FROM SOUND TO SYNTAX:
THE PROSODIC BOOTSTRAPPING OF CLAUSES

by

Kara Hawthorne

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF LINGUISTICS

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College of

THE UNIVERSITY OF ARIZONA

2013

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation

prepared by Kara Hawthorne

entitled From Sound to Syntax: The Prosodic Bootstrapping of Clauses

and recommend that it be accepted as fulfilling the dissertation requirement for the

Degree of Doctor of Philosophy

____ Date: 11/13/2012
Dr. LouAnn Gerken

____ Date: 11/13/2012
Dr. Rebecca Gomez

____ Date: 11/13/2012
Dr. Heidi Harley

____ Date: 11/13/2012
Dr. Diane Ohala

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

____ Date: 11/13/2012
Dissertation Director: LouAnn Gerken

STATEMENT BY AUTHOR

SIGNED: Kara Hawthorne

ACKNOWLEDGEMENTS

While this dissertation will go forth into the world bearing only my name, it is truly a collaborative document. I would like to thank the following people for their advice and support.

* LouAnn Gerken, my adviser, for her patience, encouragement, and unfailingly excellent advice. Each time we met, I left her office inspired.
* My committee members – Rebecca Gomez, Heidi Harley, and Diane Ohala – for being in my corner throughout this whole process.
* Reiko Mazuka for welcoming me into her lab in Tokyo and helping me finish an entire infant experiment in just ten weeks.
* Cecile McKee for teaching me much of what I know about how to conduct responsible research and how to keep preschoolers from crying. Also, Adam Ussishkin for his helpful comments at the earliest stages of this project.
* Colin Dawson for patiently answering the same six questions about statistics over and over again. And for the tea.
* Greg Key and Alex Trueman for answering my truly embarrassing questions about syntax.
* Jessamyn Schertz for reading a dissertation about Japanese prosody *written in French* on my behalf. That's a good friend.
* Carolyn Quam, Andréa Davis, and Davi Vitela for useful discussion and emotional support in the forms of lab floor time, swimming dates, baked goods, and venting sessions.
* Sara Knight, Brianna McMillan, and Juliet Minton – I've been through a lot of lab managers in my time, but each provided invaluable practical help.
* Omar Aujani, Jeannette Bell, Lauren Clough, Erin Huff, and Kelly McGinn – the undergraduate research assistants who helped with everything from scheduling appointments to dealing with undergraduate participants to acoustic analysis of hundreds of sentences.
* Mihoko Hasegawa ,Yuri Hatano, and Mari Kanamura – the lab coordinators in Japan – for introducing me to tomato ramen and for easing every aspect of conducting an experiment in a foreign country.
* Ai Kanato, Yumiko Maeda, and Yuka Miyake, for facilitating the experiment in Japan and for their excellent puppeteering.
* The babies, parents, and adults in Tucson and Tokyo, who graciously donated their time to participate in these experiments.
* Finally, to the funding agencies: NSF grant 0950601 to LouAnn Gerken, RIKEN BSI funding to Reiko Mazuka, and NSF Graduate Research Fellowship Program and East Asian and Pacific Summer Institute Grants to me.

TABLE OF CONTENTS

TABLE OF CONTENTS – *Continued*

LIST OF FIGURES

Chapter 2

Chapter 3

Chapter 4

Chapter 5

LIST OF TABLES

Chapter 2

Chapter 3

Chapter 4

ABSTRACT

It has long been argued that prosodic cues may facilitate syntax acquisition (e.g., Morgan, 1986). Previous studies have shown that infants are sensitive to violations of typical correlations between clause-final prosodic cues (Hirsh-Pasek et al., 1987) and that prosody facilitates memory for strings of words (Soderstrom et al., 2005). This dissertation broaches the question of whether children can use this information in syntax acquisition by asking if learners can use the prosodic correlates of clauses to locate syntactic constituents. One property of certain syntactic constituents in natural languages is that they can move, so learning of constituency was inferred if participants treated prosodically-grouped words as cohesive, moveable chunks. In Experiment 1, 19-month-olds were familiarized with sentences from an artificial grammar with either 1-clause or 2-clause prosody. The infants from the 2-clause group later recognized the prosodically-marked clauses when they had moved to a new position in the sentence and had a new acoustic contour. Adults in Experiment 2 showed similar learning, although their judgments also rely on recognition of perceptually-salient words at prosodic boundaries.

Subsequent experiments explored the mechanisms underlying this prosodic bootstrapping by testing Japanese-acquiring infants on English-based stimuli (Experiment 3) and English-acquiring infants on Japanese-based stimuli (Experiment 4). Infants were able to locate constituent-like groups of words with both native and non-native prosody, suggesting that the acoustic correlates of prosody are sufficiently robust across languages that they can be used in early syntax acquisition without extensive exposure to language-specific prosodic features. On the other hand, adults (Experiment 5) are less flexible, and

are only able to use prosody consistent with their native language, suggesting an age- or experience-related tuning of the prosodic perceptual mechanism. This dissertation supports prosody as an important cue that allows infants and young children to break into syntax even before they understand many words, and helps explain the rapid rate of syntax acquisition.

CHAPTER 1

INTRODUCTION TO THE RESEARCH QUESTIONS

1.1 Introduction

Human infants are born relatively helpless, yet they are capable of learning all the complexities of any human language in a short amount of time and without explicit instruction. This is a particularly amazing feat given the nature of the input that infants receive. An adult reading the words on this page can easily identify the boundaries of words because of the spaces that surround them. Similarly, a reader can use capital letters, commas, periods, and other punctuation marks to help identify the beginnings and ends of clauses and some other types of phrases. In the spoken language that infants hear, however, there are no spaces or pauses between words or between most syntactic constituents: sounds and words all run together. An infant must be able to pick words out of the speech stream before she can learn what those words mean. Similarly, she must learn to identify phrasal and clausal constituents before she can learn the hierarchical sentence structure that is crucial for sentence comprehension. This is a daunting task, yet infants can segment many words from the speech stream by 7.5 months (Jusczyk, Houston, & Newsome, 1999) and they are sensitive to the acoustic cues that occur at the boundaries of clausal constituents as early as 4.5 months (Jusczyk, 1989).

One important aspect of sentence segmentation and early syntax acquisition is learning to identify which words group together to form syntactic constituents. While human language manifests as a linear string of words, many syntactic phenomena that learners must acquire operate over hierarchically-organized variables – grammatical categories and syntactic constituents (note that linguistic theory posits both syntactic and prosodic constituents; unless otherwise noted, "constituent" will be used to refer to syntactic constituents throughout this dissertation). A constituent is a group of words that functions as a cohesive unit (see Carnie, 2002). Syntactic constituents form the building blocks upon which natural language grammar is organized.

Constituency is important even at the early stages of syntax acquisition. Very young language learners need a rudimentary concept of constituency to understand the subject-predicate distinction. It has been argued that, cross-linguistically, sentences require at least a subject noun phrase (NP) and a predicate, which is usually a verb phrase (VP) (c.f., Marantz, 1984)[1]. In example (1), the subject NP is "The little girl," and the rest of the sentence is the predicate VP. This subject-predicate distinction is important for low-level sentence comprehension of who did what to whom. Although it is possible that a child learning English could achieve the correct interpretation for the majority of English sentences by hypothesizing that the first NP is the agent of the sentence, this strategy would not work for passive sentences or for learners of languages with more flexible word order.

(1) $[_{NP}$ The little girl] $[_{VP}$ kissed all the frogs in the pond]

Once a learner can distinguish at least between subject and predicate constituents, she can begin to note the distributional properties and dependencies of the elements inside those constituents. An English-acquiring child, for example, will learn that

---

[1] Pro-drop languages, in which the subject can be omitted, present an apparent counterexample. However, pro-drop languages still make a distinction between the subject and predicate, such that the subject is the optional constituent.

determiners and adjectives precede nouns in NPs. She will also determine whether English phrases have a head-complement or complement-head structure. A learner acquiring a language with case marking will learn that constituents are the domain over which affixes conveying gender and number information must agree (c.f. Morgan, Meier, & Newport, 1987).

For a more advanced language learner, constituency is important for understanding the rules underlying the behavior of certain groups of words. For example, only constituents can be replaced by a pro-form (e.g., *it* for NPs), as demonstrated by example (2), or undergo movement operations, as shown in example (3). (Words that are bracketed together form a constituent, though not all constituents are shown; note that constituents can be embedded inside other constituents.)

(2) [The [monster [from Mars]]] eats purple people.

a) *It from Mars eats purple people.

b) It eats purple people.

(3) I will meet the monster [in [March [of [this year]]]].

a) *In March, I will meet the monster of this year

b) In March of this year, I will meet the monster.

"The monster" cannot be replaced by the pronoun "it" in sentence (2a) because it does not form a constituent on its own. Similarly, sentence (3a) is ungrammatical because "in March" does not form its own constituent, so it cannot undergo a movement operation. A child must be able to identify constituents before she can learn these sorts of syntactic operations.

Finally, constituency is important to the learner because of its role in determining sentence meaning. Syntactic constituents are organized hierarchically, and this organization influences how we interpret a linear string of words. For example, the nature of the monster and its diet in the 1958 song "Purple People Eater" by Sheb Wooley is ambiguous. The interpretation depends on which words are grouped together into constituents.

(4) a) One-eyed one-horned [flying [purple [people eater]]]

Monster: one-eyed, one-horned, flying, purple

Eats: people

b) One-eyed one-horned [flying [[purple people] eater]]

Monster: one-eyed, one-horned, flying

Eats: purple people

c) One-eyed one-horned [[[flying [purple people]] eater]

Monster: one-eyed, one-horned

Eats: flying purple people

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. One aspect of the input that has received considerable attention as a possible source of information about syntactic structures is prosody, the rhythmic and melodic aspects of speech.

The prosodic bootstrapping hypothesis (Gleitman & Wanner, 1982; Morgan, 1986; Peters, 1983) proposes that prosodic information can help with the task of identifying word, phrase, and clause boundaries and possibly even for inferring constituency and hierarchical syntactic structure. In this dissertation, I focus specifically on prosodic information as a tool for the learning of syntactic constituency. As noted above, constituents are the building blocks upon which many critical syntactic dependencies hold and over which syntactic operations operate.

Previous formulations of the prosodic bootstrapping hypothesis are introduced in Section 1.2 and a discussion of the mechanism underlying prosodic bootstrapping is presented in Section 1.3. This chapter concludes with a overview of the dissertation in Section 1.4.

1.2 Prosodic Bootstrapping Hypotheses

According to the prosodic bootstrapping hypothesis, infants make use of correlations that exist between prosodic structure and the boundaries of linguistic units like words, phrases, and clauses to locate the boundaries of those units and, in some versions of the hypothesis, to locate constituents and even to identify the hierarchical relationships between those constituents. Also known as phonological bootstrapping or bootstrapping from the signal, prosodic bootstrapping is a compelling 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 (Blake & de Boysson-Bardies, 1992; Halle, de Boysson-Bardies, & Vihman, 1991; Mampe, Friederici, Christophe, & Wermke, 2009).

Indeed, there is a large body of evidence in support of the word segmentation component of the prosodic bootstrapping hypothesis. Nine-month-old English-acquiring infants listen longer to words that conform to the trochaic strong-weak stress pattern that predominates in English (Jusczyk, Cutler, & Redanz, 1993) and 7.5-month-olds can segment trochaic but not iambic words from fluent speech (Jusczyk et al., 1999). Johnson and Seidl (2009) pitted stress cues against statistical cues and found that 11-month-olds rely more heavily on the prosodic cues to word boundaries, while others have found that larger prosodic groupings constrain the domain over which learners look for words (Gout, Christophe, & Morgan, 2004; Shukla, Nespor, & Mehler, 2007).

The idea that prosody might also be a useful cue in syntax acquisition emerged in the 1980s from the observations of several different researchers. Morgan and Newport (1981) familiarized adults with an artificial grammar in which nonsense words corresponded to unfamiliar shapes. Word/shape sequences were constructed based on a grammar that could be interpreted either as finite-state or hierarchically-organized. Participants who were provided with only distributional cues or with distributional cues plus arbitrary visual groupings of the unfamiliar shapes learned the simper, finite-state grammar. On the other hand, participants who were provided with visual groupings that correlated with the hierarchically-organized constituents learned the more complex, hierarchical grammar. While the groupings were visual in their experiment, Morgan and Newport proposed that intonational groupings may be one natural language means for locating constituents in the absence of or in conjunction with semantic and morphological information.

In a foundational article, Gleitman and Wanner (1982) note that work in prosody has shown speech-timing effects, such as pausing and segmental lengthening, at syntactically-important sites like major constituent boundaries and locations where deletion is typically thought to occur. They argue that evidence from Read and Schreiber (1982), in which 7-year-olds were successful at locating major syntactic constituents, as long as those constituents occurred with the appropriate prosodic cues, hints that prosody may be useful at even earlier stages of syntax acquisition.

A few years later, Peters (1983) and Morgan (1986, 1996) fleshed out this hypothesis in more detail. Peters argues that pauses, suprasegmentals, intonational contours, and rhythm can all help children group the speech stream into units, whether those units be words, word-like strings (such as "thank you"), or syntactic constituents. The larger initial units can then, with additional experience, be broken into smaller units by additional use of intonation, rhythm, pauses, and stress.

Morgan (1986) argues for a more ambitious role for prosody in syntax learning. It has often been assumed that simple distributional analysis of the linearly-organized input string can help the learner identify syntactic constituents. He provides the following examples (from Morgan, 1986, p. 5):

(5) a) The little boy threw a rock at the dragon.

b) A doughty warrior threw a rock at the dragon.

c) He threw a rock at the dragon.

d) Yesterday in a fit of pique someone threw a rock at the dragon.

Based on (5a-c), a learner could infer that any string of words before "threw a rock at the dragon" is a unit that serves a particular grammatical and/or semantic function. Namely, it will be a subject NP. However, this gives the rather misleading conclusion that all strings before predicate VPs are subject NPs. This will clearly fail in the case of (5d). Morgan proposes that if a learner perceives the input not just as a string of words, but as a *bracketed* string of words, she will have an easier time reaching the correct final syntactic analysis, even if the brackets lack part-of-speech labels. Other sources of bracketing information, such as grammatical morphemes (which typically occur at constituent boundaries), require prior learning before they become useful cues. Prosody is a particularly useful source of bracketing information early on because it is readily perceivable.

