this is different from the 1-clause familiarization items in Experiment 1, which had a single IP and a single PhP.

The design and task were identical to Experiment 1, except there was only one familiarization condition (2-clause) and all participants were tested on items with 1-IP prosody. The original-clauses items (DEF'ABC) had the same clause-like constituents that the infants heard during familiarization, though the clauses occurred in the opposite order and with a new prosodic contour. The scrambled-clauses items (DBF'AEC) had the boundary words in the same positions as the original-clauses items and the prosody was the same, but the clause-internal words were swapped.

3.2. Results

After looking times shorter than 2 s were excluded, the mean overall looking time was 5.56 s. Ten of the 12 infants listened longer to the original-clauses test items.

Data were analyzed using linear mixed effects modeling, as described in Section 2.2. There was no significant effect of test block ($\chi^2(9) = 6.93$, $p = .64$), so it was excluded from subsequent analyses. There was a significant main effect of test condition: the infants listened longer to the original-clauses test items ($\chi^2(1) = 3.93$, $p = .047$), even though they were pronounced with new, 1-IP prosody (see Fig. 4).

3.3. Discussion

Because the ungrammatical-movement test items in Experiment 1 introduced new words into boundary posi-

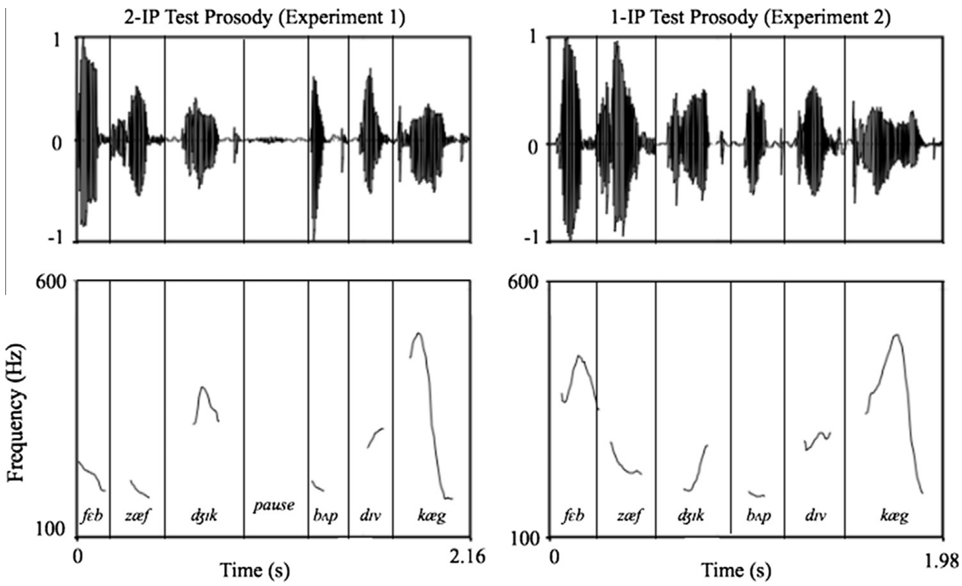

Fig. 3. Example waveforms and pitch contours for a test item from Experiments 1 and 2.

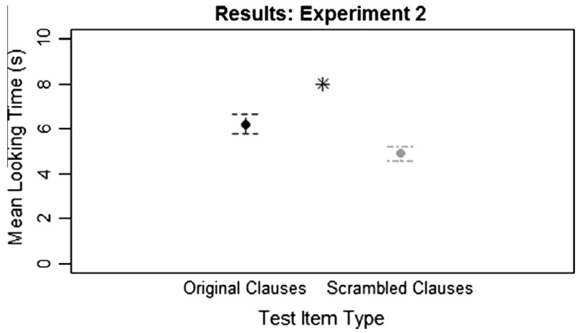

Fig. 4. Error bars represent standard error of the mean difference. Asterisk indicates a significant difference between the adjacent means.

tions – and, therefore, had new lengthened words – while the grammatical-movement test items did not, it is possible that participants in Experiment 1 were influenced by perceptually-prominent words at prosodic boundaries or by how well the test items maintained the relative durations of individual words from familiarization. In Experiment 2, participants could not rely on acoustic salience or similarity of durations, since both types of test items had the same boundary words and the same words in lengthened positions. Yet infants in Experiment 2 successfully discriminated between the original-clauses and scrambled-clauses test items.

