Cognition, 51 (1994) 237–265
0010-0277/94/$07.00

When prosody fails to cue syntactic structure: 9-month-olds' sensitivity to phonological versus syntactic phrases

LouAnn Gerken*, Peter W. Jusczyk, Denise R. Mandel

Department of Psychology and Center for Cognitive Science, Park Hall, State University of New York at Buffalo, Buffalo, NY 14260, USA

Received February 2, 1993, final version accepted September 14, 1993

Abstract

According to prosodic bootstrapping *accounts of syntax acquisition, language learners use the correlation between syntactic boundaries and prosodic changes (e.g., pausing, vowel lengthening, large increases or decreases in fundamental frequency) to cue the presence and arrangement of syntactic constituents. However, recent linguistic accounts suggest that prosody does not directly reflect syntactic structure but rather is governed by independent prosodic units such as phonological phrases. To examine the implications of this view for the prosodic bootstrapping hypothesis, infants in Experiment 1 were presented with sentences in which pauses were inserted either between the subject noun phrase (NP) and verb or after the verb. Half of the infants heard sentences with lexical NP subjects, in which prosodic structure is consistent with syntactic structure. The other half heard sentences with pronoun subjects, in which prosodic structure does not mirror syntactic structure. In a preferential listening paradigm, infants in the lexical NP condition listened longer to materials containing pauses between the subject and verb, the main syntactic constituents. However, in the pronoun NP condition, infants showed no difference in listening times for the two pause locations. To determine if other sentence types containing pronoun subjects potentially provide*

**Corresponding author*. Email: psygerkn@ubvms.bitnet.

This research was supported in part by an NSF grant (DBS91020952) to L.A.G. and an NIH grant (HD-15795) to P.W.J. We thank Ann Marie Jusczyk and Nancy Redanz for their help in testing subjects, Stephen Yeoh for his assistance in programming the experiment, and Lila Gleitman, Deborah Kemler Nelson, Elissa Newport, Alice Turk, Irene Vogel and two anonymous reviewers for helpful comments and suggestions on an earlier draft.

SSDI 0010-0277(93)00584-T

information about the syntactic constituency of these elements, infants in Experiment 2 heard yes–no questions with pronoun subjects, in which the prosodic structure reflects the constituency of the subject. Infants listened longer when pauses were inserted between the subject and verb than after the verb. Taken together, our results suggest that the prosodic information in an individual sentence is not always sufficient to assign a syntactic structure. Rather, learners must engage in active inferential processes, using cross-sentence comparisons and other types of information to arrive at the correct syntactic representation.

Introduction

Prosodic cues to syntax acquisition

Can first language learners uncover the syntactic structure of their language from acoustic information in the speech stream? One potential cue to the syntactic organization of a language is acoustically salient prosodic changes such as pausing, vowel lengthening and fundamental frequency change. It has long been noted that such prosodic information tends to coincide with syntactic boundaries (Beckman & Edwards, 1990; Cooper, 1975; Cooper & Paccia-Cooper, 1980; Crystal, 1969; Klatt, 1975; Lehiste, Olive, & Streeter, 1976; Luce & Charles-Luce, 1983; Nakatani & Dukes, 1977; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Scott, 1982; Scott & Cutler, 1984; Streeter, 1978). Based on the correlation between prosodic changes and syntactic boundaries, numerous researchers have proposed variants of a prosodic bootstrapping hypothesis, in which young learners use prosody to cue the locations of syntactically relevant units, such as phrases and clauses, and the structural relations among these units (Brown, 1973; Fisher, 1991; Gleitman, Gleitman, Landau, & Wanner, 1988; Gleitman & Wanner, 1982; Hirsh-Pasek et al., 1987; Jusczyk et al., 1992; Kemler Nelson, Hirsh-Pasek, Jusczyk, & Wright Cassidy, 1989; Lederer & Kelly, 1991; Mazuka, 1991; Morgan, 1986; Morgan, Meier, & Newport, 1987; Morgan & Newport, 1981; Peters, 1983).

The view that prosodic cues to syntactic structure are important in language learning is supported by two lines of research. First, caregiver speech to children contains prosodic cues that are much richer and potentially more reliable than adult-directed speech (e.g., Bernstein-Ratner, 1986; Broen, 1972; Fernald, 1985; Fernald & Kuhl, 1987; Fernald & Simon, 1984; Garnica, 1977; Jusczyk et al., 1992; Morgan, 1986; Snow, 1972). Second, young language learners demonstrate remarkable sensitivity to prosodic regularities in their native language. In speech perception studies, infants prefer speech in their mother's language over speech in a foreign language, even when both samples are low-pass filtered to leave intact only prosodic cues (Bahrick & Pickens, 1988; Mehler et al., 1988). Infants as young as 6 months are able to distinguish between English and Norwegian word lists that have been low-pass filtered (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993). Six-month-olds are also sensitive to configurations of prosodic cues to clause boundaries (Hirsh-Pasek et al., 1987; Kemler Nelson et al., 1989), and infants as young as 9 months are sensitive to cues for major phrase boundaries (Jusczyk et al., 1992). In the phrase and clause studies, infants attended longer to passages in which pauses appear in syntactic boundary positions (as marked by other cues such as lengthening and falling intonation) than in non-boundary positions. Similarly, in studies of speech production, researchers have demonstrated that infants as young as 8 months exhibit in their babbling the canonical prosodic patterns of their native language (de Boysson-Bardies, 1993; de Boysson-Bardies, Sagart & Durand, 1984).

The imperfect relation between prosody and syntax

The correlation between prosody and syntax, the quality of prosodic cues in caregiver speech, and young children's sensitivity to these cues support the hypothesis that children could use prosody to provide initial information about syntactic structure. To see how this might work, consider the sentence illustrated in (1). Here, a slow careful talker is likely to produce prosodic boundary cues between the subject noun phrase (NP) and the verb phrase (VP), thereby cueing a sensitive listener to the major syntactic constituents in the sentence and their structural relation.

(1) $[_S[_{NP}$ The little girl over there] / $[_{VP}$ is the one who likes snails]].

But the correlation between prosody and syntax is not perfect (Beckman & Edwards, 1990; Ferreira, 1993; Gee & Grosjean, 1983; Suci, 1967; also see Hillinger, James, Zeil, & Prato, 1976; Levelt, 1969; Martin, 1970). In particular, although the presence of prosodic changes such as pausing or fundamental frequency change may be good cues to the presence of a syntactic boundary, prosody does not reliably cue the hierarchical structure of syntax (see Chomsky & Halle, 1968; Morgan et al., 1987, for discussion). That is, although prosody *does* highlight syntactic units, the structural relations among these units are not always reflected. An often-cited example of the imperfect relation between prosody and syntax is illustrated in (2) (Chomsky & Halle, 1968; Morgan et al., 1987).

(2a) $[_S[_{NP}$ This] $[_{VP}$ is $[_{NP}$ the dog that chased the cat that bit the rat that lived in the house that Jack built]]].

(2b) This is the dog/ that chased the cat/ that bit the rat/ that lived in the house that Jack built.

(2c) $[_{S}[_{NP}$ This] $[_{VP}$ is [[[[$[_{NP}$ [[[[the dog that chased [[$[_{NP}$ the cat that bit [$[_{NP}$ the rat that lived in $[_{NP}$ the house that Jack built]]]]]]]]]]]]]]]].

As in (1), the bracketing in (2a) illustrates the major syntactic constituents in the syntactic structure of the sentence. However, unlike sentence (1), the major constituents in (2) are not reflected by the prosody. Instead, a talker is likely to produce prosodic boundary cues at the locations illustrated by slashes in (2b). Even the major syntactic constituents embedded within the VP, as illustrated in (2c), are not cued by pausing or other prosodic information. Thus, sentences like (2) that contain a string of relative clauses demonstrate that prosody does not perfectly reflect syntax. Nevertheless, the slashes used to denote prosodic boundary cues such as pausing *do* mark one type of syntactic boundary, namely the one between NPs and their relative clauses. The point here is that the hierarchical syntactic structure of sentences like (2) is not reflected by prosodic information.

How do cases such as the one illustrated in (2) bear on the prosodic bootstrapping hypothesis? Insofar as learners want simply to identify linguistically relevant units in the speech stream, prosody remains a potentially reliable cue, because prosodic boundary cues occur at syntactic boundaries. However, if learners also attempt to make inferences about syntactic *structure* from prosodic information, then examples like the one in (2) might cause a problem. This is because prosody marks some lower-level syntactic units (such as relative clauses) without marking the higher-level units which contain them (such as subjects and VPs). In response to such examples, supporters of the prosodic bootstrapping hypothesis are quick to point out that young learners are not likely to hear or interpret sentences with many embedded clauses and that in the majority of sentences children hear, prosodic information correlates well with syntactic structure (e.g., Morgan et al., 1987). However, in this paper, we will consider another example of the imperfect relation between prosody and syntax that is much more common in speech directed at young learners. An example of this is illustrated by the sentences in (3).