Morgan's argument draws from earlier work by Levy and Joshi (1978, 1982), who demonstrate that the presence of nondistributional bracketing info enhances the ability of a finite-state automaton to learn a context-free grammar. Morgan further argues that the unlabeled bracketing provided by prosody reduces the need for multiple levels of embedding in the input. Instead of the child requiring two levels of embedding (degree-2) in the input to learn the transformations of her native language (as argued, for example, by Wexler & Culicover (1980)), learners who use prosody for bracketing only need one level of embedding (degree-1). Since degree-2 input is relatively rare in children's input, Morgan argues that a theory that can account for syntax acquisition with degree-1 input is preferable, and that the bracketing information provided by prosody may even been necessary for successful syntax acquisition.

While the proposals mentioned above focus on prosody as a cue to grouping or bracketing, there are other roles that prosody could play. For example, it has been argued that prosody could play a role in setting syntactic parameters. Mazuka (1996) contends that the Branching Direction parameter, which determines the direction in which embedded clauses occur in relation to the main clause (Lust & Chien, 1984), could be set based on prosodic information. Most English sentences, for example, are right-branching, while Japanese is a left-branching language, as shown in examples (6) and (7) from Mazuka (1996, p. 315).

(6) [I met the teacher [who called the student [who had an accident].

(7) [[[Ziko-ni atta] gakussee-ni denwasita] sensee -ni atta].
an-accident had student -DAT called teacher-DAT met

Mazuka measured prosodic cues in a small sample of elicited American English speech, and found that left-branching sentences had longer pauses at the clause boundary and a larger rise in fundamental frequency at the clause boundary compared to right-branching sentences. By making clever use of this prosodic information, she argues, infants may be able to set a branching direction parameter even before they have learned their first words.

More recently, Nespor and colleagues (2008) put forth a similar proposal, arguing that infants can use the iambic-trochaic law of perception to determine whether a given PhP is head-initial or head-final (within or across languages). The iambic-trochaic law states that sequences of alternating short and long units tend to be perceived as iambs, while sequences of alternating loud and soft units are perceived as trochees (although subsequent evidence suggests that this is not an innate property of our perceptual systems: see Bion, Benavides-Varela, & Nespor (2011); Yoshida et al. (2010)).

Therefore, the acoustic correlates of prominence in a given PhP will indicate whether the PhP boundary precedes or follows the prominent syllable. If the primary acoustic correlates of prominence in particular PhPs are increased intensity and pitch, for example, the learner will recognize that the PhP is trochaic and that the PhP boundary should go before the prominent syllable. Since PhPs are tightly constrained by syntax, and since complements are typically more prominent than heads, learners can infer that a trochaic PhP has a head-final structure, thereby bootstrapping from prosodic information to syntactic information about head direction.

While these proposals certainly merit further investigation, in this dissertation I focus on prosody as a cue to grouping and bracketing, delving more deeply into the question of how far prosody can take the young language learner in her acquisition of hierarchically-organized constituents. Three possibilities are outlined below; note that evidence in favor of one possibility does not constitute evidence against the others.

First, and at minimum, prosody may be used to segment the speech stream into substrings, which are then stored in memory and/or subjected to further analysis (Peters, 1983; Gerken, 1996). For example, large prosodic groupings, such as IPs and PhPs, have been shown to influence memory for strings of words (e.g., Soderstrom, Kemler Nelson, & Jusczyk, 2005) and to constrain the domain over which learners track transitional probabilities (e.g., Gout et al., 2004). However, under this proposal, the prosodically-extracted substrings themselves are not treated as syntactically-meaningful constituents by the learner. I will call this the *grouping* version of prosodic bootstrapping.

A second, intermediate possibility, is that infants use prosody to segment the speech stream into substrings that they recognize and treat as syntactic constituents. Instead of just breaking speech into smaller chunks, the learner recognizes that certain prosodic groupings are syntactically meaningful and that they can undergo transformations. I will refer to this as the *constituents* version of prosodic bootstrapping. The experiments presented in Chapter 3 test the constituents version of prosodic bootstrapping by asking whether learners treat prosodically-grouped words as cohesive, moveable syntactic constituents.

Finally, the strongest and most controversial proposal is that prosody can help the learner bracket the speech stream into hierarchically-organized constituents that comprise at least part of a skeletal phrase structure tree (Morgan, 1996; see Gerken, 1996 for further discussion). As discussed above, this would significantly ameliorate the problems inherent to syntax acquisition, but there is little behavioral evidence to support a claim this strong. I will refer to this as the *phrase structure* version of prosodic bootstrapping. Note that, even in its strongest form, prosodic bootstrapping proposals are neutral about whether children use prosody to posit syntactic structures or if they are instead correlating innate structures with prosodically-grouped words.

1.3 Mechanisms Underlying Prosodic Bootstrapping

An important and overlooked question at the root of all variations of the prosodic bootstrapping hypothesis is at what level prosody is useful to the learner. What is perceived by an adult native speaker as prosodic information about intonation and stress may be interpreted by a naïve listener merely as acoustic information about the pitch, duration, and amplitude of sounds. While pitch, duration, and amplitude are perceivable to most anyone with a mammalian auditory system, the way that these cues interact with each other to manifest as question intonation, pitch accents, or contrastive stress varies based on the language in question and therefore must be learned.

I propose two contrasting hypotheses about the mechanisms underlying prosodic bootstrapping effects. First, the *general acoustic* hypothesis asserts that prosodic bootstrapping relies on the acoustic salience of prosodic cues. Human newborns (Nazzi et al., 1998), cotton-top tamarins (Ramus, Hauser, Miller, Morris, & Mehler, 2000) and rats (Toro, Trobalon, Sebastian-Galles, 2003) can all discriminate sentences from languages that differ along the prosodic dimension of linguistic rhythm, suggesting that the cues underlying perception of linguistic rhythm are not inherently linguistic in nature. Although experimental studies addressing the general acoustic nature of prosody have, to date, largely focused on linguistic rhythm, it is probable that other acoustic correlates of prosody are also perceivable and discriminable without reference to linguistic structures. If the general acoustic hypothesis is correct, experience with the target language should not be required for prosodic bootstrapping to occur.

While it is true that at least certain manifestations of prosody are acoustically salient to mammals, it is also true that adult speakers interpret prosodic information as linguistically meaningful. For example, an adult native speaker of English interprets rising pitch at the end of a sentence as question intonation. This knowledge is not present at birth, but is acquired by all typically-developing native speakers of English. If it is the case that prosody's role in learning is based on the learner's acquired knowledge of her native language, experience with the target language would be an important prerequisite to the prosodic bootstrapping of syntax. This is the *language-specific* hypothesis. The general acoustic and language-specific hypotheses are tested in the experiments presented in Chapter 4.

1.4 Structure of the Dissertation

This dissertation is organized as follows. In Chapter 2, I review the prosodic hierarchy, then discuss previous studies testing the prerequisites of the prosodic bootstrapping hypothesis, with a section focusing on non-prosodic cues that may also facilitate syntax acquisition. Chapter 2 concludes with a discussion of the role of language-specific experience in prosodic bootstrapping.

Chapters 3 and 4 present experiments addressing two interrelated aspects of the prosodic bootstrapping hypothesis. In Chapter 3, I ask whether infant and adult learners can use prosody to locate clause-like units in an artificial grammar. The experiments in this chapter probe the intermediate, constituents version of the prosodic bootstrapping hypothesis by testing learners on their recognition of prosodically-grouped words when the words behave like rule-abiding syntactic constituents and move to a new position in the utterance.

In Chapter 4, I present three experiments that ask whether learners can use non-native prosody for locating clauses. These experiments investigate the mechanism underlying prosodic bootstrapping: if learners can use non-native prosody to locate clause-like units, it would suggest that prosodic cues are sufficiently robust that they can be used for syntactic learning without prior specific knowledge of the correlations between syntax and prosody in the target language, as suggested by the general acoustic hypothesis. Finally, in Chapter 5, I discuss the implications and possible extensions of the empirical findings of this dissertation and raise issues to be addressed in the future.

CHAPTER 2

AN OVERVIEW OF PREVIOUS RESEARCH ON PROSODIC BOOTSTRAPPING

In this chapter, I review prior work related to the prosodic bootstrapping hypothesis and situate the experimental work presented in Chapters 3 and 4 in context. The chapter begins with an introduction to prosody in Section 2.1. Section 2.2 reviews previous work on the prosodic bootstrapping proposal, including evidence for the grouping, constituents, and phrase structure hypotheses outlined in Chapter 1. Section 2.3 discusses several other cues that may be helpful in learning to parse speech into syntactic constituents. In Section 2.4, I cover evidence pertaining to the mechanism underlying prosodic bootstrapping by exploring the role of language-specific experience. Finally, Section 2.5 summarizes the relevance of previous work to the present experiments.

2.1 The Linguistics of Prosody

2.1.1 The Prosodic Hierarchy

Prosody can be defined as "(1) acoustic patterns of F0, duration, amplitude, spectral tilt, and segmental reduction, and their articulatory correlates, that can be best accounted for by reference to higher-level structures, and (2) the higher-level structures that best account for these patterns" (Shattuck-Hufnagel & Turk, 1996, p. 196). More simply, prosody encompasses the rhythm, melody, and intensity of speech, including phenomena like stress and intonation. Prosodic and certain phonological phenomena are constrained by prosodic units, which are hierarchically organized. For example, stress in English is assigned at the level of the foot; feet are composed of syllables and are organized into prosodic words.

The units in the prosodic hierarchy can be visualized as well-formed trees, similar to syntactic trees. Unlike syntactic trees, however, prosodic trees typically obey the Strict Layering Hypothesis: a member of one level of the hierarchy is composed exclusively of at least one member of the next level down in the hierarchy. For instance, a foot is composed of at least one syllable, and a syllable is composed of at least one mora (cf. Shattuck-Hufnagel & Turk, 1996). For a more detailed perspective on constraints on the prosodic hierarchy and examples of cases where the Strict Layering Hypothesis is violated, see Selkirk (1996).

There are many competing proposals about the exact breakdown of prosodic units in the prosodic hierarchy. The basic units of the hierarchy are summarized in Figure 2.1.

<u>The Prosodic Hierarchy</u>
Utterance (U)
Intonational Phrase (IP)
Phonological Phrase (PhP)
Clitic Group (CG)
Phonological Word (Pwd)
Foot (F)
Syllable (σ)
Mora (μ)

FIGURE 2.1 This version of the prosodic hierarchy is similar to that expressed by Nespor & Vogel (2007) and Hayes (1989), with the addition of the mora (see McCawley, 1968).

The prosodic units that are relevant for describing multi-word utterances are the clitic group, phonological phrase, intonational phrase, and utterance. The clitic group

(CG) is a single content word – typically a noun, verb, or adjective – along with any adjacent function words that occur within the same syntactic constituent (Hayes, 1989).

(1) $[_{\text{CG}}$ The overripe] $[_{\text{CG}}$ and very] $[_{\text{CG}}$ brown] $[_{\text{CG}}$ bananas] $[_{\text{CG}}$ are rotting]

CGs group together to form phonological phrases (PhPs), which are tightly constrained by syntax. A PhP is comprised of one or more adjacent CGs within the same noun, verb, or adjective phrase (Nespor & Vogel, 2007). In English, a PhP is typically composed of the CG containing the head of the syntactic phrase, along with all CGs within that phrase that occur to the right of the head and, sometimes, one CG to the left of the head (Hayes, 1989). It is common for a non-pronominal subject NP and its predicate VP to be separated by a PhP boundary in English.

(2) $[_{\text{PhP}}$ The overripe and very brown bananas] $[_{\text{PhP}}$ are rotting]

PhPs, in turn, group together to form intonational phrases (IPs), which are the domain of a coherent intonational contour (Shattuck-Hufnagel & Turk, 1996). IP boundaries are also positions where a speaker may choose to pause. Certain constructions, including parentheticals (3a) and nonrestrictive relative clauses (3b) have a strong tendency to form their own IP. IP boundaries typically co-occur with major syntactic constituent boundaries, particularly the boundaries of clauses. However, there are exceptions, as in (3c), where an IP interrupts the VP.

(3) $[_{\text{IP}}$ The overripe and very brown bananas are rotting]

a) $[_{\text{IP}}$ Jessamyn] $[_{\text{IP}}$ who is a little crazy] $[_{\text{IP}}$ likes to eat unripe bananas]

b) $[_{\text{IP}}$ According to Greg] $[_{\text{IP}}$ bananas are better when they are overripe]

c) $[_{\text{IP}}$ Chelsea fell] $[_{\text{IP}}$ as you probably heard] $[_{\text{IP}}$ into a train car full of bananas]

Finally, the utterance (U) is the largest prosodic domain. Utterances are built from one or more IPs, and typically correspond to the top node of a syntactic tree (Nespor & Vogel, 2007).

(4) [$_U$ Chelsea fell, as you probably heard, into a train car full of bananas]

2.1.2 Prosody and Syntax

The link between prosodic and syntactic structure has been a matter of discussion for many years (e.g., Chomsky & Halle, 1968). It has often been noted that certain syntactic ambiguities can be disambiguated prosodically when spoken aloud, as demonstrated by the sentence from Hayes (1989), "On Tuesdays, he gives the Chinese dishes." Depending on its syntactic structure, this sentence can be interpreted with *Chinese* as modifying *people* (Figure 2.2) or *dishes* (Figure 2.3). This syntactic difference is reflected prosodically.