The sole difference between the original-clauses and scrambled-clauses items lies in whether the clause-internal words maintain their structural relationship to the other words in the clause. Infants preferred to listen to the original-clauses items, which retained the dependencies between word classes that were learned during familiarization. Importantly, they did so even though both types of test items introduced new word combinations: the clauses in the original-clauses test items used word combinations that participants did not hear during familiarization, and the scrambled-clauses test items had both new word combinations and a new structure.

It is not clear, though, whether the clause-internal structure that the infants learned is of a statistical or grammatical nature. Trigrams (clauses) were withheld to be used at test, but the component bigrams were not. For example, the ABC combination *bup div kag* was withheld, but *bup div* and *div kag* occurred during familiarization, both with a transitional probability (TP) of 0.2. Compare this to the TP of 0 for bigrams in the scrambled-clauses items. Endress and Mehler (2009) found that adults familiarized with trisyllabic nonsense ‘words’ rated new trisyllabic sequences just as highly as the familiarized ‘words’ when the new trisyllabic sequences comprised two overlapping bigrams that had occurred during familiarization. Therefore, it is possible that the infants in the current study preferred the original-clauses test items because they had learned the bigram TPs within the prosodically-signaled clauses. Note, however, that a TP of 0.2 is quite low compared to TPs typically seen in experiments that directly investigate the role of TPs in syntax acquisition (c.f., Gómez & Gerken, 1999; Takahashi & Lidz, 2007; Thompson & Newport, 2007). Moreover, we can be reasonably certain that transitional probabilities alone do not account for our effects: if prosody is factored out, the 1-clause and 2-clause groups from Experiment 1 should have behaved the same.

Alternatively, it is possible that the infants were indeed learning the underlying grammatical structure of the clauses. That is, they may have used prosodic grouping cues to deduce that a word from class A is followed by a word from class B and then a word from class C, that these abstract categories form a cohesive string, and that this string can appear in multiple positions in a sentence with either IP or PhP prosody.

Regardless, since both types of test conditions in Experiment 2 had the same words in boundary positions, the results provide evidence that participants in Experiment 1 were not merely attending to salient words. Similarly, the infants in Experiment 2 were not simply showing a preference for test sentences that maintained the relative durations of words from the familiarization phase, since the same words were in lengthened positions across both test conditions. Finally, the infants in Experiment 2 were able to recognize prosodically-marked constituent-like groupings, even when the prosody of the test items (1-IP) was markedly different from the prosody they heard during familiarization (2-IP).

4. General discussion

Together, these two experiments suggest that 19-month-olds can use prosody to extract groups of words that they treat like cohesive chunks with a consistent linear order of clause-internal elements. They do so even when those chunks appear with different prosodic contours and in different positions in a sentence. As noted above, these are hallmarks of syntactic constituents. In previous studies (e.g., Nazzi et al., 2000; Soderstrom et al., 2005), infants showed enhanced memory for strings of words that they initially heard as well-formed prosodic groupings, particularly when the prosodic contour of the target string matched at familiarization and test. The present experiments extend prior work in two important ways.

First, unlike in prior studies, participants had to generalize beyond input acoustics to recognize prosodically-grouped strings at test. They had to recognize the clauses ABC and DEF when they occurred with new prosody (Experiment 1), and they had to generalize to previously unheard trigrams/clauses and to a new prosodic contour (2-IP versus 1-IP, 2-PhP; Experiment 2). The infants were not just using prosody to chunk the incoming stimuli into unanalyzed units of memory. Rather, they appear to have been chunking the input into strings of syllable bigrams or strings of abstract categories, such that they could recognize new chunks that maintained the same bigram or category sequence at test.