(3a) Joe / kissed the dog.
(3b) He kissed (/) the dog.

In sentence (3a), just as in sentence (1), a talker is likely to produce prosodic boundary cues after the subject NP, thereby potentially indicating to the listener that the subject NP and VP are the major syntactic constituents of the sentence. But in sentence (3b), a talker is likely to either not produce any salient prosodic boundary cues or to produce boundary cues between the verb and the object NP (e.g., Fisher, 1991; Gee & Grosjean, 1983). Consistent with the prosodic analysis in (3a,b), Read and Schreiber (1982) observed that young school-aged children have difficulty in a meta-linguistic task identifying pronoun subjects as constituents (also see Ferreira & Morrison, in press). These researchers suggest that the reason for children's poor performance is that they have difficulty treating the weakly stressed pronoun as a potential constituent (see Gerken, 1993).[1] Another finding that is consistent with prosodic differences between pronoun and lexical subjects is that 2-year-olds omit more object article omissions from sentences like (3b) than (3a) (Gerken, 1991). Gerken (1993) has explained these differences by proposing that children produce sentences like (3a) with a prosodic boundary between the subject and verb but sentences like (3b) with a boundary between the verb and object. This results in the article in (3b) beginning a weak–strong metrical pattern, which children learning several languages find difficult to produce (Allen & Hawkins, 1980; Demuth, 1992; Gerken, 1991, 1993, in press-a; Gerken, Landau, & Remez, 1990; Pye, 1983; Wijnen, Krikhaar, & den Os, in press).

If learners use prosodic cues to gain evidence about the syntactic structure of their language, then sentences like (3b) could well present a problem. One potential difficulty is that they might be given no prosodic cues to syntactic structure if the talker failed to produce prosodic changes between any of the constituents. Or worse, learners might be misled to infer that the subject and verb form one major constituent and the object another, if the talker produced prosodic changes between the verb and object. Thus, insofar as the prosodic bootstrapping hypothesis postulates that learners use prosodic cues to *directly* infer syntactic structure, sentences like (3b) represent a potential problem for this approach. Thus, the main goal of our research was to determine if sentences like (3b) represent a class of input to children in which prosodic changes fail to directly cue syntactic structure.

If this is the case, then an alternative to the direct version of the prosodic bootstrapping hypothesis is offered by a current linguistic account of prosody – prosodic phonology. In this framework, the difference in prosodic patterns in sentences like (1) and (3a) on the one hand, and sentence (3b) on the other, occurs because a weakly stressed pronoun subject is phonologically joined (cliticized) to the following strongly stressed verb. Thus, prosodic boundary cues do not occur between the subject and verb, because this is not a prosodic boundary, even though it is a syntactic boundary (Hayes, 1989; Nespor & Vogel, 1986; Selkirk, 1981). Note that within the prosodic phonology framework, syntactic structure only indirectly influences sentential prosody via an independent prosodic structure, which tends to be flatter and less detailed than syntactic structure. Thus, in sentences like (2), prosodic boundary cues do not reflect the fact that each NP is embedded in the one preceding it. Prosodic structures are

[1] Gleitman (personal communication) suggests that this is the reason for young children's omission of pronoun subjects (also see Gerken, 1991).

also influenced by the phonological status of the words they contain. Thus in a sentence like (3b), the weak stress received by the pronoun subject gives it a different prosodic status than a lexical NP subject, even though both comprise the same syntactic constituent.[2] Within the prosodic phonology framework, rather than directly mapping prosodic cues onto syntactic structure, learners might use prosodic changes to infer prosodic structure and in turn infer syntactic structure from prosodic structure.

Because a secondary goal of our research is to consider the implications of prosodic phonology for our research and theories of language acquisition, it would be useful to provide a brief outline of this theory.

Introduction to prosodic phonology

As background for the research presented here, we will briefly discuss three hierarchically arranged categories of prosodic phonology that apply specifically to sentences: clitic group, phonological phrase, and intonational phrase. All three are illustrated in the structures in Fig. 1.

Let us consider a prosodic analysis of the sentence "Joe kissed the dog, and he called up his brilliant doctor" in the framework of Hayes (1989) and Nespor and Vogel (1986) (Fig. 1). The clitic group is the first prosodic category that is applicable to multiword utterances. A clitic is typically defined as a function word or morpheme that phonologically joins to an adjacent host content word, as in auxiliary contraction (e.g., Sam'*s* going to the store; Kaisse, 1983; Klavans, 1985; Selkirk, 1984; Zwicky & Pullum, 1983). A clitic group (or "prosodic word" as it is sometimes called; e.g., Selkirk & Shen, 1990) is defined as a non-clitic host and all adjacent clitics (Hayes, 1989; Nespor & Vogel, 1986). In the structures in Fig. 1, the determiners "the" and "his", the coordinator "and", the pronoun "he" and the particle "up" are clitics, and they form clitic groups with adjacent content words.

Adjacent clitic groups are joined together to form phonological phrases (Hayes, 1989; Nespor & Vogel, 1986). Phonological phrases are composed of clitic groups within a single syntactic phrase, up to and including the head of the phrase (e.g., Jackendoff, 1977). Heads must be members of the syntactic categories noun, verb or adjective (Nespor & Vogel, 1986). Thus, typical phonological phrases include lexical NPs (e.g., "Joe", "the dog", "his brilliant doctor") and verbs with their particles and auxiliaries (e.g., "kissed", "called up"). In some languages, such as Italian and English, a phonological phrase that

[2] Of course, there are circumstances in which a pronoun NP might not be cliticized to a following verb, such as when it receives contrastive stress (e.g., HE kissed the dog. SHE just watched). See Experiment 2 for further discussion of the effect of discourse on prosody.

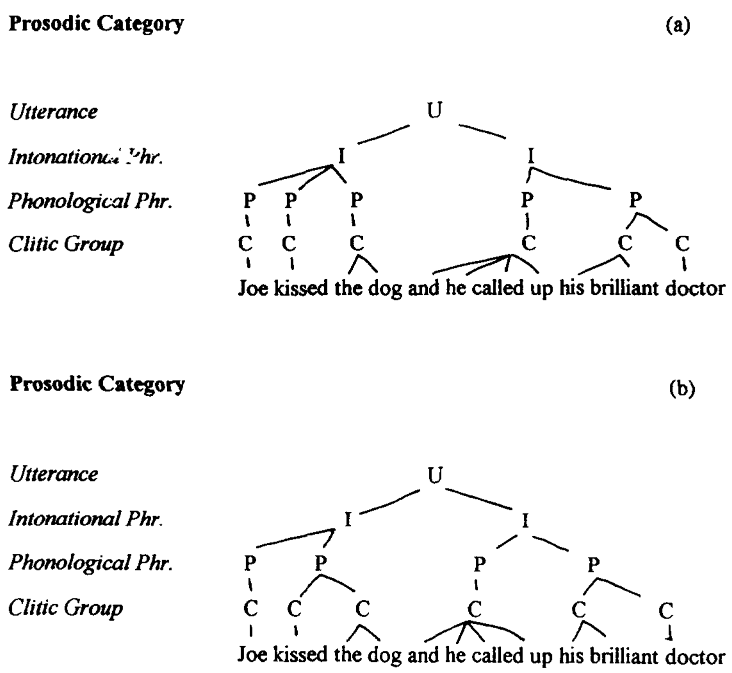

Figure 1. (a) Example of hierarchically arranged prosodic categories. (b) Phonological phrase restructuring of the phrase "kissed the dog".

comprises only a single clitic group, and that is the complement of a preceding head, may be incorporated into the phonological phrase containing the head (phonological phrase restructuring, Nespor & Vogel, 1986). Thus, an object NP comprising only a single clitic group may be included in the phonological phrase with the verb (Hayes, 1989, p. 216). The phonological phrase-restructuring option is not exercised in Fig. 1a, but it is in Fig. 1b, in which the first-clause object NP "the dog" is incorporated into the phonological phrase with the verb "kissed". The second-clause object NP "his brilliant doctor" cannot be incorporated into the phonological phrase with the verb "called up", because the object contains two clitic groups ("his brilliant" and "doctor"). Phonological phrases are joined into intonational phrases, the last prosodic level before the utterance. Figures 1a,b both show the sentence in question divided into two intonational phrases.

Within the prosodic phonology framework, prosodic structure, and not syntactic structure, serves as the domain of a variety of phonological processes. For example, word-final segment /v/ can be deleted when the following word in the clitic group begins in a non-syllabic consonant (Hayes, 1989; Selkirk, 1972). This is illustrated in (4a) (subscript "C" indicates that the bracketed material constitutes a clitic group). However, as illustrated in (4b), if there is no following word contained within the same clitic group, /v/-deletion is not allowed.