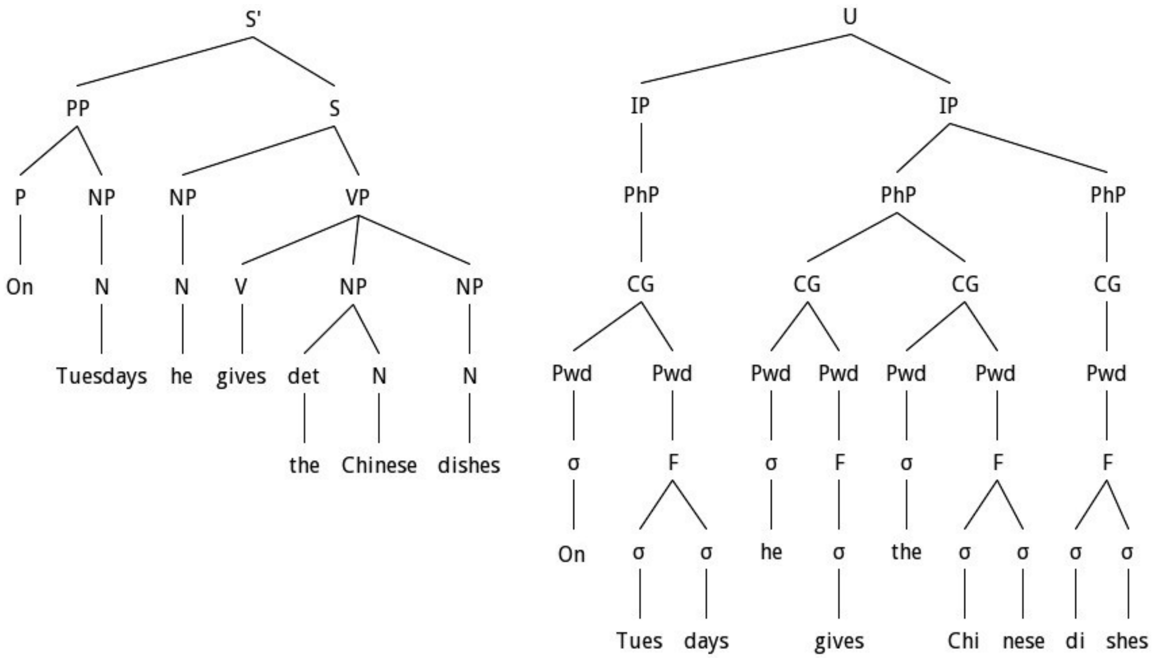

FIGURE 2.2 Syntactic and prosodic trees for the "*On Tuesdays, he gives dishes to the Chinese*" interpretation. Example adapted from Hayes, 1989, pp. 202-203. Trees drawn using RsyntaxTree (Hasebe, 2012).

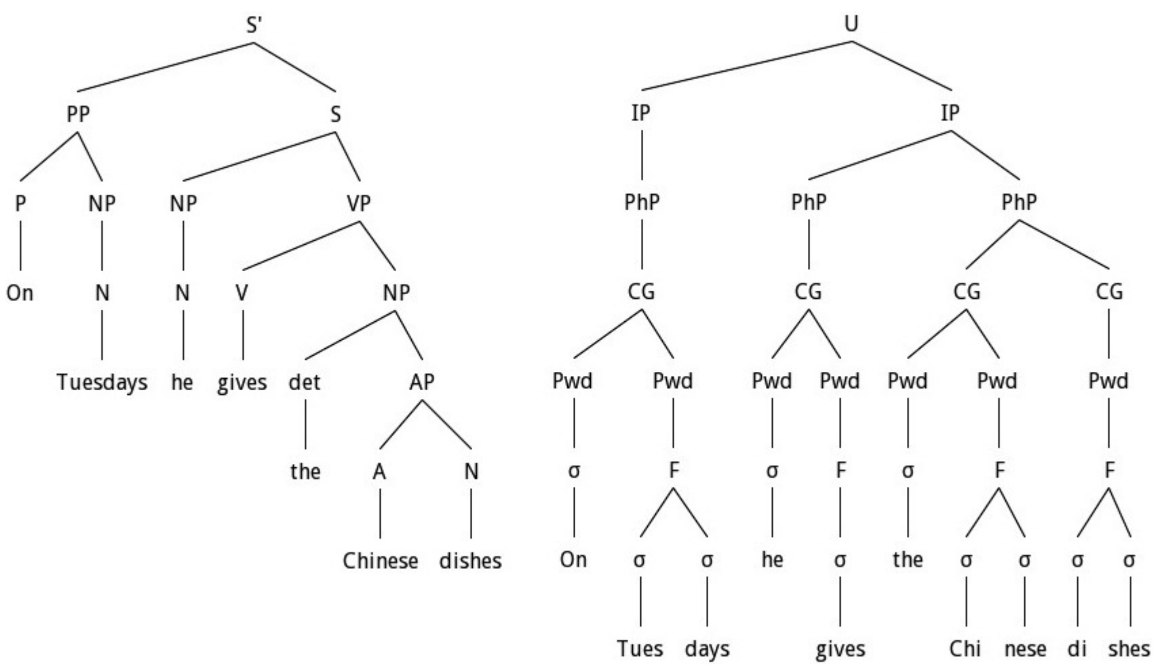

FIGURE 2.3 Syntactic and prosodic trees for the "*On Tuesdays, he gives Chinese dishes to people*" interpretation. Example adapted from Hayes, 1989, pp. 202-203. Trees drawn using RsyntaxTree (Hasebe, 2012).

Since prosody can disambiguate some kinds of syntactic ambiguities, syntactic structure clearly interacts with prosodic structure to a certain extent. In fact, the definitions for CG, PhP, IP, and U above all make reference to syntactic structures. However, prosodic structure is also at least partially independent from syntax, and syntactic and prosodic boundaries are often mismatched, as in the sentence "This is the cat that caught the rat that stole the cheese" (Chomsky & Halle 1968). Example (5a) shows the syntactic parsing and (5b) the prosodic groupings.

(5) a) [This is [the cat [that caught [the rat [that stole [the cheese]]]]

b) [This is the cat] [that caught the rat] [that stole the cheese]

There are many non-syntactic factors that influence prosodic choices. For example, speech rate and utterance length influence how a speaker breaks an utterance into IPs. In rapid speech, a speaker may pronounce (6) as a single IP, while a speaker who is talking more slowly will likely break it into several IPs with intervening pauses and/or pitch resets.

(6) a) [$_{IP}$ Alex works on a big beautiful and productive banana plantation]

b) [$_{IP}$ Alex works on a big] [$_{IP}$ beautiful] [$_{IP}$ and productive] [$_{IP}$ banana plantation]

Focal structure also influences a speaker's prosodic choices. Sentence (7), for instance, can be continued in two different ways, depending on the location of contrastive stress. Note that the sentence is not syntactically ambiguous. The appropriate completion for this sentence depends on its prosody.

(7) He gave me orange dishes.

a) He gave me ORANGE dishes...

...but I was hoping for blue dishes.

#...but I was hoping for orange placemats.

b) He gave me orange DISHES...

...but I was hoping for orange placemats.

#...but I was hoping for blue dishes.

Speakers can even choose between different prosodic parsings of a given sentences for purely stylistic reasons. Shattuck-Huffnagel & Turk (1996, p. 201) provide the following examples from *Sesame Street*.

(8) a) [$_{IP}$ Sesame Street is brought to you] [$_{IP}$ by the Children's Television Workshop]

b) [$_{IP}$ Sesame Street is brought to you by] [$_{IP}$ the Children's Television Workshop]

Therefore, while prosody is constrained by syntax, it is also influenced by other factors. Given a particular syntactic structure, a speaker can make prosodic choices based on speech rate, focal structure, and stylistic preferences. It is instructive, then, to examine the reliability of prosodic cues at syntactic boundaries, as is done in Section 2.1.3, in order to determine whether prosody is sufficiently correlated with syntax to facilitate syntax acquisition.

2.1.3 Prosodic Correlates of Syntactic Boundaries

There is a large body of work discussing prosodic cues at the boundaries of major syntactic constituents. Clauses, cross-linguistically, can be identified based on pauses (e.g., Cruttenden, 1986), and clause boundaries are commonly marked by pitch resets and final-syllable lengthening (e.g., Beckman & Edwards, 1990; Beckman & Pierrehumbert, 1986a; Cooper & Paccia-Cooper, 1980). In infant-directed speech (IDS), or "motherese," these prosodic features are exaggerated. Pauses are longer and occur more reliably at clause boundaries in IDS (Broen, 1972; Morgan, 1986). IDS also has exaggerated final-syllable lengthening (Bernstein Ratner, 1986; Morgan, 1986) and greater overall pitch deviations (Garnica, 1977), as compared to adult-directed speech (ADS).

While most studies looking at prosody in IDS have collected data in a controlled laboratory setting, there are two important exceptions. Fernald and colleagues (1989) conducted a cross-linguistic comparison of prosodic exaggeration in IDS by recording and analyzing spontaneous IDS at participants' homes. They examined IDS from native speakers of French, Italian, German, Japanese, British English, and American English – five mothers and five fathers from each language. Parents of all six languages had a higher mean fundamental frequency (F0) and greater F0 variability when speaking to infants; they also used shorter utterances and longer pauses in IDS. However, there was an interesting cross-cultural difference: parents who spoke American English had significantly more extreme intonational exaggerations in IDS, even compared to British English-speaking parents. This suggests that, despite prosodic exaggeration in IDS across many languages, there is also cultural variability in natural IDS prosody.

More recently, Soderstrom Blossom, Foygel, and Morgan (2008) analyzed the prosody of two mothers speaking to their 9-month-old infants in a natural, at-home setting. Importantly, they considered both utterance-final and utterance-internal clause boundaries, and found that both types of clauses are reliably marked by pauses and final-syllable vowel lengthening. However, the peak pitch difference between pre- and post-boundary vowels was only significant for one of the two mothers. Thus, natural (American English) IDS contains reliable acoustic cues to clause boundaries, though some cues may be more consistent across speakers than others.

Phrases below the level of the clause are less reliably marked. In controlled American English laboratory speech, even relatively minor phrase boundaries are marked with a fall-rise pitch pattern across the boundary (Cooper & Sorensen, 1981). In more natural speech, sub-clausal phrase boundaries are prosodically marked only if certain conditions are met. For example, the subject-predicate boundary is typically marked only if the subject is not an unstressed pronoun, as seen in example (9). However, in natural speech to 9-month-olds, the two mothers studied by Soderstrom and her colleagues (2008) only produced reliable subject-predicate boundary cues in interrogative sentences. The prosody of sub-clausal phrases also varies across languages (e.g., Fisher & Tokura, 1996a; Gerken, 1996; Jusczyk, 1997).

(9) a) [$_{PhP}$ The girl] [$_{PhP}$ kissed the boy]

b) [$_{PhP}$ She kissed the boy]

Although there is a large body of evidence attesting to the existence of prosodic cues at clause and some phrase boundaries, other researches have pointed out that non-syntactic factors also affect the manifestation of prosodic boundary cues. Segmental lengthening does not only occur in final-syllable position: it is also affected by mood, emphasis, stress, phonetic category, etc. (Fernald & McRoberts, 1996; Klatt, 1976). Similarly, pausing is highly influenced by speech rate and utterance length, with speakers tending to pause near the middle of an utterance, even if that occurs inside a syntactic constituent (Gee & Grosjean, 1983). Even pauses that occur at syntactic boundaries are influenced by speech rate and constituent length, with more pauses occurring after long subject NPs than shorter subject NPs and hardly any pauses occurring after pronominal subject NPs (Gerken, Jusczyk, & Mandel, 1994; Nespor & Vogel, 2007). Finally, intonational contours are affected by emphasis (Eady, Cooper, Klouda, Mueller, & Lotts, 1986) and social variables (Apple, Streeter, & Krause, 1979).

Given this variability in the manifestation of prosodic cues, Fernald and McRoberts (1996) point out that it is important to consider not only the probability of prosodic cues given the syntactic structure in question, but also the probability of that structure given a particular prosodic cue. For example, the probability of final-syllable lengthening at a major syntactic boundary is quite high, but the probability of a major syntactic boundary occurring immediately after a lengthened syllable is only about 65% (Klatt, 1976). Thus, an infant who posits a syntactic boundary after every lengthened syllable is going to have a lot of false positives. The same holds for pauses – some studies have found that only about 50% of pauses occur at major syntactic boundaries, with the rest occurring as the result of hesitations, speech errors and other non-syntactic phenomena (Maclay and Osgood, 1959).

However, these non-syntactic influences on prosody are reduced in IDS. Broen (1972) found that 93% of IDS sentence boundaries were followed by a pause (29% in

ADS), and that 99% of pauses occurred at boundaries (54% in ADS). Of interest in this dissertation are clause boundaries, of which sentence boundaries are just a subset, but Broen's findings suggest that non-syntactic influences on prosody may not be as problematic as they seem during acquisition.

The imperfect relationship between syntactic and prosodic structure clearly rules out prosody as a sufficient cue for complete syntactic parsing of sentences. However, as Morgan (1996) argues, infants are likely able to subtract the irrelevant influences on prosody in order to narrow in on the relevant prosodic cues. In fact, infants have been shown to do the reverse, filtering out syntactic influences on prosody to determine the positive or negative affect of an utterance (Fernald, 1993).

No theory of prosodic bootstrapping posits that prosody is the sole source of grouping or bracketing information. It is merely one tool that the learner can use to acquire the grammar of her language, and other learning mechanisms can likely take over where prosody fails. Other sources of information that may be used for locating syntactic constituents in conjunction with or instead of prosody are reviewed in Section 2.3.

2.2 Prosodic Bootstrapping of Syntax

While evidence attesting to the reliability of prosodic cues at clause boundaries in IDS is presented above, the mere presence of these cues will not benefit the infant if she is unable to perceive subtle changes in pitch, duration, and intensity or if she is unable to use this information for grouping the speech stream, locating constituents, or inferring phrase structure. The literature regarding infant perception of prosody at syntactic boundaries is reviewed in Section 2.2.1 and the influence of prosody on memory and grouping is covered in Sections 2.2.2. Adult use of prosodic cues is discussed in Section 2.2.3.

2.2.1 Infants' Perception of Prosodic Cues at Syntactic Boundaries

Infants' sensitivity to the prosodic correlates of syntactic boundaries has been assessed by measuring looking time preferences for speech in which a short pause was inserted at a syntactic boundary versus inside a syntactic constituent. In a foundational study (Hirsh-Pasek et al., 1987), a mother was recorded speaking to her 19-month-old child. Pauses longer than 450ms were removed in order to neutralize the effect of naturally-occurring pauses, and 1-second pauses were inserted either at clause boundaries (clause-coincident; example (10a), slashes indicate inserted pauses) or inside the clauses (clause-internal; example (10b)).