Second, our results suggest that prosody is useful for parsing speech into constituents or the underpinnings thereof. Participants in the 2-clause group of Experiment 1 not only used prosody to extract clause-like substrings from the speech stream, they then recognized them when they 'moved' to a new part of the sentence. On a conservative interpretation, these data provide evidence that infants treat prosodically-grouped words as more cohesive than consecutive words interrupted by a prosodic boundary. Regardless of whether this is a purely perceptual phenomena or a structure-induction process, parsing speech into cohesive chunks is a key task underlying the acquisition of a fuller understanding of syntactic constituents and is the sort of behavior we would expect from a learner with syntactic proclivities.

More broadly, our data are consistent with an interpretation in which prosody serves as a cue to syntactic constituents themselves. Constituents are cohesive strings of words that act together in sentences (e.g., by moving together), and infants appear to treat strings of words with clause-like prosody as cohesive groups of words that can appear in multiple positions of a sentence. Of course, prosody is not likely to be a reliable cue to all constituent boundaries. Sentences, clauses, and the subject-predicate boundary (if the subject is not a pronoun) are reliably marked in infant-directed speech, but there is little evidence that the prosodic cues to lower-level constituent boundaries are sufficiently robust or reliable to be useful to the language learner. Further complicating matters, non-syntactic factors such as style and speech rate also influence prosodic phrasing (c.f. Shattuck-Hufnagel & Turk, 1996).

Nonetheless, our results suggest that infants can use prosody to chunk the speech stream into cohesive chunks on which they perform further analysis (calculating bigrams or learning the relationship between abstract word classes) – abilities which underlie full acquisition of a concept of constituency – in the absence of lexical, contextual, morphological, and statistical sources of information. Regardless of the extent to which the infants' syntactic faculties were recruited in the tasks presented here, the results are what we would expect from individuals with syntax learning capabilities and provide continued support for prosodic bootstrapping proposals.

Further work can take two tacks to investigate how much syntax infants glean from prosody. First, we could look for failure in an experiment similar to the ones reported here[3]: there is ample evidence that young infants are sensitive to boundary words (an ability that is intimately tied with perception of prosodic boundaries) (e.g., Seidl & Johnson, 2006) and that infants can calculate TPs under a broad variety of conditions (e.g., Gómez & Gerken, 1999), so such a failure would suggest that these surface cues alone cannot explain the learning seen here. We did, in fact, fail to find significant results when replicating Experiment 1 – albeit with a shorter familiarization period – with a full complement of younger infants (17-month-olds; $n = 24$). Since both the participants' age and the familiarization duration were changed vis-à-vis Experiment 1, we cannot draw firm conclusions from this failure. However, if the younger infants were given the full familiarization period and the 2-clause group still failed to learn, it would suggest that more sophisticated processing is required to discriminate between grammatical- and ungrammatical-movement items.

Second, we could look for clearer positive evidence of prosody-driven learning in the syntax domain. One more stringent test would be to ask if infants can use prosody to bootstrap a hierarchically-organized grammar (Hawthorne & Gerken, in prep), since hierarchical organization is arguably a more integral aspect of syntax than the movement-based task employed here. However, we can only look for evidence of prosodic bootstrapping in those places where syntactic and prosodic structure overlap. This makes it nigh impossible to distinguish between learning of prosodic versus syntactic constituents, at least without the aid of neuroimaging equipment.

Evidence reported here supports prosody as a cue to syntactically relevant groupings. The extent to which young learners connect the prosodic groupings they so readily perceive with higher-level syntactic concepts remains an open question.

[3] Further thanks to the reviewer for this suggestion.

Acknowledgments

This research was supported by National Science Foundation Grant 0950601 and NICHD #R01 HD042170 to LouAnn Gerken and a National Science Foundation Graduate Research Fellowship to Kara Hawthorne. Thank you to the anonymous reviewers, as well as to Carolyn Quam and Rebecca Gomez for comments on this work. I am also grateful to Sara Knight and other members of the Tweety Lab, as well as the families whose participation made this work possible.