(4a) [$_C$ Please] [$_C$ leave me] [$_C$ alone].
$v \rightarrow 0$

(4b) [$_C$ Please] [$_C$ leave] [$_C$ men] [$_C$ alone].
$^*v \rightarrow 0$

Let us now consider an account of the prosody of sentences (3a,b), within the framework of prosodic phonology that was just outlined. In this framework, the lexical subject NP of sentence (3a) comprises a single phonological phrase. In contrast, the pronoun subject in (3b) forms a clitic group with the following verb; the subject and verb together comprise the first phonological phrase of the sentence. Therefore, if the phonological phrase is the domain of prosodic changes such as pausing, then a pause cannot be inserted between the subject NP and verb in (3b), even though these are the major syntactic constituents of the sentence. Now consider why prosodic boundary cues can be optionally produced between the verb and object NP in (3b). Exercising the phonological phrase-restructuring option, the object NP in (3b) is incorporated into the phonological phrase containing the verb (and pronoun subject). Under this scenario, there would be no prosodic boundary cues in the sentence. However, if the option is not exercised, the subject-plus-verb comprises one phonological phrase and the object NP another. Consequently, prosodic boundary cues would be produced between the two phonological phrases. Conditions that affect whether or not phonological phrase restructuring occurs include how many other non-optional phonological phrase boundaries there are in the utterance, the length of the object NP, and the placement of focal stress (Ferreira, 1993; Gee & Grosjean, 1983; Martin et al., 1971; Vogel & Kenesei, 1990).

Implications for prosodic bootstrapping

Let us now turn to the implications of the imperfect relation between syntax and prosody for the viability of the prosodic bootstrapping hypothesis. As noted earlier, previous research has demonstrated that infants as young as 9 months of age are sensitive to prosodic cues to phrases (Jusczyk et al., 1992). In that study, a woman was recorded talking to an 18-month-old child. Pauses were inserted in the sentences either between the subject and verb or after the verb. Nine-month-old infants who were presented with each type of stimulus in a preferential listening paradigm demonstrated a significant preference for sentences in which the pause was inserted between the subject and verb. That is, they listened longer to sentences in which prosody reflected major syntactic constituents. However, because the study did not separately examine sentences with lexical NP and pronoun subjects, it is possible that infants' apparent preference for syntactic constituency was due mainly to cases in which prosody and syntax coincide. In fact, Jusczyk and his colleagues used two different types of stimulus materials: spontaneous speech samples from a mother to her child, and storybook samples read to a child. The latter samples were chosen explicitly to have long subject NPs. Consequently, none of the sentences in these samples had pronoun subjects. With respect to the spontaneous speech samples, only about 15% of the sentences had structures of the type found in example 3b.[3] In Experiment 1, we used the same procedure as Jusczyk and his colleagues, but we compared infants' responses to sentences with lexical NP versus pronoun subjects.

EXPERIMENT 1

Experiment 1 was designed to determine whether 9-month-old infants are sensitive to prosodic differences between sentences with lexical NP and pronoun subjects. To investigate these issues, we constructed two sets of stimuli. One set was composed entirely of sentences with lexical NP subjects. The second set was identical to the first except that pronouns were substituted for all lexical NP subjects. Based on the findings of Jusczyk et al. (1992), we expected that, for sentences with lexical NP subjects, infants would listen longer when pauses were inserted between the subject and verb than when pauses were inserted after the verb.

The predictions for the sentences with pronoun NP subjects depend on whether syntactic structure or prosodic structure governs prosodic changes, and in the latter case, whether or not the speaker exercises the phonological phrase-restructuring option. If, contrary to the theory of prosodic phonology, syntactic boundaries directly control the production of prosodic boundary cues, there should be boundary cues after the pronoun subject. Therefore, just as in the case of lexical NP subjects, infants should listen longer to sentences with pauses inserted between the pronoun subject and verb over sentences with pauses inserted after the verb.

In contrast, a different pattern of results is predicted by a theory such as prosodic phonology, in which prosodic boundaries, not syntactic boundaries, control the production of boundary cues. Recall that simple declarative sentences with pronoun subjects like (3b) can be produced either with or without prosodic boundary cues between the verb and object NP. This is because a talker may or may not choose to incorporate the verbal complement into the phonological phrase containing the verb (and subject). If the option is exercised, there should be no boundary cues between the verb and its complement. If the talker does not

[3]Interestingly, about 27% of the spontaneous samples consisted of yes–no questions in which an auxiliary preceded a pronoun subject (e.g., "Can you find it?"). We will return to this point in Experiment 2.

exercise phonological phrase restructuring and leaves the verbal complement in a separate phonological phrase from the verb, there should be boundary cues after the verb. Whichever prosodic structure the talker chooses, infants should not detect boundary cues after the subject. They should therefore not listen longer to sentences with a pause inserted after the pronoun subject, as they should in sentences with lexical NP subjects. If this prediction is borne out, the prosodic bootstrapping hypothesis needs to be revised to explain how learners are able to assign syntactic structures to sentences with pronoun subjects.

Method

Subjects

Forty-eight American infants of approximately 9 months of age were tested. The infants had an average age of 39 weeks, 6 days (range: 36 weeks, 3 days to 42 weeks, 5 days). Seven additional infants were tested but not included for the following reasons: failed to look for an average of at least 3 s to each side ($n = 4$), cried ($n = 2$), and parent failed to center the infant on her lap ($n = 1$).

Stimuli

The materials were prepared from storybooks in a fashion similar to the one used by Jusczyk et al. (1992). The sentential materials were modifications of texts from two children's storybooks. Sentences were rewritten to provide samples that consisted of approximately five clauses. For the lexical NP samples, every sentence began with either a proper noun ("Sammy"), or a determiner–noun sequence ("the caterpillar"). The pronoun NP samples were identical except that the pronoun "he" was substituted for lexical NP subjects. Examples of a lexical NP and pronoun version of a sample are shown in Table 1.[4] In all, there were eight lexical NP and eight pronoun NP samples. The new text was glued into the storybooks covering the original words so that they appeared in conjunction with the appropriate pictures. These materials were given to a college-aged woman to

[4]The reader might note that there is a difference in the number of syllables between the lexical NP and pronoun subject versions of the stories. Data from Gee and Grosjean (1983) on adult speakers and data from Gerken (1993) on 2-year-old speakers suggest that it is the lexical status of the subject (noun vs. pronoun) and not the number of syllables that determines whether it will be prosodically marked as a constituent. However, even if it were the case that monosyllabic subjects were less likely to be prosodically marked as constituents than subjects containing more syllables, this would not be important for the main question under consideration in this paper – whether there is a set of sentences in which prosody fails to provide infants with cues to syntactic constituency. Prosodic phonology suggests that declarative sentences with pronoun subjects constitute such a set, but if the set is in fact larger than this our main point would not be affected.

read. She was not informed as to the purpose of the experiment but was simply instructed to read the passages aloud in an animated way as if speaking to a 2-year-old. Her readings of the passages were recorded on audio tapes in a sound-attenuated room with a Shure microphone (SM10A) and a Revox (A77) tape recorder.

The tape recording was used to prepare the stimulus materials for the experiment. Each of the 16 sequences (i.e., 8 lexical NP samples and 8 pronoun NP samples) was digitized and stored on a VAXStation 3176 computer using a 12-bit A/D converter.[5] Two versions were prepared for each sequence by inserting 1 s pauses into the utterances at different locations. Pauses were inserted only at zero crossings in the waveform so as not to produce transients in the signal. For the after subject versions, the pauses were inserted just after the subject NP and before any auxiliaries. The after verb versions were prepared by inserting the pause immediately after the main verb. Finally, in order to avoid inserting a pause immediately after a sample began, each sample began with either a long lead-in phrase or with a lead-in sentence in which no pauses were inserted. Except for the placement of the pauses, the after subject and after verb versions of each sample were identical. For the samples presented on the test trials, the mean duration was 17.14 s, with a range of 14.78–21.88 s. Two versions of one sample are shown in Table 2.

Apparatus

The digitized files were transferred from the VAXStation to a PDP 11/73 computer. During the experiment, the PDP 11/73 controlled the presentation of

Table 1. *Sample of lexical and pronoun NP materials*

Lexical NP subject version
Out of the egg came a tiny and very hungry little animal. The caterpillar ate four strawberries. But the caterpillar was still hungry. The caterpillar ate five oranges. But the caterpillar was still hungry.

Pronoun subject version
Out of the egg came a tiny and very hungry little animal. He ate four strawberries. But he was still hungry. He ate five oranges. But he was still hungry.

[5]In the Hirsh-Pasek et al. (1987) study and subsequent investigations, pauses longer than 400 ms were removed from the spontaneous speech samples. This was done, because many of the samples contained multiclausal sentences with substantial pausing. Thus, inserting new pauses in the non-coincident versions of the samples would have resulted in one version containing many more pauses than the other. We had considered following the same procedure in the current study, but the relatively short length of our sentences and the fact that they were read resulted in virtually no long pauses in any of the samples, making it unnecessary.