(10) a) Cinderella lived in a great big house, / but it was sort of dark / because she had this mean, mean, mean stepmother. / And, oh, she had two stepsisters / that were so ugly. / They were mean, too.

b) ...in a great big house, but it was / sort of dark because she had / this mean, mean, mean stepmother. And, oh, she / had two stepsisters that were so / ugly. They were mean, / too. They were...

Using the Head-Turn Preference Procedure, the authors found that 7- to 10-month-olds listened significantly longer to speech when the artificial pauses were coincident with clause boundaries rather than in the middle of the clauses. Subsequent work found that the effect obtained with IDS, but not ADS (Kemler Nelson, Hirsh-Pasek, Jusczyk, & Wright-Cassidy, 1989). A similar preference for speech with clause-coincident pauses was found in infants as young as 4.5 months, and, at 6 months, when the speech was low-pass filtered (Jusczyk, 1989). Low-pass filtering removes most information that identifies phonetic segments, but preserves the prosodic features. These studies suggest that infants are sensitive to violations of typical correlations between pauses and other prosodic boundary cues, at least with the exaggerated prosody of IDS.

This work has been extended to both Japanese and German infants. Hayashi, Tamekawa, Deguchi, and Kiritani (1996) tested Japanese infants on Japanese sentences with clause-coincident and clause-internal pauses and found the 10-month-old but not 6-month-old infants discriminated between the two types of stimuli, a delay of several months compared to the American infants tested by Jusczyk (1989). Also, the Japanese infants showed a novelty preference, preferring to listen to sentences with clause-internal pauses, contrary to the American infants' consistent familiarity preference for clause-coincident pauses. A follow-up study by Hayashi and Mazuka (2002) found that Japanese infants showed no preference from 4 to 7 months, a novelty preference from 8 to 11 months, and, finally, a familiarity preference from 12 to 14 months. Thus, Japanese infants are able to detect deviations from typical correlations of prosodic cues, but this ability may develop differently than it does in infants acquiring American English, perhaps because American IDS is particularly exaggerated.

On the other hand, it is also possible that the apparent difference between Japanese and American infants is an artifact of the experimental stimuli or methods. In fact, in a variation of Hirsh-Pasek et al,'s (1987) design, Fernald and McRoberts (1996) tested American infants using an auditory preference procedure. Unlike Jusczyk's (1989) results, 4-month-olds showed no preference for clause-coincident or clause-internal pauses, while 7-month-olds listened longer to sentences with clause-internal pauses, and 10-month-olds listened longer to speech with clause-coincident pauses. This closely resembles the pattern seen with Japanese infants. Regardless, it is important to note that the direction of preference is not of utmost importance, since a preference for either type of sentence indicates that the infant can discriminate between pauses that occur in conjunction with other prosodic cues at clause boundaries and pauses that occur inside the clause.

Schmitz (2009; Schmitz, Hohle, & Weissenborn, 2003) further extended this line of research to German 6-month-olds, who discriminated between German sentences with clause-coincident and clause-internal pauses, showing a novelty preference. Overall, although there are differences in infants' direction of preference across studies and differences in the age at which discrimination effects are seen (summarized in Table 2.1), converging data from three different languages suggest that infants are sensitive to the prosodic correlates of clauses.

<table>
<tr><th>Age
(in months)</th><th>American
(Hirsh-Pasek et al., 1987; Jusczyk, 1989; Jusczyk et al., 1992; Kemler Nelson et al., 1989)</th><th>American
(Fernald & McRoberts, 1996)</th><th>Japanese
(Hayashi et al., 1996; Hayashi & Mazuka, 2002)</th><th>German
(Schmitz et al., 2003)</th></tr>
<tr><td>4</td><td>Familiarity</td><td>None</td><td rowspan="4">None</td><td rowspan="2"></td></tr>
<tr><td>5</td><td></td><td rowspan="2"></td></tr>
<tr><td>6</td><td>Familiarity
(low-pass filtered speech)</td><td>Novelty</td></tr>
<tr><td>7</td><td rowspan="4">Familiarity</td><td>Novelty</td><td rowspan="8"></td></tr>
<tr><td>8</td><td rowspan="2"></td><td rowspan="4">Novelty</td></tr>
<tr><td>9</td></tr>
<tr><td>10</td><td>Familiarity</td></tr>
<tr><td>11</td><td rowspan="4"></td><td rowspan="4"></td></tr>
<tr><td>12</td><td rowspan="3">Familiarity</td></tr>
<tr><td>13</td></tr>
<tr><td>14</td></tr>
</table>

TABLE 2.1 Summary of pause-insertion experiments focusing on infant perception of clause-level prosodic cues in their native language. Results from studies using ADS are not included here. “Familiarity” indicates a preference for the familiar, clause-coincident pauses. “Novelty” indicates a preference for the novel, clause-internal pauses. “None” indicates that no listening preference was found.

The pause-insertion technique has also been used to investigate infants’ sensitivity to the subject-predicate boundary. In a series of experiments, Jusczyk and his colleagues found that 9-month-olds, but not 6-month-olds, listened significantly longer to speech in which a pause was artificially inserted at the boundary between the subject and predicate phrases rather than elsewhere in the sentence, even when the speech was low-pass filtered (Jusczyk et al., 1992). Interestingly, while Kemler Nelson and colleagues (1989) failed to find a clause-level effect with ADS, infants were able to discriminate phrase-coincident and phrase-internal pauses in ADS. This discrepancy may be due to the fact that the adult-directed stimuli from Jusczyk et al. (1992) were recordings of a woman reading from a storybook that was originally designed to elicit child-directed speech, while the adult-directed stimuli used by Kemler Nelson and colleagues were taken from a recording of a conversation between two adults. Nevertheless, Jusczyk et al. (1992) provides evidence that infants can perceive relatively subtle prosodic cues that tend to correlate with the subject-predicate boundary.

Together, these 'pause-insertion' experiments are often taken as evidence that infants can perceive clause and subject-predicate boundaries. However, there is an important distinction between perceiving the syntactic boundary itself and perceiving the prosodic correlates of that boundary. Gerken, Jusczyk, and Mandel (1994) tackled this question by testing infants on sentences with unstressed pronominal subjects (like (11)), for which prosodic and syntactic boundaries were mismatched. The authors found that 9-month-olds only preferred to listen to sentences in which pauses occur at the subject-predicate boundary if there is also a PhP boundary (like in (12)), suggesting that infants do not perceive the syntactic boundaries themselves, but rather are reacting to the unnaturalness of non-coincident stimuli.

(11) a) $[_{PhP}$ He ate four strawberries]

b) $[_{NP}$ He] $[_{VP}$ ate four strawberries]

(12) a) $[_{PhP}$ The caterpillar] $[_{PhP}$ ate four strawberries]

b) $[_{NP}$ The caterpillar] $[_{VP}$ ate four strawberries]

There are two primary remaining interpretations for results from pause-insertion studies. First, infants' discrimination between coincident and non-coincident (straddling) pauses could be an effect of the physiological unnaturalness of non-coincident pauses – a continuous vocalization does not suddenly drop into silence. Alternatively, it could be based on infants' awareness of the behavior of prosodic boundary cues – pauses, syllable lengthening, and pitch resets tend to correlate. Morgan, Swingley, and Miritai (1993) addressed these possibilities through a slight variation of the experimental methods. Instead of measuring looking time preferences, they trained 10-month-olds to turn their heads whenever they heard a buzzing noise. Buzzes were inserted during either clause-final or clause-medial syllables in a passage of fluent speech. If the previous findings were due to an aversion to non-coincident pauses, the infants should be equally good at reacting to buzzing noises at clause boundaries and inside of the clause. However, the participants had a higher success rate when the noises occurred at a clause boundary. This supports previous findings and suggests that infants are indeed sensitive to the prosodic correlates of clauses. While this is an improvement over the pause-insertion design, note that naturally-occurring pauses were still deleted from the stimuli passages (J. Morgan, personal communication, July 23, 2012), so the typical constellation of pausing, syllable lengthening, and pitch resets at clause boundaries was not maintained.

Finally, event-related brain potential (ERP) data also attest to infant perception of prosodic boundaries. Pannekamp, Weber, and Friederici (2006) played German sentences to 8-month-old German infants. The critical part of the sentence either contained one (13a) or two (13b) IPs. The infants showed a reliable positive-going waveform at the boundary between IPs in the 2-IP condition. This is known as a closure positive shift (CPS; see Section 2.2.3 for further discussion of this effect). Since the infants were only 8 months old, the CPS was likely elicited by the IP boundary, and was not an effect of the syntax or semantics. This provides further evidence that infants can perceive at least some IP boundaries.

(13) a) [$_{IP}$ Kevin verspricht Mama zu schlafen]...

[$_{IP}$ *Kevin promises Mom to sleep*]...

b) [$_{IP}$ Kevin verspricht] [$_{IP}$ Mama zu küssen]...

[$_{IP}$ *Kevin promises*] [$_{IP}$ *to kiss Mom*]...

Overall, evidence from a long line of pause-insertion experiments suggests that infants are sensitive to deviations from typical correlations between prosodic cues at major syntactic boundaries before the end of their first year of life and in at least three languages. Additional evidence (Morgan et al., 1993; Pannekamp et al., 2006) lends further support for infant perception of prosodic boundary cues – an important prerequisite to prosodic bootstrapping.

2.2.2 Infants' Use of Prosody in Speech Processing and Grouping

Once the pause-insertion work of the late 1980s and early 1990s established that clause-final, and sometimes phrase-final, prosodic cues are perceivable by infants, researchers turned their attention to the next logical question: can learners use these cues to facilitate syntax acquisition? Starting with Mandel, Jusczyk, and Kemler Nelson (1994), the next decade of inquiry into the prosodic bootstrapping of syntax focused on the influence of prosodic structures on memory, drawing from older work suggesting that imposing linguistic structure enhances memory in adults. For example, Epstein (1961) found that adults were able to learn and recall a sequence of nonsense syllables more quickly if the syllables contained grammatical inflections (see also Forster, 1966; Miller & Isard, 1963; Suci, 1967). Therefore, Mandel and colleagues argue, if prosodically well-structured speech is easier to remember than prosodically ill-formed speech, it would suggest that prosody influences processing of speech. Using the High Amplitude Sucking Procedure, the authors found that 2-month-olds were more successful at detecting a phonetic substitution, as in (14a) versus (14b), if the string of words occurred with a single well-formed sentential prosodic contour rather than as a list ("the, rat, chased, white, mice") or as two well-formed sentence fragments during familiarization. Similarly, 2-month-olds were better able to detect a word order change, as in (15a) versus (15b), when the words were initially presented as a single well-formed sentence rather than as two well-formed sentential fragments (Mandel, Kemler Nelson, & Jusczyk, 1996). These studies suggest that prosody impacts how infants as young as 2 months process and encode incoming acoustic stimuli, but they do not directly address the role of prosody in syntax acquisition.

(14) a) The *r*at chased white mice.

b) The *c*at chased white mice.

(15) a) Cats *would jump* benches.

b) Cats *jump wood* benches.

Nazzi, Kemler Nelson, Jusczyk, and Jusczyk (2000) further built on the assumption that memory is indicative of how infants process and encode linguistic input. They familiarized 6-month-olds with two short strings of words, (16) and (17). Each infant heard one of the passages as a prosodically well-formed sentence and one as a prosodically ill-formed sequence (e.g., a participant might be familiarized with (16a) and (17b)). Later, the infants were tested on longer passages of fluent speech containing the familiarization sequences. The prosody of the target strings in the test items matched the prosody with which the infant initially heard that string. Infants listened significantly longer to the passage containing the familiarization sequence that they originally heard as a prosodically well-formed sentence. This provides additional support for the idea that prosody affects how infants remember strings of words, but, since the target sequences were initially presented in isolation, the infants did not have to use prosody to extract a group of words from the speech stream. The results, therefore, do not speak to the grouping, constituents, or phrase structure versions of prosodic bootstrapping.

(16) rabbits eat leafy vegetables

a) Rabbits eat leafy vegetables.

b) …rabbits eat. Leafy vegetables...

(17) leafy vegetables taste so good

a) Leafy vegetables taste so good.

b) ...leafy vegetables. Taste so good....

Soderstrom, Seidl, Kemler Nelson, and Jusczyk (2003) replicated Nazzi et al.'s (2000) methodology, looking at phrases versus phrase-straddling sequences. Sentences (18a) and (18b) both contain the target sequence “design telephones,” but in (18a), the target is entirely in the predicate, while it straddles the subject-predicate boundary in (18b). Both 6- and 9-month-olds listened longer to test passages that included the well-formed phrase, extending Nazzi and colleagues' (2000) results below the level of the clause. However, this study suffers from the same shortcoming, namely that the infants were familiarized with pre-extracted snippets of speech, so it is not clear that the memory effect relates to speech processing or grouping.

(18) a) [$_{NP}$ Inventive people] [$_{VP}$ design telephones at home]

b) [$_{NP}$ The director of design] [$_{VP}$ telephones at home]

Soderstrom, Kemler Nelson, and Jusczyk (2005) addressed this through a variation of Nazzi et al.'s (2000) experiment. Six-month-olds were familiarized with either passage (19) or (20), both of which contain the sequence "rabbits eat leafy vegetables taste so good." Both passages have a well-formed clause ((19a) and (20a)) and a clause-straddling sequence, in which the target words are split across a clause boundary ((19b) and (20b)). The well-formed clause target for one group was the clause-straddling target for the other. (Note that, while the authors refer to the well-formed strings as "clauses," they are full utterances.)

(19) Many animals prefer some things. Rabbits eat leafy vegetables. Taste so good is rarely encountered.

a) Rabbits eat leafy vegetables.

b) ...leafy vegetables. Taste so good...

(20) John doesn't know what rabbits eat. Leafy vegetables taste so good. They don't cost much either.

a) Leafy vegetables taste so good.

b) ...rabbits eat. Leafy vegetables...