References

Beckman, M. E., & Edwards, J. (1990). Lengthening and shortening and the nature of prosodic constituency. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology: I. Between grammar and the physics of speech* (pp. 152–178). Cambridge, England: Cambridge University Press.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook, 3*, 255–309.

Boersma, P., & Weenink, D. (2008). Praat: Doing phonetics by computer (Version 5.3.04) [Computer program]. <http://www.praat.org> Retrieved 16.10.08.

Broen, P. A. (1972). The verbal environment of the language-learning child. *Monograph of American Speech and Hearing Association, 17*.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.

Cruttenden, A. (1986). *Intonation*. Cambridge, England: Cambridge University Press.

Endress, A. D., & Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language, 60*, 351–367.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-linguistic study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16*, 477–501.

Garnica, O. (1977). Some prosodic and paralinguistic characteristics of speech to young children. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children* (pp. 63–88). Cambridge, England: Cambridge University Press.

Gervain, J., Macagno, F., Cogoi, S., PenA, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences of the United States of America, 105*(37), 14222–14227.

Gleitmen, L. R., & Wanner, E. (1982). Language acquisition: The state of the state of the art. In E. Wanner & L. R. Gleitman (Eds.), *Language* *acquisition: The state of the art* (pp. 3–48). Cambridge: Cambridge University Press.

Gómez, R. L., & Gerken, L. A. (1999). Artificial grammar learning by one-year-olds leads to specific and abstract knowledge. *Cognition, 70*, 109–135.

Gout, A., Christophe, A., & Morgan, J. L. (2004). Phonological phrase boundaries constrain lexical access II. Infant data. *Journal of Memory and Language, 51*, 548–567.

Halle, P. A., de Boysson-Bardies, B., & Vihman, M. (1991). Beginnings of prosodic organization: Intonation and duration patterns of disyllables produced by Japanese and French infants. *Language and Speech, 34*, 299–318.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Wright Cassidy, K., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition, 26*, 268–286.

Johnson, E. K., & Seidl, A. H. (2009). At 11 months, prosody still outranks statistics. *Developmental Science, 12*, 131–141.

Jusczyk, P. W. (1989). *Perception of cues to clausal units in native and non-native languages*. Paper presented at the biennial meeting of the Society for Research in Child Development, Kansas City, Missouri.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64*, 675–687.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. A. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development, 18*, 111–116.

Langus, A., Marchetto, E., Bion, R. A. H., & Nespor, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language, 66*, 285–306.

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology*.

Morgan, J. L. (1986). *From simple input to complex grammar*. Cambridge, MA: MIT Press.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learners: Contributions of prosodic and morphological of phrases to the acquisition of language. *Cognitive Psychology, 19*, 498–550.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1989). Facilitating the acquisition of syntax with cross-sentential cues to phrase structure. *Journal of Memory and Language, 28*, 360–374.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 756–766.

Nazzi, T., Kemler Nelson, D. G., Jusczyk, P. W., & Jusczyk, A. M. (2000). Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. *Infancy, 1*, 124–147.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris Publications.

Newport, E., Gleitman, H., & Gleitman, L. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. Snow & C. Ferguson (Eds.), *Talking to children: Language input and acquisition* (pp. 109–149).

Peters, A. M. (1983). *The units of language acquisition*. Cambridge: Cambridge University Press.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Seidl, A., & Johnson, E. K. (2006). Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science, 9*(6), 565–573.

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language, 57*, 24–48.

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*, 193–247.

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences, 108*(15), 6038–6043.

Soderstrom, M., Kemler Nelson, D. G., & Jusczyk, P. W. (2005). Six-month-olds recognize clauses embedded in different passages of fluent speech. *Infant Behavior and Development, 28*, 87–94.

Soderstrom, M., Seidl, A., Kemler Nelson, D. G., & Juscayk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language, 49*, 249–267.

Takahashi, E., & Lidz, J. (2007). Beyond statistical learning in syntax. *Proceedings of generative approaches to language acquisition*.

Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development, 3*, 1–42.