Table 2. *Sample of pauses inserted after subjects and after verbs in Experiment 1*

Pause after subject
The next day was Sunday again. The caterpillar / ate one nice green leaf, and after that the caterpillar / felt much better. The caterpillar / wasn't hungry anymore.

Pause after verb
The next day was Sunday again. The caterpillar ate / one nice green leaf, and after that the caterpillar felt / much better. The caterpillar wasn't / hungry anymore.

the lists and recorded the observer's coding of the infant's responses. The audio output for the experiment was generated from the digitized waveforms of the samples. A 12-bit D/A converter was used to recreate the audio signal. The output was fed through anti-aliasing filters and a Kenwood audio amplifier (KA 5700) to 7-inch Advent loudspeakers mounted on the side walls of the testing booth. The experiment was conducted in a three-sided test booth constructed out of pegboard, with panels of 4 feet by 6 feet on three sides and open at the back. This made it possible for an observer to look through one of the existing holes to monitor the infant's head turns. Most of the pegboard was backed with white cardboard to guard against the possibility that the infant might respond to movements behind the panel. However, there was a small area of about 30 cm^2 without backing behind the center panel to permit the experimenter to observe the infant. In addition, there was a hole of about 8 cm in diameter cut into the pegboard about 8 cm below a green light which was mounted at the infant's eye level. A JVC compact video camera (GR-303) was aligned with the hole behind the pegboards. The camera was used with the existing lighting in the room. Each of the side panels also had a red light and a loudspeaker mounted at the infant's eye level. A white curtain suspended around the top of the booth shielded the infant's view of the rest of the room. A computer terminal and response box were located behind the center panel out of the infant's view. The response box, which was connected to the computer, was equipped with a series of buttons that started and stopped the flashing center and side lights, recorded the direction and duration of head turns, and terminated a trial when the infant looked away for more than 2 s. Information about the direction and duration of head turns and the total trial duration was stored in a data file on the computer and all sessions were videotaped to permit reliability checks on judgments about the durations of head turns.

Design and procedure

Half of the infants were assigned randomly to the lexical NP condition and the other half to the pronoun NP condition. The procedure was a modified version of one originally developed by Fernald (1985). Each infant was held on a parent's lap. The parent was seated in a chair in the center of the test booth. The infant completed a 4-trial familiarization phase (the after subject and after verb versions of two samples) and a 12-trial test phase. The after subject versions of the samples were consistently played through the loudspeaker on one side panel, and the after verb versions through the loudspeaker on the other side panel. (The side of presentation was counterbalanced across subjects.) The familiarization phase was intended to acquaint the infant with the assigned position of each type of sample. The ordering of the stimuli during the test trials was random, subject to the constraint that no more than two samples of the same type could occur in a row. During the test phase, the infant heard six samples of each type, each one in its two versions. Each trial began by blinking the green light on the center panel until the infant had oriented in that direction. Then, the center light was extinguished and the red light above the loudspeaker on one of the side panels began to flash.[6] When the infant made a head turn of at least 30° in the direction of the loudspeaker, the next sample appropriate to that side began to play and continued until its completion or until the infant failed to maintain the 30° head turn for 2 consecutive seconds (e.g., if the infant turned back to the center or the other side, looked at the mother, the floor or the ceiling). If the infant turned briefly away from the target by 30° in any direction, but for less than 2 s, and then looked back again, the time spent looking away was not included in the orientation time. During the familiarization trials, the red light was extinguished when the list began, but during the test trials the light remained on for the entire duration of the trial. An observer hidden behind the center panel looked through a peephole and recorded the direction and duration of the infant's head turns using a response box. The observer was not informed as to which loudspeakers played the after subject and after verb versions of the samples. This was possible because the assignment of the versions to the left or right side was determined by the computer and not revealed to the observer until the completion of the test session. The loudness levels for the samples were set by a second assistant, who was not involved in the observations, at 72 ± 2 dB (C) SPL using a Quest (Model 215) sound level meter. In addition, both the observer and the infant's parent listened over Sony (MDR-V600) headphones to continuous music, which proved to be an excellent masking stimulus (see below). Parents and observers reported that with this background they were unaware of either the location or the nature of the stimulus on the trial.

[6]During familiarization trials, the blinking red light was extinguished as soon as the infant oriented to the side and the sample began to play. However, during the test trials, the blinking light remained on until the trial ended. Extensive pilot testing convinced us that this was the best way to handle the lights during the procedure. Leaving the flashing light on during familiarization trials seemed to habituate the infants to lights and resulted in very short orientation times during the test trials. Moreover, the infants were less likely to complete the full set of test trials under these circumstances.

We performed two procedures to ensure that the observer who controlled the beginnings and ends of trials did not bias the results. The first was reliability coding of the videotapes made during each session. Videotapes of 40 of the 48 subjects tested were available for reliability coding in which a different observer from the one who made the original observations viewed the videotapes with the soundtracks turned off. The second observer recorded looking times from the videotapes with the same type of response box used in the original test session. For a given infant, the second observer provided looking time measures for each test trial. A Pearson product–moment correlation was performed on the difference in looking times for the after subject and after verb samples for each infant noted by the "live" and "videotape" observers ($r_{39} = .947$). It indicated very high agreement across these observations. As a further check on observer agreement, we took a closer look at possible discrepancies between the live and videotape looking time judgments by inspecting differences in recorded times on a trial-by-trial basis. On 312 trials (65%), the discrepancy between the recorded time of the live observer and a different observer viewing the videotape was less than 0.5 s. On 82 trials (17%), there was a discrepancy of 1 s or more. For these discrepancies, we examined the pattern of differences between the two observers to determine whether there is any systematic tendency on the part of the live observer to overestimate or underestimate the looking times on the after subject or after verb trials for either the lexical or pronoun materials. There was no systematic difference between observers across these trials for either the lexical NP ($t_{36} = 1.15$, $p > .25$) or pronoun NP ($t_{42} = 0.76$, $p > .45$) materials. Finally, we note that the pattern of significant results reported for the live and videotape observers in the present study is identical.

The high agreement between the times recorded during the live observations and those made from videos without soundtracks is an indication that the experimenter's judgments of looking times were not unconsciously biased by the possibility that she may have heard a portion of the soundtrack despite the masking noise. Nevertheless, as a further check on the effectiveness of the masking music in blocking out the stimulus materials, we also performed an experiment with 8 normal-hearing adult listeners who were presented with the stimuli from the lexical NP condition. Each listener was seated in the chair in the middle of the test apparatus. The nature of the stimulus materials was carefully explained to the listeners. They were told that some of the samples had pauses after the subjects and that other samples had pauses after the verbs. Furthermore, the listeners were told that all of the samples of a given type would be played on the same side of the room. Their task was to circle the correct response (i.e. "after subject" or "after verb") on an answer sheet after each trial. Each listener heard a different order of 12 test trials. In the first phase of the experiment, the listeners heard the stimuli with no masking noise. Not surprisingly, performance was virtually perfect (one listener missed the first trial but correctly identified the samples on all 11 remaining trials). Thus, it is clear that the listeners understood the nature of the task. For the next phase, they were instructed that they would be presented with the same materials but in a new test order. They were told that this time they would be wearing headphones over which loud masking music would be played, but that their task was still to identify the type of sample that would be playing. Listeners were told that they would likely have great difficulty in hearing the stimuli and to use the flashing light on the side panel as an indication of when each sample was playing. They were told to respond on each trial, even if they had to guess. This listening situation was meant to mimic that of the experienced observer in the infant experiment. Upon completion of the experiment, all listeners reported that they were unable to hear the stimuli over the masking music. Their subjective impressions were reflected in their performance on the task because there was no evidence that they significantly discriminated the two types of samples when masking music was present ($t_7 = 0.23$, $p > .80$). Four of the listeners reported that they employed a guessing strategy in which they decided to choose one side throughout to associate with the "after subject" pauses, and the other side with the "after verb" pauses. Of these 4 listeners, 2 guessed correctly and achieved perfect scores; the other 2 guessed incorrectly and achieved no correct answers. Across all listeners, there were 51 correct responses and 45 incorrect responses. The results of this experiment strongly suggest that the observer in the infant experiment could not discriminate the two versions of the stimuli over the masking music, thereby making it highly unlikely that this individual could unconsciously influence the results.

Results

The amount of time that each infant oriented to the loudspeaker on each trial was recorded. In the lexical NP condition, the average looking time was 8.54 s (SD = 3.66 s) when pauses occurred after the subject and 6.37 s (SD = 3.09 s) when pauses occurred after the verb. Seventeen of the 24 infants had longer average looking times for the stimuli in which pauses occurred after the subject. In the pronoun NP examples, the average looking time was 7.51 s (SD = 3.11 s) when pauses occurred after the subject and 7.21 s (SD = 2.8 s) when pauses occurred after the verb. Only 12 of the 24 infants had longer average looking times for the stimuli in which pauses occurred after the subject.