At test, the infants listened longer to the passage that contained the well-formed clausal target sequence from familiarization. They also listened longer to passages in which the target sequence matched its initial familiarization prosody (i.e., they listened longer to a clause-straddling test item if it was initially presented with clause-straddling prosody), such that the condition with the longest looking times was the one in which a well-formed familiarization sequence occurred with well-formed prosody at test. This study goes beyond previous work by suggesting that prosody plays a role in extracting groups of words from short passages of fluent speech during on-line processing. This is evidence in favor of the simplest version of prosodic bootstrapping – grouping, since the infants used prosody to chunk the speech stream into units of memory. However, it is important to keep in mind that the infants were also relying on acoustic similarity of extracted groupings, an issue that is considered in depth in Section 3.1.7.

Seidl (2007) also replicated the Nazzi et al. (2000) experiment, but manipulated the availability of prosodic cues. She found that infants preferred the prosodically well-formed target sequence when pauses or lengthening cues were removed, but not if the pitch contour was flattened. This suggests that pitch may be the only necessary cue for obtaining prosodic bootstrapping effects, at least in American English IDS. However, pitch information is not sufficient – when both pauses and final-syllable lengthening were removed, infants had no preference for well-formed versus ill-formed sequences. This is an important finding, since it suggests that individual variation (e.g. Soderstrom et al., 2008) and cross-linguistic variation in the manifestation of prosodic boundary cues would not derail the prosodic bootstrapping mechanism (see Section 2.4.1 for further discussion of the effect of cross-linguistic prosodic differences on learning).

A final memory study was conducted by Johnson and Seidl (2008). They replicated Nazzi et al.'s (2000) results with Dutch-acquiring 6-month-olds listening to Dutch stimuli, providing evidence that prosody facilitates memory in at least one other language. Unlike the English infants in Seidl (2007), however, Dutch infants showed no preference for well-formed versus ill-formed sequences if pauses were neutralized (see Section 2.4 for discussion of this finding).

A second line of evidence that infants use prosody for grouping has taken an entirely different approach. Gout, Christophe, and Morgan (2004) trained 10- and 13-month-olds to turn their head when they heard the word *paper* or *beacon*. At test, the infants were more likely to turn their head if the word occurred within a PhP (21a), than if the syllables were separated by a PhP boundary (21b). Similar results were found with French-acquiring 16-month-olds listening to French words that either straddled a PhP boundary or occurred inside a PhP (Millotte et al., 2011). These experiments provide further support for the hypothesis that infants use prosody for grouping the speech stream, and it even suggests that they do indeed use those groupings as units for further analysis – in this case, finding word boundaries.

(21) a) [$_{PhP}$ The scandalous paper] [$_{PhP}$ sways him]....

b) [$_{PhP}$ The outstanding pay] [$_{PhP}$ persuades him]...

Together, these studies offer evidence that prosody influences the way infants remember and process linguistic input (Johnson & Seidl, 2008; Mandel et al., 1994; Mandel et al., 1996; Soderstrom et al., 2003; Seidl, 2007), and even that infants use prosody to group the speech stream (Gout et al., 2004; Millotte et al., 2011; Soderstrom et al., 2005), providing support for the grouping version of the prosodic bootstrapping hypothesis. The experiments in Chapter 3 extend this work to ask whether infants map prosodically-extracted groups of words onto syntactic constituents.

2.2.3 Adult Studies

While the prosodic bootstrapping hypothesis relates to syntax acquisition, not syntax processing, it should be noted that several studies have presented evidence that adults use prosody for grouping-type tasks in online language processing. For example, adults can disambiguate phrases like (22) (from Aasland & Baum, 2003), using duration cues (Aasland & Baum, 2003; Scott, 1982), as well as pitch information (Streeter, 1978). Disambiguating between (22a) and (22b) relies on achieving the correct phrasal groupings.

(22) pink and black and green

a) [pink and black] and [green]

b) [pink] and [black and green]

It is therefore not surprising that adults, like infants, show a preference for speech in which pauses are inserted at prosodic boundaries. In a variation of Hirsh-Pasek et al's (1987) study, Jusczyk and colleagues (1992) asked adult participants to rate speech in which pauses were inserted at the subject-predicate boundary versus elsewhere in the sentence. Stimuli were low-pass filtered so that participants could not rely on lexical information. In a forced-choice task, adults rated speech with phrase-coincident pauses as more natural, both for IDS and ADS.

ERP evidence also suggests that adults are sensitive to the prosodic features of speech. In a precursor to the Pannekamp et al. (2006) study discussed in Section 2.2.1, Steinhauer, Alter, and Friederici (1999; c.f Steinhauer, 2003) played German sentences to native German speakers. The sentences had either two or three IPs, and participants showed a reliable CPS at IP boundaries. However, IP boundaries always corresponded to clause boundaries in this study, so it is unclear whether the CPS was related to prosodic or syntactic structure.

Later work, however, suggests that the CPS is indeed an effect of prosodic processing. Pannekamp, Toepel, Alter, Hanhe, and Friederici (2005) again played sentences with two and three IPs to German adults, but the stimuli were controlled for possible sources of the CPS. Each sentence was in one of four conditions: normal sentences, jabberwocky sentences (in which content words (nouns, verbs, adjectives) were replaced by nonsense words to remove semantic content), pseudoword sentences (in which both function and content words were replaced by nonsense words to remove syntactic structure), and hummed sentences (in which the intonational contour was preserved, but there was no phonemic content). An example two-IP sentence for each condition is given in (23) (a = normal, b = jabberwocky, c= pseudoword, d = hummed).

(23) a) [$_{IP}$ Kevin verspricht Mama zu schlafen] [$_{IP}$ und ganz lange lieb zu sein]

[$_{IP}$ *Kevin promises Mom to sleep*] [$_{IP}$ *and to be a good boy for awhile*]

b) [$_{IP}$ Der Bater verklicht Onna zu labeiken] [$_{IP}$ und das Rado zu nupen]

[$_{IP}$ *The bater rables Onna to lubol*] [$_{IP}$ *and the rado to nupe*]

c) [$_{IP}$ Bater saklimm Onna ko labei keg] [$_{IP}$ nug som Rado lie nupes]

d) [$_{IP}$ mm mmm mmmm mm mmmm] [$_{IP}$ mmm mmm mmm mmm mmmmm]

While there were differences in the localization of the CPS on the scalp based on experimental condition, for all types of sentences, the two-IP sentences elicited a CPS at the internal IP boundary, and the three-IP sentences elicited CPSs at both internal IP boundaries. ERP evidence is difficult to interpret, but this study lends credence to the hypothesis that adults use prosody, at least at the IP level, to group the speech stream.

Recent work on word segmentation in an artificial grammar provides additional evidence that adults use prosody for grouping. Shukla, Nespor, and Mehler (2007) familiarized Italian adults with trisyllabic nonsense words (e.g., (24a) and (24b)) embedded in a string with seven other random syllables. These 10-syllable sequences were synthesized to have natural IP-like intonational contours and final-syllable lengthening, although there were no pauses between IPs: each 10-syllable sequence was immediately followed by another. The target trisyllabic words either were contained entirely inside the IP, as for (24a) in example (24), or spanned an IP boundary, as for (24b). Transitional probabilities between adjacent syllables in the word targets were 1, while transitional probabilities between syllables throughout the rest of the stimuli averaged about 0.1.

(24) … [$_{IP}$ ti-ro-mi-<u>ni-da-fo</u>-pa-po-mi-sa-du-<u>te-ki][$_{Ip}$ me</u>-sa-pa...

a) ni-da-fo

b) te-ki-me

At test, participants were given a forced-choice task, in which they were asked to judge whether a word target (either IP-straddling or IP-internal) or a non-word sequence of syllables belonged to the familiarization language. Participants were significantly above chance at recognizing IP-internal words, but performed at chance for IP-straddling words, suggesting that adults do not consider high transitional probability sequences of syllables to be possible candidates for words if the sequence is interrupted by an IP boundary. Thus, in line with Gout et al.'s (2004) infant results, adults grouped the stimuli into prosodic phrases and then used those groupings as the domain over which they searched for words.

Only two studies to date have begun probe the role of prosody in syntax learning beyond grouping the speech stream. Langus, Marchetto, Bion, and Nespor (2012) familiarized Italian adults with sentences composed of six nonsense syllables organized into two phrases. PhP-like phrases were marked with final lengthening and IP-like sentences were marked with a declining pitch contour. There were no pauses between or within sentences, and all prosodic features were artificially synthesized. Both the phrases and the sentences had long-distance dependencies, whereby the first syllable of a phrase predicted the last syllable of that phrase, and the first phrase in the sentence predicted the second phrase in the sentence. At the phrase level, groupings had the form AXB, such that if the first syllable (A) was *to*, the last syllable (B) was always *fe*. The phrases are listed in (25a-c). There were three middle syllable options for each phrase (X = *se*, *ko*, and *lu*).

(25) $[_{PhP}$ AXB]

a) $[_{PhP}$ to-X-fe]

b) $[_{PhP}$ mu-X-gi]

c) $[_{PhP}$ ba-X-pe]

The phrases were organized into three two-phrase sentences, shown in (26a-c). Since the first phrase in the sentence predicted the second phrase in the sentence (for example, sentences beginning with *to X fe* always ended with *mu X gi*), there was an adjacent dependency between the last syllable of the first phrase and the first syllable of the second phrase (*fe* was always followed by *mu*) For clarity, I will refer to the first phrase as A_1XB_1 and the second phrase as A_2XB_2, but note that the A_1XB_1 phrase for one sentence serves as the A_2XB_2 phrase for another.

(26) $[_{IP}$ $[_{PhP}$ A_1XB_1] $[_{PhP}$ A_2XB_2]]

a) $[_{IP}$ $[_{PhP}$ to-X-fe] $[_{PhP}$ mu-X-gi]]

b) $[_{IP}$ $[_{PhP}$ ba-X-pe] $[_{PhP}$ to-X-fe]]

c) $[_{IP}$ $[_{PhP}$ mu-X-gi] $[_{PhP}$ ba-X-pe]]

Participants were tested using a forced-choice task; half made judgments about phrases, and half made judgments about sentences. Prosody was neutralized for all test items. Participants making judgments about phrases were presented with A_1YB_1 (27a) and B_1A_2X or XB_1A_2 (27b) items and were asked to judge which was more likely to be part of the familiarization grammar. The A_1YB_1 items were prosodically-marked phrases from familiarization, although a new middle syllable was introduced (e.g., *mu*). Thus, the transitional probability of the A_1YB_1 strings was 0. The B_1A_2X and XB_1A_2 items were part-phrases. They were valid strings from familiarization and, therefore, had non-0 probability, but they crossed the PhP boundary. Participants in this condition chose the phrases that were prosodically-marked during familiarization more often than the part-phrases, suggesting that participants used PhPs to group the stimuli into smaller chunks. Similar to Shukla et al.'s (2007) findings, participants either did not track or did not heavily weight transitional probabilities across PhP boundaries.

(27) a) to-Y-fe; mu-Y-gi; ba-Y-gi

b) fe-mu-X; X-fe-mu; gi-ba-X; X-gi-ba; pe-to-X; X-pe-to

The second group of participants chose between sentences that had novel syllables in medial position ($A_1YB_1A_2YB_2$; (28a)) and a sentences in which the dependency between the first phrase and the second phrase was violated, such that phrases from familiarization occurred in different combinations ($A_1XB_1A_3XB_3$; (28b)). Participants chose the sentences with novel syllables in medial position as better exemplars of the familiarization grammar more often than sentences that violated the dependency between the first and second phrases. The authors argue that, since participants used prosody both to group the speech into phrases and to group it into sentences (albeit, this was tested between-participants), this suggests that adult learners can use prosody to learn hierarchical structure – phrases embedded in sentences.

(28) a) ba-Y-pe-to-Y-fe; mu-Y-gi-ba-Y-pe; to-Y-fe-mu-Y-gi

b) ba-X-pe-mu-X-gi; mu-X-gi-to-X-fe; to-X-fe-ba-X-pe

However, there is an alternative explanation for these results. During familiarization, the phrases always occurred in the same order – phrase 1 (25a), then phrase 2 (25b), then phrase 3 (25c), then phrase 1, phrase 2, etc., as in (29), where each number represents the phrase number. Therefore, sentences of the form $A_1XB_1A_3XB_3$ (e.g., phrase 1 followed by phrase 3 or phrase 2 followed by phrase 1) did not occur as possible strings in the familiarization items. Effectively, the participants were asked to choose between $A_1YB_1A_2YB_2$ strings, in which the variable medial words introduces additional variability and $A_1XB_1A_3XB_3$ strings, in which phrases are grouped into sentences that do not occur either within an IP or spanning an IP boundary during familiarization.

(29) … $[_{IP}[_{PhP}1][_{PhP}2]]$ $[_{IP}[_{PhP}3][_{PhP}1]]$ $[_{IP}[_{PhP}2][_{PhP}3]]$ $[_{IP}[_{PhP}1][_{PhP}2]]$ $[_{IP}[_{PhP}3][_{PhP}1]]$ $[_{IP}[_{PhP}2][_{PhP}3]]$ …

A stronger test for ascertaining whether adults are indeed learning an embedded grammar would be to ask participants to choose between $A_1YB_1A_2YB_2$ and $A_2XB_2A_3XB_3$ strings, the later of which occur in the familiarization items, but do not form a single IP contour. Nonetheless, is still notable that participants rated the $A_1YB_1A_2YB_2$ strings as better examples of the familiarization grammar than $A_1XB_1A_3XB_3$ strings, since it suggests that they are learning a dependency that crosses a PhP boundary. However, we cannot infer that participants were acquiring an embedded, hierarchical structure.

The second study to look beyond prosody as a cue to grouping was conducted by Morgan, Meier, and Newport (1987; Experiment 1). The authors asked whether adults could use prosody as a cue to hierarchically-organized structures by replicating Morgan and Newport (1981) (see Section 1.2) with the addition of prosodic cues. Participants were familiarized with an artificial grammar in which nonsense words corresponded to unfamiliar-shape referents. The grammar underlying sequences of shape/word pairs could be construed as a finite-state grammar or as hierarchically-organized. Unlike Morgan and Newport (1981), Morgan et al. (1987) manipulated input prosody as a cue to constituent groupings. All participants were exposed to the same set of sentences, but some heard the sentences with list prosody, some with arbitrary prosodic phrasing (i.e., prosodic phrases did not correspond to constituents), and some with prosodic phrasing that was consistent with constituents.