To investigate the reliability of the apparent differences between the lexical NP and pronoun NP conditions, listening times were subjected to a 2 Sentential Subject (lexical NP vs. pronoun NP) × 2 Pause Location (after subject vs. after verb) analysis of variance by subjects and by items. Infants in the lexical NP and pronoun NP conditions did not differ significantly in their overall listening times ($F_s(1, 46) = 1.84$, n.s.; $F_i(1, 5) = 0.55$, n.s.). However, infants listened longer to samples that contained pauses after the subject than to samples that contained pauses after the verb; this effect was significant in the analysis by subjects and marginal in the analysis by items ($F_s(1,46) = 6.25$, $p = .02$; $F_i(1,5) = 6.18$, $p = .056$). The interaction between sentential subject and pause location was marginal in the analysis by subjects and highly significant in the analysis by items ($F_s(1,46) = 3.13$, $p = .08$; $F_i(1,5) = 30.13$, $p = .003$). Planned t-tests revealed significantly longer listening times for samples in which pauses were inserted after the subject in the lexical NP condition ($t_s(23) = 2.54$, $p < .02$; $t_i(5) = 3.62$, $p < .02$). However, there was no significant listening difference between the two pause locations in sentences with pronoun subjects ($t_s(23) = 0.67$, n.s.; $t_i(5) = .73$, n.s.).

In order to gain some indication of just what kind of information their subjects might be responding to in the speech signal, Jusczyk et al. (1992) conducted acoustic analyses of their samples. Consistent with a number of previous reports (e.g., Cooper & Paccia-Cooper, 1980; Grosjean & Gee, 1987; Klatt, 1975; Martin, 1970; Nakatani & Dukes, 1977; Price et al., 1991), they found evidence of significant changes in the pitch and duration of syllables in the vicinity of major phrase boundaries. In particular, Jusczyk et al. reported that pitch tended to drop and syllable durations increased over the last several syllables just prior to the boundary between the subject and predicate phrases. We conducted similar analyses for our lexical NP and pronoun subject materials. Unlike Jusczyk et al., we found that infants appeared to respond to *combinations* of statistically reliable changes in both pitch and syllable duration, *regardless of the direction of these changes*. Below, we will present the pitch and vowel change data for the two pause locations in each version of the story. Following Jusczyk et al., we measured the average pitch of vowels of the last few syllables just prior to the phrasal boundary. Average pitch was determined by counting individual pitch pulses per unit time from the center of each vowel. Whenever possible, we sampled at least 50 ms from each vowel. However, our samples were unlike the Jusczyk et al. samples in that they typically contained only two syllables prior to the phrasal boundary, as opposed to the long subject NPs used in the earlier study. For this reason, we report only changes over the last two syllables.[7]

In the lexical NP condition, there was a significant pitch fall (31 Hz) over the

[7]It is appropriate to inject a notion of caution regarding the kinds of conclusions that we can draw from these acoustic analyses. As Jusczyk et al. (1992) pointed out, samples like those used in the present study are not optimal for providing a very informative analysis of syllable durations. Our stimuli were designed with perception, rather than production, experiments in mind. Ideally, in studying how differences are conveyed in production, one would want to construct the samples in such a way that would permit measurements of identical items at boundary and non-boundary locations within the phrases (e.g., Morgan, 1986; Price et al., 1991), or employ computational algorithms that normalize vowel durations across different phonetic contrasts. Only in this way can one truly investigate how the presence of a particular boundary influences the overall duration of syllables in a particular word. We were able to do this to a limited extent by comparing durational differences between the pronoun subjects in Experiments 1 and 2 (see Experiment 2).

last two syllables just prior to the boundary between the subject and verb ($t(21) = 2.28$, $p < .05$). There was also a significant decrease in syllable duration (122 ms) in the syllables prior to the boundary between subject NP and verb ($t_{21} = 5.24$, $p < .001$). At the verb–complement boundary, there was no evidence of a significant pitch change ($t_{21} = 1.65$, n.s.), but there was a significant increase in syllable duration (84 ms, $t(21) = 3.08$, $p < .01$).[8]

In the pronoun NP condition, in contrast to the lexical NP materials, there was no indication of a significant pitch change over the last two syllables prior to the boundary between the subject and verb ($t_{10} = 1.52$, n.s.), nor was there a significant duration change in the vicinity of the boundary between the subject NP and VP ($t_{10} = 1.74$, n.s.). There was a significant pitch drop (64 Hz) between the verb and complement ($t_{21} = 2.94$, $p < .01$), but the pitch change was not accompanied by a significant change in duration ($t_{21} = 0.85$, n.s.).

Discussion

We will first discuss the data from the lexical NP condition and then compare these data to those from the pronoun NP condition. Infants in the lexical NP condition listened longer to samples in which a pause was inserted between the subject NP and verb than to samples in which a pause was inserted after the verb. This finding replicates that of a previous study using the pause insertion technique, although the previous study did not separately examine sentences of different types (Jusczyk et al., 1992). We interpret infants' different responses to the pause locations as an indication that they detected other prosodic boundary cues, such as vowel lengthening and fundamental frequency change, between the lexical NP subject and verb. When a pause was inserted in this location, it was consistent with the presence of these other prosodic cues. In contrast, a pause inserted after the verb did not coincide with other prosodic boundary cues, causing infants to reject these sentences.

This interpretation of the data is supported by the acoustic analyses we performed. Pitch and duration both changed significantly prior to the boundary between the subject and verb, whereas there were not reliable changes in both cues at the verb–complement boundary. Concurrent changes in pitch and duration at the subject–verb boundary were also found by Jusczyk et al. (1992),

[8]In accordance with a suggestion from an anonymous reviewer, we also compared the duration of the final syllable of the lexical NP subject with a non-boundary control syllable exhibiting the same stress level and vowel quality. Consistent with the finding of a significant *decrease* in length, we found that the final syllable of the lexical NP was *shorter* than the control syllable. However, we also found that the final syllable of the verb was *shorter* than the control syllable, suggesting that our initial finding of lengthening in this position was due to the fact that our verbs were most often monosyllables which were preceded by weak syllables from the subject NPs.

but in that study the direction of the duration change was in the opposite direction from the one seen here. The reason for the decline in duration in the current experiment is probably attributable to the fact that this syllable was inevitably an unstressed syllable that followed a stressed one. It is possible that any significant change in pitch and duration, regardless of direction, is sufficient to indicate the presence of a boundary. However, given the problems inherent in making acoustic measurements from stimuli such as ours, this conclusion is highly speculative (see footnote 7).

The most important new information added by the current experiment is that infants responded differently to sentences with lexical NP versus pronoun subjects. Consistent with adult pausing data (e.g., Gee & Grosjean, 1983) and with the theory of prosodic phonology within linguistics (Hayes, 1989; Nespor & Vogel, 1986), infants in the pronoun NP condition failed to demonstrate a significant listening bias for materials with pauses inserted after the subject. In contrast, recall that infants did demonstrate a bias for sentences with pauses after the subject in the lexical NP condition. This finding suggests that infants' apparent bias for pauses between major syntactic constituents in the previous study by Jusczyk et al. (1992) was due to the fact that the majority of syntactic boundaries coincided with detectable prosodic boundaries.

The fact that infants detected prosodic boundary cues after the subject in the lexical NP condition but not in the pronoun NP condition suggests that our talker employed the phonological phrase-restructuring option and produced sentences with pronoun subjects as a single phonological phrase. This was not the case in the 2-year-olds studied by Gerken (1993), who appeared to divide sentences like (3b) into two phonological phrases: one containing the pronoun subject and verb and the other containing the object NP. One possible explanation for this difference is that the sentences used by Gerken (1993) were shorter than many of those used in the current experiment. Thus, if a talker is inclined to produce one-clause sentences as two phonological phrases whenever possible (Ferreira, 1993; Gee & Grosjean, 1983; Martin et al., 1971), then a sentence like "Every Saturday, he plays baseball" will exhibit boundary cues only after "Saturday", which receives focal stress.[9]

Another possible explanation for the apparent differences in phonological phrase assignment by children and adults concerns the placement of focal stress. The sentences with pronoun subjects that were imitated by children in the Gerken (1991) study were modeled with focal stress on the verb. Vogel and Kenesei (1990) have proposed that a phonological phrase boundary must occur after a word receiving focal stress. This would result in children producing sentences with pronoun subjects as two phonological phrases with a phonological phrase boundary after the verb as shown in (3b). In contrast, there is no reason to

[9]We thank an anonymous reviewer for this suggestion.

believe that our talker read the pronoun NP sentences with focal stress on the verb, thereby allowing her to produce these sentences as a single phonological phrase.