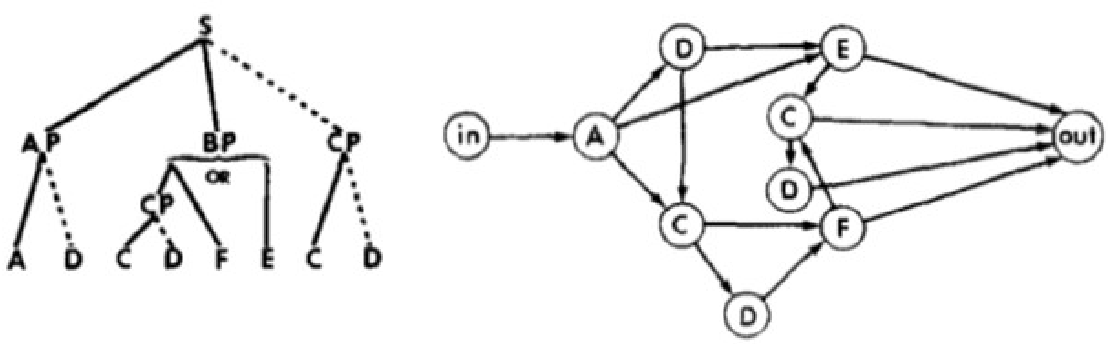

FIGURE 2.4 Hierarchical (left) and finite-state (right) grammars from Morgan and Newport (1981), Morgan et al., (1987), and Morgan et al. (1989). Figures from Morgan et al. (1987).

Participants were given several forced-choice tests to assess their learning of the experimental grammar. The two most relevant tests were the fragment test and the movement test (called the "transformational" test by the authors). In the fragment test trials, participants were asked to judge between two sentence fragments. The sentence fragments all contained possible strings from familiarization. Some strings were phrasal groupings, such as "AD," while other strings, such as "DC," violated the hierarchical structure but not the phrase structure grammar. In the movement test, participants judged between sentences in which a phrase (e.g. "AD") or a non-phrase string (e.g. "DC") moved to a new position in the sentence. In both the fragment and movement tests, participants in the list prosody and the arbitrary prosody conditions did not score reliably above chance in identifying the phrasal groupings, while participants in the constituent-consistent prosody condition identified the phrasal groupings significantly more often than chance would dictate. This suggests that participants whose input was not prosodically structured (or whose input contained misleading prosodic cues) learned a simpler finite-state grammar, while participants whose input was prosodically-grouped according to syntactic constituents may have acquired a hierarchical grammar.

It is also possible, if not probable, however, that participants in the constituent-consistent prosody condition were merely learning to group the stimuli sentences into constituent-like units. In order to demonstrate learning of a hierarchical grammar, it is necessary to show that participants are aware that smaller constituents are embedded inside larger constituents, which this experiment does not do. However, Morgan et al.'s (1987) work does go beyond grouping: for a learner to prefer sentences in which prosodically-grouped words move to a new position in the sentence over sentences in which a non-phrase string of words (that was nonetheless a valid string from familiarization) moves, it suggests that the learner is treating the prosodically-grouped words as constituent-like units. This issue is further addressed in the introduction to Chapter 3.

Thus, evidence suggests that adults do use prosody for grouping in online speech processing and in artificial grammar learning tasks. Morgan et al. (1987) even provides initial evidence for the constituents version of the prosodic bootstrapping hypothesis.

2.3 Other Useful Cues for the Learning of Constituency and Phrase Structure

The evidence reviewed above suggests that prosodic and syntactic boundaries co-occur with regularity in IDS, that infants and adults are sensitive to these co-occurrences, and that learners can use prosody for grouping the speech stream. The experiments in Chapter 3 test whether learners can use prosodic information to bootstrap knowledge of syntactic constituency. By no means, though, is prosody the only available cue for grouping the speech stream or inferring that certain strings of words are syntactic constituents. Other potential sources of information about grouping, constituency, and maybe even phrase structure include functional morphology, transitional probabilities, semantics, and cross-sentential comparisons. These cues may function independently or interact in the learner's acquisition process.

First, function morphemes have often been hypothesized to help infants identify constituent boundaries. Function morphemes are frequent and phonologically distinct, making them easy for the learner to recognize. For example, English function morphemes tend to be monosyllabic words with a reduced vowel, such as /ə/ or /ɨ/. Indeed, even newborns are able to discriminate between function and content morphemes (Shi, Werker, & Morgan, 1999). Function morphemes also typically occur at either the beginning or end of a constituent (Clark & Clark, 1977). For example, the English morphemes *the*, *of*, and *will* introduce NPs, VPs, and PPs (prepositional phrases), respectively. Thus, positing a boundary before freestanding function morphemes would be a useful strategy for an English learner learning to group the speech stream or locate constituents.

One aspect of functional morphology that may be particularly useful for grouping speech or locating constituents is concord morphology or morphological agreement. In many languages, affixes marking gender or number information must agree within a constituent, as shown in the Spanish example in (30). (Note that there are also morphological agreement processes that cross constituent boundaries, such as subject-verb agreement.)

(30) [La chica] [comió [muchas cerezas negras]]
the-fem child-fem ate many-fem-pl cherries-fem-pl black-fem-pl
The girl ate many black cherries

Morgan et al. (1987, Experiment 2) examined the role of concord morphology in learning constituency in an artificial grammar. They replicated the design of Experiment 1 (see Section 2.2.3), but instead of manipulating the availability and validity of prosody as a cue to constituents, participants listened to sentences in which there were no morphological cues, in which function morphemes (affixes) were arbitrarily attached to words, or in which function morphemes agreed throughout constituents (e.g., the suffix *-ro* was attached to the both words in the constituent "AD"). There were no significant differences in performance across the three conditions in the fragment test. However, the participants from the constituent-consistent morphology condition preferred test sentences in which a constituent grouping had moved to a new position in the sentence significantly more often than participants in the arbitrary morphology condition. This suggests that adults are able to use concord morphology to group speech into constituent-like units. However, note that participants in Experiment 1, in which constituency was marked with prosody, showed learning in both the fragment and movement tests, hinting that prosody may be a stronger cue to constituency than concord morphology, at least for adult speakers of English, which has an impoverished system of inflectional morphology.

A second source of information that learners could potentially use to bootstrap syntactic knowledge is transitional probabilities. Takahashi and Lidz (2007) investigated the role of transitional probabilities in how adult participants learn constituency (a similar study was conducted by Thompson and Newport (2007)). Listeners were exposed to sentences from one of two grammars, which differed only in constituent structure. For example, both grammars contained the sentence ABCDE (with each letter representing a class of (nonsense) words). However, in Grammar 1, A and B formed a constituent, while B and C formed a constituent in Grammar 2. Constituency could be inferred based on the transitional probabilities between the word classes. The transitional probability within a constituent was always 1, while the transitional probability across constituent boundaries was much lower. Experiment 1 also contained cross-sentential cues to constituency, such as movement and pro-form replacement of certain constituents, while Experiment 2 only used transitional probability.

After familiarization, participants were given a forced-choice test. For each trial, participants heard a pair of word sequences and were asked to select the sentence that came from the grammar they had just heard. None of the test sentences occurred during familiarization. Some of these trials required the participants to differentiate between grammatical sentences in which constituents were moved and ungrammatical sentences in which non-constituent sequences were moved. Both in Experiment 1, which contained multiple cues to constituency, and in Experiment 2, which only used transitional probability to mark constituent structure, participants were able to differentiate grammatical and ungrammatical movements. This suggests that transitional probability can facilitate grouping the speech stream into constituent-like units.

Semantics is a third source of information about syntactic phenomena. While the semantic bootstrapping hypothesis (Pinker, 1984) largely deals with acquisition of grammatical category information, semantics may also be useful for locating constituents because constituents tend to be meaningful units. However, this will only be a useful cue for learners with more advanced lexical knowledge.

Functional morphology, transitional probabilities, and semantics are all locally-available cues to constituency. A final source of information comes from comparing the behavior of groups of words across multiple sentences. As discussed in Section 1.1, phenomena such as pro-form replacement (31a) and movement (31b) depend on syntactic constituents. Many constituents are also optional, as in (31c), and constituents can occur as sentence fragments (31d). Sentence fragments, in particular, are common in IDS (Fisher & Tokura, 1996b). IDS also frequently contains repetitions and rephrasings, as in (32) (from Brown (1973); cited in Morgan, Meier, and Newport (1989). The mother in (32) provides several cues that “your foot” is a constituent: she uses it as a stand-alone sentence fragment, she refers to it with the pro-form *it*, and she places it in multiple sentences.

(31) [My brother] [travels [by train] [for business]]

a) He travels by train for business.

b) For business, my brother travels by train.

c) My brother travels.

d) *How does he travel?* By train.

(32) Eve: Oh foot.

Mother: *Your foot*? Where is *your foot*?

Eve: Foot [unintelligible] chair.

Mother: There *it* is. There's *your foot*.

Morgan, Meier, and Newport (1989) tested the role of cross-sentential comparisons, using the same ambiguous (hierarchically-organized or finite-state) grammar as Morgan et al. (1987). Instead of marking constituents with prosody or concord morphology, constituents were marked through cross-sentential means. In one condition, constituents were signaled through occasional pro-form substitutions; in a second condition, constituents were marked through movement operations. Finally, one group of participants did not receive either cross-sentential cue. As in Morgan et al. (1987), participants were tested on sentence fragments that either formed a constituent from the hierarchically-organized grammar or a non-constituent sequence from the simpler finite-state grammar. Participants were also tested on sentences in which a constituent or non-constituent string moved to a new position in the sentence. Although differences between the conditions were not statistically reliable, only participants who received either pro-form or movement cross-sentential cues performed above chance at inferring the hierarchically-organized grammar, suggesting a potential role for cross-sentential cues in identifying syntactic constituents.

Prosody, function morphemes, transitional probabilities, semantics, and cross-sentential comparisons are by no means the only potential cues for grouping the speech stream, locating constituents, or inferring phrase structure. Nor do they operate alone. By making clever use of all available cues, the learner can avoid the pitfalls of any given strategy. For instance, Gerken, Jusczyk, and Mandel (1994) propose that the infant may be able to use cross-sentence comparisons to bolster prosodic bootstrapping, which would fail in cases where prosody and syntax are mismatched. For example, an infant may locate a lexical NP subject in a declarative sentence and then locate a pronominal subject in a yes-no question (where pronoun subjects *are* prosodically-marked; c.f. Gerken et al. (1994)). The learner could then infer that pronoun subjects are also constituents in declarative sentences. Thus, while this dissertation focuses on prosody's role in syntax acquisition, it is important to keep in mind that learners are likely making use of multiple cues in a more natural learning situation.

2.4 The Role of Language-Specific Experience: Bootstrapping from Unfamiliar Prosody

Most work looking at the role of prosody in syntax acquisition has focused on infants and adults listening to their native language. However, the experiments presented in Chapter 4 ask whether prosody is useful in syntax acquisition because of its acoustic salience or because of the learner's acquired knowledge of and experience with her native language's prosodic patterns (for an example of the type of cross-linguistic prosodic variation at issue, see Chapter 4, which includes a detailed comparison of English and Japanese prosody). To answer this question, it is necessary to test learners' ability to perceive and use novel manifestations of prosodic cues. I review infant perception and use of non-native prosody in Section 2.4.1, and I discuss relevant adult studies in Section 2.4.2.

2.4.1 Infant Studies

The literature on language discrimination in newborns, cotton-top tamarins, and rats suggests that prosody is a very salient aspect of speech, even in the absence of language-specific experience (Nazzi, Bertoncini, & Mehler, 1998; Ramus, Hauser, Miller, Morris, & Mehler, 2000; Toro, Trobalon, Sebastian-Galles, 2003). On the other hand, there is also evidence that native-language prosody begins to influence speech perception and production very early on. For example, Friederici, Friedrich, and Christophe (2007) present ERP evidence suggesting a processing advantage for words with native-language stress patterns in French and German infants as young as 4 months, and Halle, de Boysson-Bardies, and Vihman (1991) found evidence for native-language influence on production of disyllabic babbles in French and Japanese 18-month-olds.

Of particular interest here, though, is how infants perceive and interpret the prosodic correlates of major syntactic boundaries in an unfamiliar language. Jusczyk (1989) extended Hirsh-Pasek and colleagues' (1987) foundational pause-insertion experiment, testing English-acquiring infants on Polish. Pauses were inserted either at clause boundaries or inside of clauses. He found that American 4.5-month-olds preferred to listen to speech with clause-coincident pauses in Polish, but that 6-month-olds showed no preference. Six-month-olds even failed to show a preference when the speech had been low-pass filtered, suggesting that their lack of discrimination was not influenced by non-native phonemes or phonotactics. Jusczyk speculates that the difference between infants at 4.5 and 6 months may be because 6-month-olds are actively engaged in other aspects of language learning, perhaps because infants begin to babble around 6 months. Alternatively, the decline in this ability between 4.5 and 6 months suggests that experience may play a role in tuning perception.

While Jusczyk found that young American infants are able to discriminate clause-coincident and clause-internal pauses in Polish, Fernald and McRoberts (1996) tested American infants on German sentences with clause-coincident and clause-internal pauses, and failed to find a listening preference at 4, 7, or 10 months. It is not clear what underlies the different results from the two research groups. In Jusczyk's study, stimuli were played contingent on the infant's headturns, while participants in Fernald and McRoberts' experiments were not required to make headturns – the sounds played during the infant's fixation to a central display. Fernald and McRoberts argue that their procedure, which did not require the infant to turn her head, reduced task demands, and that their procedure is therefore more sensitive to small differences in looking time. However, it is possible that the procedure used by Jusczyk is actually more sensitive, perhaps because the relationship between headturns and stimulus duration may be more obvious to the infant with a headturn-based procedure. This could explain why only Jusczyk found discrimination at 4 months. If the ability to perceive non-native prosodic boundary cues declines between 4 and 6 months, this would be consistent with the lack of discrimination seen in both studies for the older age groups (see Table 2.2 for a comparison of these two studies).