To summarize, the data from Experiment 1 suggest that infants are able to detect prosodic boundary cues in sentences with lexical NP subjects, in which prosody reflects major syntactic constituency, but not in simple declarative sentences with pronoun subjects, in which prosody fails to reflect syntax. How do our data bear on the prosodic bootstrapping hypothesis? The data suggest that when prosody boundary cues reflect syntactic constituency, these cues are readily available to young learners. Such cases are the bread and butter of prosodic bootstrapping. In contrast, for at least one case in which prosodic phonology predicts that prosody will not reflect syntactic constituency, learners do not appear to detect this constituency. If learners cannot employ prosodic cues to discover the syntactic constituency of declarative sentences with pronoun subjects, how do they come to realize that pronoun subjects, like their lexical NP counterparts, are separate syntactic constituents? One possibility is that prosodic information about the syntactic constituency of pronoun subjects is available from other sentence types. One such sentence type might be yes–no questions such as the one in (5).

(5) Did HE / kiss the dog?

In such sentences, it is possible to give the pronoun subject focal stress without producing a contrastive meaning, while this is not possible in declarative sentences with pronoun subjects (see footnote 2).[10] As noted earlier, Vogel and Kenesei (1990) have proposed that a phonological phrase boundary occurs after any word that receives focal stress, resulting in the prosodic analysis indicated by the slash in (5). If yes–no questions exhibiting the prosodic analysis shown in (5) are sufficiently frequent in caregiver input, they might provide infants with a way to infer that a pronoun subject comprises a constituent separate from the verb. This is because the pronoun is no longer in the same phonological phrase as the verb, and because receiving focal stress may give the pronoun potential status as an independent linguistic unit. An examination of the spontaneous caregiver speech used in the study by Jusczyk et al. (1992) revealed that 27% of their samples consisted of yes–no questions with pronoun subjects. We do not know whether these sentences were produced with focal stress on the verb. However, the fact that the sentences were so frequent in the Jusczyk et al. materials, and the fact that infants in that study showed such a strong listening bias for pauses after the subject, suggest that their talker indeed produced many yes–no

[10] We thank an anonymous reviewer for alerting us to the focal stress analysis of this sentence.

questions with boundary cues after the subject. We examined this issue in Experiment 2.

EXPERIMENT 2

In Experiment 2, we again used the pause insertion technique to determine if infants detect a prosodic boundary between the subject and verb of yes–no questions such as the one in (5). We predicted that, if they do, they should exhibit longer listening times for sentences in which a pause is inserted between the subject and verb than for sentences in which a pause is inserted after the verb. If infants are able to detect these prosodic boundary cues, then they have potentially useful information for identifying pronominal subjects as syntactic constituents.

Method

Subjects

Twenty-four American infants of approximately 9 months of age were tested. The infants had an average age of 40 weeks (range: 37 weeks, 1 day to 43 weeks, 2 days). Three additional infants were tested but excluded for the following reasons: crying ($n = 2$), and failure to complete the full set of test trials ($n = 1$).

Stimuli

The pronoun NP sentences from the previous experiment were altered to create yes–no questions. Examples of the after subject and after verb versions of a typical pronoun question sample are shown in Table 3. The same female talker who recorded the samples used in Experiment 1 recorded the new samples. As in the previous experiment, the new samples were digitized and pauses were inserted either after the subject (after subject versions) or after the verb (after verb versions).

Apparatus, design and procedure

With the exception that infants were tested on the pronoun question samples, all aspects of the experiment were the same as in Experiment 1.

Table 3. *Sample of pauses inserted after subjects and after verbs in Experiment 2*

Pause after subject
The next day was Sunday again. Did he / eat one nice green leaf, and after that did he / feel much better? Did he / feel hungry anymore?
Pause after verb
The next day was Sunday again. Did he eat / one nice green leaf, and after that did he feel / much better? Did he feel / hungry anymore?

Reliability data on the observer were taken from the video recordings. Videotapes of 18 of the 24 subjects tested were available for reliability coding. Reliability checks were performed in the same manner as described in the previous experiment. A Pearson product–moment correlation was performed on the difference in looking times for the before and after verb samples for each infant by the "live" and "videotape" observers (r_{17} = .986). It indicated very high agreement across these observations. We also examined possible discrepancies between the live and videotape looking time judgments by inspecting differences in recorded times on a trial-by-trial basis. On 159 trials (70%), the discrepancy between the recorded time of the live observer and a different observer viewing the videotape was less than 0.5 s. On 27 trials (12%), there was a discrepancy of 1 s or more. For these discrepancies, we examined the pattern of differences between the two observers to determine whether there was any systematic tendency on the part of the live observer to overestimate or underestimate the looking times on the before verb or after verb trials. There was no systematic difference between observers across these trials ($t_{25} = 0.91$, $p > .35$). Finally, the pattern of significant results reported for the live and videotape observers in Experiment 2 is identical.

As in the previous experiment, we also tested a group of 8 adult normal-hearing listeners to determine whether it was possible for the experimenter to discriminate the stimuli despite the presence of the music masking noise. The same test procedures were followed as in the previous experiment. Performance was nearly perfect on the discrimination task when judging the samples without any masking noise present (one listener again missed the first item but answered all others correctly). As was found in Experiment 1, there was no evidence that they significantly discriminated the two types of samples when masking music was present ($t_7 = 0.33$, $p > .70$). Across all listeners, there were 46 correct responses and 50 incorrect responses. The results of this experiment strongly suggest that the observer in the infant experiment could not discriminate the two versions of the stimuli over the masking music, thereby making it highly unlikely that this individual could unconsciously influence the results.

Results

The amount of time that each infant oriented to the loudspeaker on each trial was recorded. The average looking time was 8.71 s (SD = 3.53 s) for the after subject versions and 7.16 s (SD = 3.34 s) for the after verb versions. Overall, 16 of the 24 infants had longer looking times for the after subject versions. Paired t-tests by subjects and by items verified that infants listened longer to yes–no questions with a pause inserted between the subject pronoun and verb than sentences with a pause inserted after the verb ($t_s(23) = 2.31$, $p < .05$; $t_i(5) = 2.79$, $p < .05$).

In order to investigate which potential cues to phrase boundaries infants might be responding to, we conducted the same acoustic analyses for pitch and durational changes as in Experiment 1. We again found that infants listened longer to the condition in which significant changes occurred in both pitch and syllable duration. In particular, there was a significant rise in pitch (107 Hz; $t_{20} = 3.51$, $p < .005$) and a significant increase in duration (52 ms; $t_{20} = 2.20$, $p < .05$) at the boundary between the subject and verb phrase. The acoustic analyses of the boundary between the verb and complement showed a significant pitch drop (121 Hz; $t_{20} = 4.63$, $p < .001$) but no accompanying changes in duration ($t_{20} = 1.41$, n.s.).

We have postulated that talkers are likely to give focal stress to the subject in yes–no questions. If our talker gave focal stress to the pronoun subject in Experiment 2 but not in Experiment 1, we should expect that a comparison of the acoustic properties of the pronouns in the two experiments will indicate greater pitch change and longer duration for the pronoun in Experiment 2. This was indeed the case. The pitch of the pronouns in the yes–no question contexts in Experiment 2 proved to be significantly higher (61 Hz) than the same pronouns in declarative contexts in Experiment 1 ($t_{20} = 3.63$, $p < .02$, one-tailed). The duration of pronouns in yes–no questions was marginally longer (32 ms) than in declarative sentences ($t_{20} = 1.71$, $p = .052$, one-tailed). Thus, there are some indications of distinct pitch and durational changes of the same lexical item depending on the discourse context in which it occurs (see Grosjean & Gee, 1987, for further discussion), and these changes are consistent with the predictions of prosodic phonology.

Discussion

Infants listened longer to sentences with pauses inserted between the subject NP and verb than to sentences with pauses inserted after the verb. This finding is consistent with the one from the study by Jusczyk and his colleagues, in which about a quarter of the spontaneous speech samples were of the type employed in

Experiment 2 (Jusczyk et al., 1992). As in the lexical NP condition of Experiment 1, we interpret infants' listening bias in the current study to indicate that they detected a configuration of prosodic boundary cues, such as vowel lengthening and fundamental frequency change prior to the subject NP and verb. Thus, when a pause was inserted in this location, it was presumably consistent with the presence of a phonological phrase boundary. In contrast, a pause inserted after the verb did not coincide with a phonological phrase boundary.

As in Experiment 1, infants exhibited a listening bias for the condition in which reliable changes in both pitch and duration were found in the acoustic analyses. Although we interpret the acoustic data with caution, the fact that the same relation between infant preference and concurrent acoustic changes appeared in both experiments is potentially important. Vaissiere (1983) notes that some speakers appear to mark prosodic boundaries with fundamental frequency change, while others rely more heavily on durational changes. Perhaps one characteristic of child-directed speech is that speakers produce both changes concurrently. As we saw in the comparison of duration data in Experiment 1 and the previous study by Jusczyk et al. (1992), the direction of change in boundary cues does not seem to matter as much as the existence of the change. In Experiment 2, there was a significant pitch increase prior to the boundary between the subject and verb, whereas in the lexical NP materials in Experiment 1 there was a significant decrease in pitch for the same boundary. Different fundamental frequency characteristics for statements and yes–no questions were reported by Lieberman (1967). Perhaps rapid changes in fundamental frequency, regardless of direction, cue infants (and adults) to the presence of a prosodic boundary (see General Discussion).

Infants' apparent ability to detect prosodic boundary cues after the pronoun subject, as well as the greater pitch and duration exhibited by the pronoun in Experiment 2 versus Experiment 1, suggest that yes–no sentences with pronoun subjects might provide learners with more evidence about the lexical and syntactic status of pronoun subjects than did the declarative sentences used in Experiment 1. We will continue with this notion in the General Discussion.

GENERAL DISCUSSION AND CONCLUSION

The data from Experiments 1 and 2 support the view that young language learners have available in the speech stream prosodic cues to syntactic structure. But the data also suggest that the syntactic structure of language is not worn on its prosodic sleeve. Rather, learners must engage in active inference in order to determine the correct syntactic structure from prosodic information. The learner's problem can be seen in two aspects of our findings. First, the acoustic analyses of our materials suggest that, although there are detectable acoustic changes to prosodic boundaries, not every acoustic change implies a boundary, and not every boundary is marked by an identical configuration of acoustic changes. That is, we found significant changes in pitch or syllable duration, but unless both types of changes occurred together infants appeared not to detect a boundary. We also found that infants appeared to detect a prosodic boundary when both pitch and duration decreased significantly (Experiment 1) and when pitch decreased and duration increased significantly (Experiment 2). Recall that Jusczyk et al. (1992) found that infants detected a boundary at the point where pitch decreased and duration increased. Thus, across three experiments, infants have shown sensitivity to boundaries at which three of the four possible relations between pitch and duration were exhibited. As we indicated earlier, making acoustic measurements on materials that were not designed specifically for this purpose is problematic, and therefore we must be extremely cautious in interpreting these measurements. However, it is important to note that our findings do not suggest a universal mapping between a particular configuration of acoustic cues and the existence of a linguistic boundary.[11]

The second aspect of our data that highlights the learner's problem is that not all sentences provide prosodic cues to syntactic structure. In declarative sentences with lexical subjects and in yes–no questions with pronoun subjects, major syntactic constituency is reflected by prosodic boundary cues, and infants appear to be highly sensitive to these cues. Thus in these cases, learners can read a great deal of syntactic structure directly from prosodic information. However, in declarative sentences with pronoun subjects, syntactic constituency is not reflected by prosodic boundary cues. This is evidenced by the fact that our listeners gave no indication of detecting any cues to syntactic boundaries at all for these types of sentences. Furthermore, declarative sentences with pronoun subjects are probably much more frequent in the input to language learners than some other cases in which prosody fails to cue syntactic constituency, such as sentences with strings of relative clauses.

A still worse situation for the learner might be another type of sentence with pronoun subjects, in which prosody may cue the wrong syntactic constituency (instead of cueing no constituency). Consider the sentences in (6).

(6a) Joe / kissed / the big dog.
(6b) He kissed / the big dog.

As noted earlier in the brief discussion of the theory of prosodic phonology, complex object NPs, such as those containing an adjective, cannot be incorporated into the phonological phrase containing the verb. Thus, in (6a) and (6b),

[11] We thank an anonymous reviewer for pointing out the difficulty raised by the lack of a one-to-one acoustic–linguistic mapping.

the object NP "the big dog" forms its own phonological phrase. In (6b), this results in the only prosodic boundary (marked by the slash) obscuring the syntactic constituency of the subject pronoun. Note that in this case, unlike in (3b), the prosodic boundary is obligatory, not optional. We are planning to use the pause insertion technique to determine if infants prefer a pause inserted between the verb and object NP, and therefore the wrong syntactic constituency, in sentences like (6b).

How do learners overcome exposure to missing or misleading prosodic cues to syntactic structure? Within the prosodic phonology framework, learners must discover the mapping relation between acoustically cued prosodic structures on the one hand and syntactic structures on the other. Perhaps learners at some stage in development are able to do this through the use of a combination of prosodic cues, articles, concord morphology, etc. (Morgan et al., 1987). Cross-sentence comparisons of prosodically cued linguistic structures provide another potential source of information about the prosody/syntax mapping. For example, declaratives with lexical NP subjects potentially cue the subject as a constituent, as do yes–no questions with pronoun subjects. Learners might infer from these two cases that a pronoun subject of a declarative sentence also forms a syntactic constituent, even if this information is not directly available from the prosody of the sentence itself.[12] A recent study suggests that 9-month-olds have a listening bias for just those sets of sentences that would allow them to make useful cross-sentence comparisons, namely those sentences in which the lexical items remain fixed while the syntactic structure changes (Gerken, in press-b; Jusczyk & Kemler Nelson, in press). Future research is needed to examine more precisely the nature of learners' structural representations in order to determine if and when they might begin to use cross-sentence information to extract the syntactic structure of their language.

The notion that learners somewhat older than the ones we studied are able to make inferences about the structure of one sentence type from information detected in another type is supported by data collected by Newport, Gleitman, and Gleitman (1977). These investigators conducted an extensive correlational study examining specific properties of the input of several mothers on the subsequent development of their 12- to 27-month-old children's linguistic abilities. They found that one of the only two significant positive correlations was between the frequency of yes–no questions in the input and children's subsequent use of auxiliaries. Newport and her colleagues proposed that moving auxiliaries to sentence-initial position and producing them in an uncontracted form made these elements more salient to learners in yes–no questions than in other contexts. In

[12]Of course, this assumes that the learner has some way of recognizing the occurrence of a particular pronoun across two or more sentences. Recent findings by Jusczyk and Aslin (1993) suggest that even 7-month-olds have some ability to recognize the same content words in different sentential contexts.

the current framework, it is possible that extracting the auxiliary from a clitic group containing the verb to a separate phonological phrase also adds to its salience. More important for the notion that cross-sentence comparisons might be useful for discovering syntactic structure, Newport et al. found that children whose mothers used more auxiliaries in yes–no questions also used more auxiliaries in other types of constructions. That is, learners may have first noted the existence of auxiliaries in questions, but once having done so, they were able to more rapidly discover the proper use of auxiliaries in other sentences as well. Similarly, children exposed to pronoun subjects as constituents separate from the verb in yes–no questions may begin to assign this syntactic analysis in declaratives as well.

A number of years ago, McNeill (1966) considered the prospect that "intonation . . . might be the first vehicle on which children arrive at the rudiments of syntax" (p. 53). At the time, he dismissed this possibility chiefly on the basis of a study by Lieberman (1965) which demonstrated that, using intonation alone, linguists perceived speech more in terms of its acoustic contours rather than prosodic contours that correlate with grammatical structures. McNeill viewed the language learner as being in a position comparable to the linguists. Although he may have gone too far in dismissing the role of prosody in discovering syntactic structures, the present findings support his general contention that the language learner requires additional information to induce the correct constituent structures from the input.

By delineating some of the limits of prosodic cues to provide information about phrase structures, the present study extends the findings from a similar study by Jusczyk et al. (1992). In particular, the main finding of the earlier study – that infants listen longer to samples which preserve the integrity of major phrasal units – was replicated with two new sets of materials (i.e., lexical NP subject sentences and pronoun question sentences). At the same time, the present results indicate that infants' listening preferences for segmentations of the input that preserve phrasal units may hold only for situations in which the prosodic and syntactic structures coincide. Thus, as Jusczyk and his colleagues suggested, prosodic packaging may provide the type of perceptual precategorization that allows an infant who is able to make use of such information to discover syntactically relevant units. It is probable that prosodic boundary cues work in concert with cross-sentence comparisons and with other distributional cues (Morgan et al., 1987) in informing listeners of the constituent structure of utterances.

In sum, we believe that infants' failure to detect prosodic boundary cues after the pronoun subject in Experiment 1 suggests a more active role for the young language learner than is suggested by many prosodic bootstrapping accounts. Although our view is certainly not at odds with the general flavor of the prosodic bootstrapping hypothesis, we do take issue with the sometimes expressed, often implied notion that learners can directly read syntactic structure from prosodic boundary cues (e.g., Lederer & Kelly, 1991). What we have attempted to show here is that prosodic boundary cues, although a potentially rich and reliable sourcc of information to the child, do not mirror syntactic structure. Therefore, researchers must begin to consider mechanisms by which learners might employ prosodic boundary cues and other types of information to actively construct syntactic representations.

References

Allen, G., & Hawkins, S. (1980). Phonological rhythm: definition and development. In G.H. Yeni-Komshian, J.F. Kavanagh, & G.A. Ferguson (Eds.), *Child phonology: Vol. I: production* (pp. 227-256). New York: Academic Press.

Bahrick, L.E., & Pickens, J.N. (1988). Classification of bimodal English and Spanish language passages by infants. *Infant Behavior and Development*, *11*, 277-296.

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

Bernstein-Ratner, N. (1986). Durational cues which mark clause boundaries in mother-child speech. *Journal of Phonetics*, *14*, 303-309.

Broen, P. (1972). The verbal environment of the language learning child. *ASHA Monograph*, *17*.