Age (in months)	**American (listening to Polish)** (Jusczyk, 1989)	**American (listening to German)** (Fernald & McRoberts, 1996)
4	Familiarity	None
6	None (replicated with low-pass filtered speech)	
7		None
10		None

TABLE 2.2 Summary of pause-insertion experiments focusing on infant perception of clause-level prosodic cues in an unfamiliar language. "Familiarity" indicates a preference for the familiar, clause-coincident pauses. "None" indicates that no listening preference was found.

Regardless, it is instructive to consider Fernald and McRoberts' results with German stimuli in comparison to their results with English stimuli, since the experiments were methodologically identical. As discussed in Section 2.2.1, they found that American 7- and 10-month-olds discriminated between clause-coincident and clause-internal pauses. Since 7- and 10-month-olds did not discriminate with German stimuli, this suggests a perceptual advantage for native language prosody by around 4 months.

Christophe, Mehler, and Sebastian-Galles (2001) took a different approach to investigate infant perception of PhP-like boundaries in an unfamiliar language. Using the high-amplitude sucking procedure, they habituated French newborns on disyllabic sequences extracted from Spanish sentences. The sequence *la-tí* was recorded in carrier sentences in which it occurred either in the middle of a word (e.g., (33a)) or straddling a word boundary (e.g., (33b)). Since a PhP typically contains a single content word, sequences straddling two content words also tend to straddle a PhP boundary. As with prior findings from French infants listening to French disyllabic sequences (Christophe, Dupoux, Bertoncini, & Mehler, 1994), French infants dishabituated to the Spanish disyllabic sequence when the prosody switched from word-internal to word/PhP-straddling, but the infants did not notice when word-internal or word/PhP-straddling tokens switched to a new token of the same type. While some studies have suggested that experience plays a role in shaping infant perception and processing of prosody (Fernald & McRoberts, 1996; Friederici et al., 2007; Jusczyk, 1989), Christophe and colleague's (2001) results support the general acoustic nature of prosodic cues, at least for very young listeners.

(33) a) ...gelatína...

b) ...gorí la] [$_{PhP}$ tísico...

Very little work has been done to look at unfamiliar prosody as a cue to grouping. However, Johnson and Seidl's (2008) results (see Section 2.2.2) provide further evidence that infants are highly tuned to the prosodic features of their native language. They found that Dutch infants are unable to recognize strings of words that form a well-formed prosodic grouping at familiarization when they later occur in fluent speech *unless* pause cues are available at the boundaries. This is in contrast to Seidl's (2007) study, which found that American English-acquiring infants do not require pauses for the same task. This may because English has a wider pitch range than Dutch (Collins & Mees, 1981) and particularly exaggerated pitch cues in IDS (Fernald et al., 1989), leading to increased reliance on pitch cues in American infants. Regardless, it suggests that experience with a given language influences the sorts of prosodic cues learners require for prosody-driven learning.

Johnson and Seidl also conducted a cross-linguistic version of the experiment, familiarizing English-acquiring infants with Dutch passages that were prosodically well-formed or ill-formed and familiarizing Dutch-acquiring infants with English passages that were well-formed or ill-formed. At test, neither the Dutch nor the English infants showed a listening preference for passages of fluent speech containing the well-formed or ill-formed words (these studies are unpublished, but are mentioned in Johnson & Seidl, 2005; 2008). Thus, despite evidence that prosody is salient to other species and to human newborns, non-native prosody does not cause the same memory effects as native prosody. However, it is important to note that null results from a single experimental paradigm hardly resolve the question of whether infants can use non-native, unfamiliar prosody for bootstrapping into syntax.

Overall, the evidence to date does not give a clear picture of the role of experience in shaping how learners perceive and use the prosodic correlates of syntactic boundaries over the first year of life. Studies suggest that infants' perception of prosody, including prosodic boundary cues, is influenced by their native language very early on (Friederici et al., 2007; Fernald & McRoberts, 1996; Jusczyk, 1989) but also that prosodic boundaries are perceivable without prior experience (Christophe et al., 2001; Nazzi et al., 1998). These two findings are not necessarily incompatible. Rather, it appears that prosody is acoustically salient to a naïve listener, but that, with experience, infants learn to filter out prosodic information that is not relevant in their native language. Indeed, the studies reviewed here that found effects with non-native prosody have all involved infants 4 months or younger (Jusczyk, 1989; Nazzi et al., 1998). Given that the experiments discussed here involved quite different tasks and measures, however, it is difficult to draw firm conclusions.

2.4.2 Adult Studies

Evidence from adults shows a similarly complex relationship between experience and the perception of prosody. In a precursor to the pause-insertion studies used by Hirsh-Pasek and her colleagues (1987), Pilon (1981) familiarized American adults with Korean sentences for 27 minutes, then tested them on those sentences with artificial pauses inserted at the subject-predicate boundary, within a constituent, or within a word. Participants rated sentences in which the pause occurred at the subject-predicate boundary as more natural than sentences with pauses inside a constituent or word, suggesting that adults are sensitive to the prosodic correlates of the subject-predicate boundary, at least in one non-native language. Similar results were found by Wakefield, Doughtie and Yom (1974) and, at the clause level, with Polish stimuli by Jusczyk (1989). Thus, adults are at least sometimes able to perceive the prosodic correlates of the subject-predicate boundary in some unfamiliar languages.

On the other hand, there is also evidence that native language influences adults' perception of prosodic cues. Lehiste and Fox (1992; 1993) conducted a series of experiments looking at the perception of duration and intensity as cues to syllable prominence in Estonian, Swedish, and English listeners. Participants listened to a short sequence in which a syllable was repeated four times, with one of the syllables longer in duration or higher in amplitude than the others. In some conditions, a second syllable was also marked with a prominence cue, so that duration and amplitude cues were pitted against each other. For speakers of all three languages, duration and amplitude both influenced perceptions of syllable prominence, but the particular weighting that listeners gave the two cues differed. English listeners relied more heavily on amplitude cues, while Estonian and Swedish listeners focused on duration cues, perhaps because vowel length is phonemic in Estonian and Swedish. This suggests that, though adults are able to perceive at least some non-native prosodic boundary cues, they also interpret prosody relative to language-specific prosodic expectations.

Two recent studies have looked at adult use of non-native prosody in learning an artificial language. Recall that Shukla and colleagues (2007) found that Italian adults were able to use Italian prosodic boundary cues to constrain the domain over which they looked for words – evidence that native prosody can be used to group the speech stream into substrings that are then further analyzed for word boundaries (see Section 2.2.3 for a more detailed description). In a final experiment, they used Japanese-based prosodic features and found that Japanese, like Italian, IP boundaries serve to delineate groups of syllables over which Italian adults look for words. While the prosody was synthesized and simplified, this suggests that adults are willing to entertain unfamiliar manifestations of prosody in a grammar learning task.

Langus et al. (2012) also tested Italian adults on their ability to use unfamiliar prosody. In their first three experiments (see Section 2.2.3), they established that adult

Italian speakers could use Italian prosody to group speech into PhPs and IPs. PhPs were marked with final-syllable lengthening, and two PhPs grouped together to form an IP, which was marked with a falling intonational contour. They argue that these experiments are evidence that participants can simultaneously track two levels of prosodic cues to learn a hierarchical grammar (although see Section 2.2.3 for an alternative explanation). In a final experiment, the authors reversed the prosodic cues: PhPs were marked with pitch declination, and IPs were marked with final-syllable lengthening, which is contrary to typical Italian prosody. Despite the unusual prosodic contour, Italian adults still showed evidence of grouping the speech stream based on prosody at both the phrase (PhP) and sentence (IP) level. This experiment suffers from the same flaw discussed in Section 2.2.3 – namely, that it is difficult to interpret the results at the sentence level – but, as with Shukla et al. (2007), it does suggest that adults are able to use unfamiliar prosody for grouping the speech stream into smaller units.

Compared to the infant evidence reviewed above, the role of language-specific experience in the perception and use of non-native prosody in adults is a bit clearer. While the listener's native language does influence perception of certain prosodic cues (Lehiste & Fox, 1992; 1993), adults have, overall, demonstrated flexibility in their perception (Pilon, 1982) and use (Langus et al., 2012; Shukla et al., 2007) of unfamiliar prosody. However, both Langus et al. (2012) and Shukla et al. (2007) used simplified, synthesized prosody in their experiments, so it is unknown whether this ability extends to learning tasks that involve natural prosody from an unfamiliar language. This is addressed by the experiments in Chapter 4.

2.5 Situating the Present Work in Context

The literature offers evidence that syntax and prosody are reliably correlated at the clause level for both utterance-internal and utterance-final clauses in IDS and that infants are sensitive to the prosodic correlates of clauses and certain other phrases. However, the prosodic bootstrapping hypothesis argues that children not only perceive, but actually use, these cues to make inferences about grouping, constituents, or phrase structure. Previous studies provide evidence that prosody affects how infants remember and process incoming acoustic stimuli. These studies show that units that are grouped together prosodically are more likely to be remembered together, but they do not speak to whether the learners treat prosodically-grouped words as cohesive syntactic constituents. The experiments presented in Chapter 3 are an initial step towards determining whether infants and adults can use prosody for learning about syntactic constituency. This question is addressed by testing whether infant and adult learners treat prosodically-grouped words as cohesive, moveable, constituent-like units.

Chapter 4 explores a second question: is prosody useful in syntax acquisition because it is acoustically salient and perceivable without prior language-specific knowledge or is it useful only after the learner has deduced information about the interactions between prosodic and syntactic boundaries in the target language? Previous work provides little information that bears on this question, though cross-linguistic evidence suggests both that prosody is acoustically salient and that native-language experience shapes prosodic perception before the end of the first year of life. This dissertation addresses this issue by testing infants and adults on whether they treat prosodically-grouped words as constituent-like groupings when the groupings are marked with unfamiliar, non-native prosody.

CHAPTER 3

THE PROSODIC BOOTSTRAPPING OF CLAUSES

This chapter probes the constituents version of the prosodic bootstrapping hypothesis through two experiments asking whether learners can use prosody to locate groups of words that they treat like syntactic constituents. Experiment 1 tests 19-month-old infants on their ability to use prosodic cues to locate clause-like units in an artificial grammar and then to recognize those clauses when they have been moved to a new position in the utterance. Experiment 2 investigates the same questions with adult participants. In both experiments, participants were familiarized with sentences with 1-clause (ABCDEF) or 2-clause (ABC, DEF) prosody and were then tested on items with 2-clause prosody of the forms DEF, ABC and EFA, BCD. The DEF, ABC items were consistent with the clauses from the 2-clause familiarization, while the EFA, BCD items were not. If participants from the 2-clause groups use prosody to locate clausal constituents, they should differentiate between the two types of test items (for the infants) or recognize the consistent test items as being better examples of the familiarization language (for the adults). Infants in the 1-clause group are not expected to discriminate between the consistent and inconsistent test items, since their familiarization sentences contained no internal cues to clause boundaries.

In these experiments, learning of syntactic constituency is inferred based on the participant's ability to recognize prosodically-grouped words when they have moved within the utterance. Syntactic movement is a traditional test of constituency: words that move together belong to the same constituent at some level of the syntactic hierarchy (e.g., Carnie, 2002, p. 52). Adjunct clauses, on which the stimuli sentences were modeled, can be reordered within an utterance, as seen in (1) and (2) (clauses are indicated with brackets).

(1) [While I ate], [John came home]

[John came home], [while I ate]

(2) [If he likes overripe bananas], [Greg will like banana bread]

[Greg will like banana bread], [if he likes overripe bananas]

Several studies have used movement as a test for learning of constituency or even phrase structure in adults, including Morgan and Newport (1981), Morgan, Meier, and Newport (1987; 1989), and Takahashi and Lidz (2007). Since only words that are in the same constituent can 'move' together in natural languages, a learner who applies this (innate or learned) knowledge about the behavior of syntactic constituents to the experimental grammar is going beyond merely grouping the speech stream into units – she is applying a syntactic rule to those grouped units. At the very least, she is treating prosodically-grouped strings of words as more cohesive than strings of words that contain a prosodic break. Cohesiveness is an identifying property of within-constituent words, as seen through the movement of constituents, pro-form replacement of constituents, etc. Therefore, if a learner recognizes prosodically-grouped words when they have moved within a sentence, we can infer that she is treating the grouping like a cohesive syntactic constituent.[2]

[2] Throughout this dissertation I refer to clauses as "moving." This reflects a terminological choice, not an assumption of transformational grammar. Whether clauses move or simply appear in distinct locations within a sentence is not crucial for the interpretation of the results that follow.

A second and equally important aspect of the movement test lies in the acoustics: the prosodic contour of a moved constituent is different from when that constituent appears in its original position. In the 2-clause familiarization sentences, the clause ABC has initial clause prosody, while it has final clause prosody in the consistent test items. Therefore, if the learner can recognize prosodically-grouped words once they have moved, we can infer that she has generalized beyond the input. She is not just recognizing an acoustic chunk that she has heard before; she has acquired a representation of the structure of the experimental grammar that goes beyond what can be directly observed in the familiarization items.

3.1 Experiment 1: Infants

3.1.1 Participants

Twenty-four English-acquiring 19-month-olds from the Tucson area were tested (16 male, 9 female). The mean age was 580 days, with a range of 557 to 610 days. The participants were all of normal birth weight (at least 5lb, 7oz) and term (at least 37 weeks) and had no more than ten hours a week of foreign language exposure. Data from an additional eleven infants were not included due to equipment problems or experimenter error (n = 3) and fussiness (n = 8). Data from a final participant were excluded because her overall mean looking time was more than 2.5 standard deviations above the mean.

The 24 infants included in the final sample were tested by Experimenters 1 and 2. A third experimenter tested six infants (four of whom successfully completed the experiment, one of whom fell asleep, and one for whom there was an equipment malfunction). During the course of the experiment, it quickly became apparent that Experimenter 3 was obtaining aberrant data – the participants she tested had a mean overall looking time of almost twice that of the participants tested by Experimenters 1 and 2 (6.75 seconds for Experimenters 1 and 2 versus 12.25 seconds for Experimenter 3). This is reflected statistically: there is a significant overall effect of experimenter ($\chi^2(2) = 11.08$, $p = .0039$)[3], with a significant difference between Experimenters 1 and 3 $\chi^2(1) = 12.43$, $p < .001$) and a nearly significant difference between Experimenters 2 and 3 ($\chi^2(1) = 3.79$, $p = .052$) (the lack of significance at the .05 level is likely due to the small number of data points – Experimenter 2 had six infants included in the sample and Experimenter 3 had four). The experimenter effect disappears when Experimenter 3 is excluded ($\chi^2(1) = 1.03$, $p = .31$). Since Experimenter 3 was new to the experiment and to running the Head-Turn Preference Procedure, we terminated her involvement in the experiment and excluded all data from the infants she tested.