Brown, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Cooper, W. (1975). *Syntactic control of speech timing*. Unpublished PhD dissertation, MIT.

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

Crystal, D. (1969). *Prosodic systems and intonation in English*. London: Cambridge University Press.

de Boysson-Bardies, B. (1993). Ontogeny of language-specific syllabic productions. In B. de Boysson-Bardies, S. de Schoen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Development neurocognition: speech and face processing in the first year of life* (pp. 353-364). Dordrecht: Kluwer.

de Boysson-Bardies, B., Sagart, L., & Durand, C. (1984). Discernible differences in the babbling of infants according to target language. *Journal of Child Language*, *11*, 1-15.

Demuth, K. (1992, October). *Competence of performance? What phonology shows about children's emerging syntax*. Paper presented at the Boston University Conference on Language Development.

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

Fernald, A., & Kuhl, P.K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*, 279-293.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mother's speech to newborns. *Developmental Psychology*, *20*, 104-114.

Ferreira, F. (1993). Creation of prosodic structure during sentence production. *Psychological Review*, *100*, 233-253.

Ferreira, F., & Morrison, F. (in press). Children's knowledge of syntactic constituents: effects of ages and schooling. *Developmental Psychology*.

Fisher, C.L. (1991). Prosodic cues to phrase structure in infant directed speech. *Papers and Reports on Child Language Development*, *30*.

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

Gee, J., & Grosjean, F. (1983). Performance structures: a psycholinguistic and linguistic appraisal. *Cognitive Psychology*, *15*, 411-458.

Gerken, L.A. (in press-a). A metrical template account of children's weak syllable omissions. *Journal of Child Language*.

Gerken, L.A. (in press-b). Phonological and distributional cues to syntax acquisition. In J. Morgan & K. Demuth (Eds.), *Signal to syntax: bootstrapping from speech to grammar in early acquisition*. Cambridge, MA: MIT Press.

Gerken, L.A. (1991). The metrical basis for children's subjectless sentences. *Journal of Memory and Language*, *30*, 431–451.

Gerken, L.A. (1993). Young children's representation of prosodic phonology: evidence from English-speakers' weak syllable omissions. *Journal of Memory and Language*, *32*.

Gerken, L.A., Landau, B., & Remez, R.E. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology*, *27*, 204–216.

Gleitman, L., Gleitman, H., Landau, B., & Wanner, E. (1988). Where learning begins: initial representations for language learning. In F. Newmeyer (Ed.), *The Cambridge linguistic survey*. New York: Cambridge University Press.

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

Grosjean, F., & Gee, J.P. (1987). Prosodic structure and spoken word recognition. *Cognition*, *25*, 135–155.

Hayes, B. (1989). The prosodic hierarchy in meter. In P. Kiparsky & G. Youmans (Eds.), *Phonetics and phonology: rhythm and meter* (pp. 201–260). San Diego: Academic Press.

Hillinger, M., James, C., Zeil, D., & Prato, L. (1976). The influence of prescriptive and subjective phrase markers on retrieval latencies. *Bulletin of the Psychonomics Society*, *8*, 353–355.

Hirsh-Pasek, K., Kemler Nelson, D.G., Jusczyk, P.K., Wright Cassidy, K., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for prelinguistic infants. *Cognition*, *26*, 269–286.

Jackendoff, R. (1977). *X'-syntax: a study of phrase structure*. Cambridge, MA: MIT Press.

Jusczyk, P.W., & Ashin, R.N. (1993). Recognition of familiar patterns in fluent speech by $7\frac{1}{2}$-month-old infants. Manuscript submitted for publication.

Jusczyk, P.W., Friederici, A.D., Wessels, J., Svenkerud, V.Y., & Jusczyk, A.M. (1993). Infants' sensitivity to the sound pattern of native language words. *Journal of Memory and Language*, *32*, 402–420.

Jusczyk, P.W., Hirsh-Pasek, K., Kemler Nelson, D., Kennedy, L., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, *24*, 252–293.

Jusczyk, P.W., & Kemler Nelson, D. (in press). Syntactic units, prosody, and psychological reality during infancy. In J. Morgan & K. Demuth (Eds.), *Signal to syntax: bootstrapping from speech to grammar in early acquisition*. Cambridge, MA: MIT Press.

Kaisse, E. (1983). The syntax of auxiliary reduction in English. *Language*, *59*, 93–122.

Kemler Nelson, D., Hirsh-Pasek, K., Jusczyk, P.W., & Wright Cassidy, K. (1989). How prosodic cues in motherese might assist language learning. *Journal of Child Language*, *16*, 53–68.

Klatt, D. (1975). Vowel lengthening is syntactically determined in connected discourse. *Journal of Phonetics*, *3*, 129–140.

Klavans, J. (1985). The independence of syntax and phonology in cliticization. *Language*, *61*, 95–120.

Lederer, A., & Kelly, M.H. (1991). Prosodic correlations to the adjunct/complement distinction in motherese. *Papers and Reports on Child Language Development*, *30*.

Lehiste, I., Olive, J.P., & Streeter, L. (1976). The role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America*, *60*, 1199–1202.

Levelt, W. (1969). *The perception of syntactic structure*. Heymans Bulletin No. HB-69-31B. Groningen, The Netherlands: Groningen University.

Lieberman, P. (1965). On the acoustic basis of the perception of intonation by linguists. *Word*, *21*, 40–54.

Lieberman, P. (1967). *Intonation, perception and language*. Cambridge, MA: MIT Press.

Luce, P.A., & Charles-Luce, J. (1983, May). *Contextual effects on the consonant/vowel ratio in speech production*. Paper presented at the 105th meeting of the Acoustical Society of America, Cincinnati.

Martin, J.E., Kolodziej, B., & Genay, J. (1971). Segmentation of sentences into phonological phrases as a function of constituent length. *Journal of Verbal Learning and Verbal Behavior*, *10*, 226–233.

Martin, J.G. (1970). Toward an analysis of subjective phrase structure. *Psychological Bulletin*, *74*, 153–166.

Mazuka, R. (1991). How can a grammatical parameter be set before the first word? *Papers and Reports on Child Language Development*, *30*.

McNeill, D. (1966). *Developmental psycholinguistics*. Cambridge, MA: MIT Press.

Mehler, J., Jusczyk, P.W., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 143–178.

Morgan, J. (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 learning. *Cognitive Psychology*, *22*, 498–550.

Morgan, J., & Newport, E. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior*, *20*, 67–85.

Nakatani, L., & Dukes, K. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, *62*, 714–719.

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

Newport, E., Gleitman, H., & Gleitman, L. (1977). Mother, I'd rather do it myself. In C. Snow & C. Ferguson (Eds.), *Talking to children* (pp. 109–147). Cambridge, UK: Cambridge University Press.

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

Price, P.J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, *90*, 2956–2970.

Pye, C. (1983). Mayan telegraphese. *Language*, *59*, 583–607.

Read, C., & Schreiber, P. (1982). Why short subjects are harder to find than longer ones. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: the state of the art* (pp. 78–101). Cambridge, UK: Cambridge University Press.

Scott, D.R. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America*, *71*, 996–1007.

Scott, D.R., & Cutler, A. (1984). Segmental phonology and the perception of syntactic structure. *Journal of Verbal Learning and Verbal Behavior*, *23*, 450–466.

Selkirk, E. (1972). *The phrase phonology of English and French*. Doctoral dissertation, MIT. Published 1980 by Garland publishing, New York.

Selkirk, E. (1981). On the nature of phonological representation. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech* (pp. 379–388). Amsterdam: North-Holland.

Selkirk, E. (1984). *Phonology and syntax*. Cambridge, MA: MIT Press.

Selkirk, E., & Shen, T. (1990). Tone deletion in Shanghai Chinese. In S. Inkelas & D. Zecs (Eds.), *The phonology–syntax connection* (pp. 313–337). Stanford, CA: Chicago University Press.

Snow, K. (1972). Mother's speech to children learning language. *Child Development*, *43*, 549–565.

Streeter, L. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America*, *64*, 1582–1592.

Suci, G. (1967). The validity of pause as an index of units in language. *Journal of Verbal Learning and Verbal Behavior*, *6*, 26–32.

Vaissiere, J. (1983). Language-independent prosodic features. In A. Cutler & D.R. Ladd (Eds.), *Prosody models and measurements* (pp. 53–66). New York: Springer-Verlag.

Vogel, I., & Kanesei, I. (1990). Syntax and semantics in phonology. In S. Inkelas & D. Zecs, (Eds.), *The phonology–syntax connection* (pp. 339–363). Stanford, CA: Chicago University Press.

Wijnen, F., Krikhaar, E., & den Os, E. (in press). The (non)realization of unstressed elements in children's utterances: a rhythmic constraint? *Journal of Child Language*.

Zwicky, A.M., & Pullum, G.K. (1983). Cliticization vs inflection: English n't. *Language*, *59*, 502–513.