Although 17-month-olds are old enough to learn a simple phrase structure grammar based on brief exposure to a novel language (Gerken, Wilson, & Lewis, 2005) and to comprehend sentences that contain movement (see Seidl & Hollich, 2002), pilot testing with 17-month-olds found that they struggled with the task. For these reasons, a slightly older age group was used.

3.1.2 Materials

The stimuli consisted of 6-word sentences, with either 1-clause (ABCDEF) or 2-

[3] See Section 3.1.6 for a description of the statistical model.

clause (e.g., 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." The prosodic features of the stimuli are considered in detail in Section 3.1.2.1.

There were two words from each class (A, B, etc), as seen in Table 3.1. The words were all monosyllabic nonsense CVC sequences, and they were selected to include a variety of consonants and vowels. Each of the twelve words was pseudo-randomly assigned to one of the six classes.

Word Class	**A**	**B**	**C**	**D**	**E**	**F**
	bʌp	dɪv	kæg	fɛb	hɑs	dʒɪk
	nɪm	pɛl	ɹʌd	tɑz	zæf	vɑt

TABLE 3.1 Stimuli words (in IPA)

For both familiarization conditions, the transitional probability from each word class to the next was 1, since the word classes always appeared in the same order (e.g., B words were always preceded by an A word; p(B|A) = 1). The probability of each transition (e.g., from a particular token of class A to a particular token of class B) was .5. For example, p(dɪv|bʌp) = .5 and p(pɛl|bʌp) = .5. Thus, transitional probabilities between word classes and between particular word tokens were equal across the two familiarization conditions.

It may be the case that the infants also track transitions between words and prosodic boundaries. Word A is both utterance- and clause-initial and word F is both utterance- and clause-final for both types of familiarization stimuli. On the other hand, C and D occur at clause boundaries only for the 2-clause items. Therefore, transitional probability between words and prosodic boundaries was equal across the two familiarization conditions, except for the probability of a particular word class appearing at the internal clause boundary, which necessarily differed between the 1-clause and 2-clause conditions.

Before moving on, it is important to note that there are several converging lines of evidence suggesting that these sentences will be interpreted as containing six monosyllabic words, not three di-syllabic or two tri-syllabic words. First, since the transitional probabilities between each pair of syllables are equal, the statistics of the stimuli are most consistent with parsing the sentences into six monosyllabic words. Second, each monosyllabic word contains a full stressed vowel, and equally stressed syllables do not appear next to each other within a word in English (Hayes, 1995). It has even been proposed that placing a word boundary before each stressed syllable is an effective technique for word segmentation in English (Cutler & Norris, 1988). Finally, corpus analysis of infant-directed English found that there is an average of 1.17 syllables per word (Gambell & Yang, 2006), suggesting that infants should not find a string of six monosyllabic words to be unnatural.

It is also important to note that multi-clause utterances are sufficiently common in the infants' input that we can reasonably expect them to be familiar with utterances that are comparably complex to the stimuli sentences. Soderstrom, Blossom, Foygel, and Morgan (2008) analyzed the speech of two mothers interacting with their infants. Between 6 and 15% of the mothers' utterances contained more than one clause, even though the infants were only 6-10 months old at the time of the study. A typical 19-month-old, therefore, should be quite familiar with multi-clause utterances.

The sentences were read by a young adult native American-English speaking female from the Midwest, who was naïve to the purpose of the experiment. The speaker was asked to produce the 2-clause items as if she were saying “While I ate, John came home.” For the 1-clause utterances, the speaker was asked to think of “I ate at home with John”; the speaker naturally produced the 1-clause utterances with slight contrastive intonation on the fourth word (corresponding to “I ate at **home** with John”). The speaker produced the stimuli in a natural and happy-sounding way, using IDS. The 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, pause length, and amplitude, so that the sentences retained a natural amount of variability without any of the sentences sounding distractingly different. The tempo adjustments were made by measuring the duration of each sentence and calculating the mean and standard deviation for each stimuli type separately (1-clause familiarization items, 2-clause familiarization items, consistent test items, and inconsistent test items). Any sentence whose duration was more than one standard deviation from the mean for that type of item was compressed or stretched to the mean duration for that stimuli type using the “Change Tempo” feature of Audacity (Mazzoni, 2009).

The inconsistent items averaged 2240ms, with a standard deviation of 110ms. This is comparable to the 2-clause familiarization items, which had a mean of 2230ms, with a standard deviation of 100ms. The consistent test items, however, had a mean of 2070ms, with a standard deviation of 100ms. The difference in duration between the consistent and inconsistent test items is statistically significant ($t(31) = 7.20$, $p < .0001$). It is possible that measurement error plays a role: it is impossible to accurately measure the onset of a sentence-initial stop closure, for example, or to determine the end of an unreleased sentence-final stop. However, neither of these measurement-related factors appears to account for the difference. The difference may be due to patterns resulting from combinations of particular consonants in particular prosodic positions; for instance, the speaker tended to produce a shorter pause (by 40ms) at the clause boundary if the second clause started with *nɪm*, which only occurred in the consistent test items. Since the source of this durational difference is unclear, the consistent items were all stretched by 80ms and the inconsistent items were all compressed by 80ms to control for any potentially confounding effects of durational differences. After this manipulation, the two types of test sentences had the same mean value.

Pause length adjustments were made using Praat in much the same way. The pause length was measured for all three types of 2-clause stimuli (the 2-clause familiarization items and both types of test items; the 1-clause stimuli did not have a pause), the mean and standard deviation were calculated, and outliers beyond a standard deviation were changed to the mean value.

Amplitude adjustments were made by ear using the “Amplify” feature of Audacity, so that all of the stimuli were approximately the same volume. In a few cases, a particularly loud plosive burst was reduced in amplitude to avoid a noticeable popping sound.

After the stimuli were standardized, they were played to several English-speaking linguists naïve to the purposes of the experiment, who all agreed that the stimuli sounded natural and that it was impossible to tell which sentences had been altered.

The stimuli sentences were arranged into lists (as described in Section 3.1.3, below). In all of the lists, the sentences were separated by a 1-second pause.

3.1.2.1 Acoustic Analysis of the Prosodic Features

As discussed in Section 2.1.3, clause boundaries are typically marked with a pause, final-syllable lengthening, and pitch resets. In order to verify that the expected prosodic markers occur in the experimental stimuli, five sentences each were analyzed from the 1-clause and 2-clause items. The items were matched, such that the same strings of words were analyzed for each type of prosody. Therefore, unless otherwise noted, the variables discussed below were all tested for significance using paired-samples t-tests. The measurements were all made by examining the waveforms and spectrograms in Praat.

The 2-clause sentences were an average of 2134ms in duration, while the 1-clause sentences lasted an average of 1773ms. This 361ms difference is statistically significant ($t(4) = 12.05$, $p < .001$), and is partially accounted for by the pause that occurred between the clauses (i.e., between words C and D) in the 2-clause items. The average inter-clause pause length was 173ms. There were no measurable pauses in any of the 1-clause sentences. Thus, pausing is a reliable indicator for clause-like constituents in the experimental stimuli.

In order to test for final-syllable lengthening, each word from each sentence was analyzed for vowel duration and t-tests were conducted, comparing each word class across condition. Complete results are presented in Table 3.2.

Word Class	2-Clause Mean	1-Clause Mean	t-value (df = 4)	Significant?
A (bʌp, nɪm)	80	76	0.50	no
B (dɪv, pɛl)	104	88	1.89	no
C (kæg, ɹʌd)	207	82	9.70	p < .001
pause	173	no pause		n/a
D (fɛb, tɑz)	122	122	0.045	no
E (hɑs, zæf)	104	121	2.74	no
F (ʤɪk, vɑt)	146	135	0.46	no

TABLE 3.2 Vowel duration measurements (in milliseconds)

The vowel duration for word class C was significantly longer for the 2-clause items (where word C is clause-final) than for the 1-clause items (where it is clause-internal). The other word classes do not significantly differ in duration across familiarization condition, because words A, B, D, and E are clause-internal, and word F is clause-final, across both types of familiarization items.

The durations of the vowels in clause-final and non-clause-final words were also compared (A, B, D, and E versus C and F for the 2-clause items; A, B, C, D, and E versus F for the 1-clause items). In an unpaired t-test, clause-final vowel duration was significantly longer than the non-clause-final vowel duration ($t(16.57) = 4.57$, $p < .001$)[4]. Therefore, final-syllable lengthening is a reliable indicator of both intermediate and sentence-final clause boundaries in the experimental stimuli.

[4] The non-integer degrees of freedom are due to the fact that sample sizes were unequal. The Welch approximation to degrees of freedom was used.

Finally, the stimuli were also examined to determine whether intonational cues were consistent with the intended interpretation. Intermediate clauses typically end with a fall-rise intonational contour, often called a continuation rise, followed by a drop in pitch across the clause boundary (Ladd, 2009). A continuation rise can clearly be seen in word C of the 2-clause example in Figure 3.2, and the pitch drops for word D at the beginning of the second clause. On the other hand, the pitch in the 1-clause sentences rises from word B through a peak at word D before falling off at the end of the sentence. This pattern is consistent with the presence of an internal clause boundary between words C and D for the 2-clause items only.

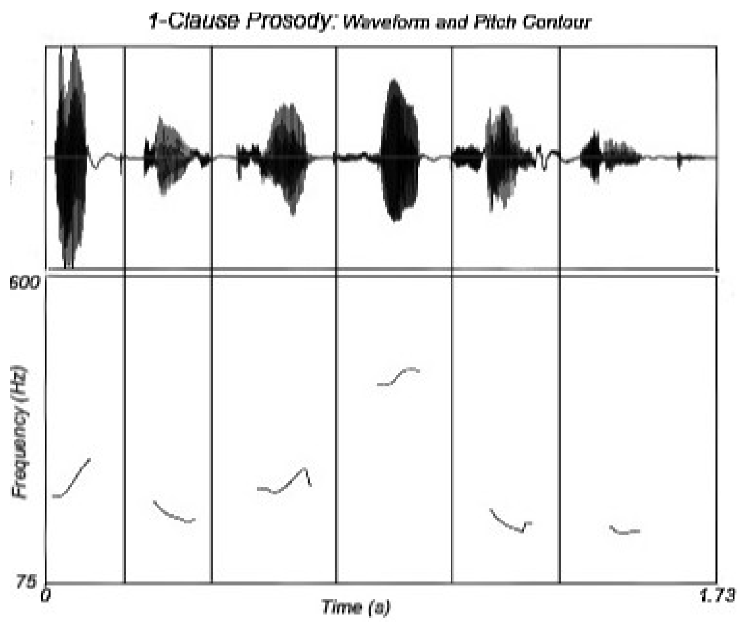

FIGURE 3.1 1-clause prosody. Example sentence: *bʌp dɪv kæg fɛb zæf ʤɪk*. Vertical lines represent word boundaries. Figure captured from Praat (Boersma & Weenink, 2008).

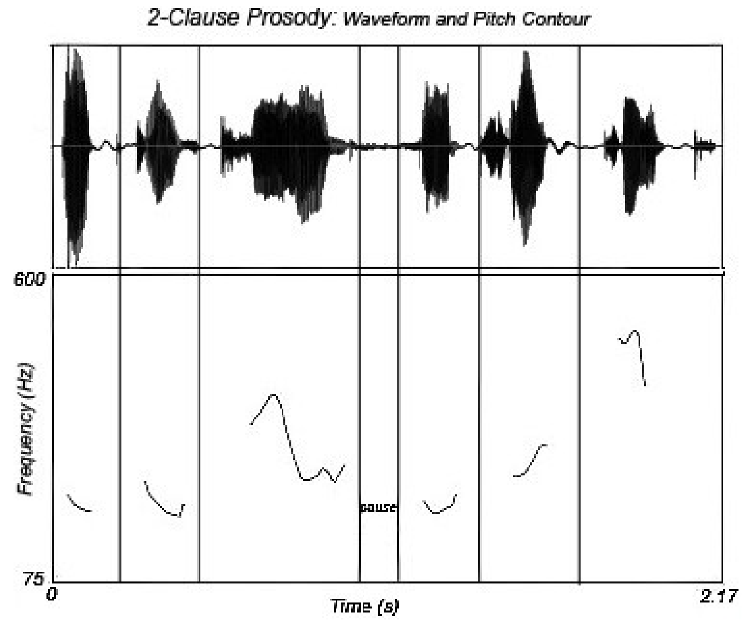

FIGURE 3.2 2-clause prosody. Example sentence: *bʌp dɪv kæg, fɛb zæf ʤɪk*. Figures captured from Praat (Boersma & Weenink, 2008).

Therefore, the experimental stimuli contain reliable prosodic cues to clauses. In the 2-clause utterances, the clause-like chunks are marked with a pause between words C and D and with clause-final-syllable lengthening of word C. The 2-clause sentences also have a fall-rise contour on word C, followed by a characteristic pitch drop at the internal clause boundary between words C and D. If infants are able to use prosody to locate clause-like constituents in general, they should be able to do so with these stimuli.

3.1.3 Design

The infants were divided into two groups, based on the prosody of the items they heard during the familiarization phase. For those in the 1-clause group, the familiarization items were of the form ABCDEF (i.e., "I ate at home with John"), and for those in the 2-clause group, the familiarization items were of the form ABC, DEF (i.e., "While I ate, John came home"). The infants from both groups heard the familiarization sentences in the same order.

During pilot testing, 17-month-old participants displayed evidence of learning only during the last block of test trials. Since they had only two minutes of familiarization, the infants in the present study were given two familiarization phrases: "long" and "2-minute." The initial "long" familiarization phase was conducted in a small room, where the infant, a parent, and an experimenter played quietly while familiarization stimuli played in the background.

The children in the 2-clause group heard a total of 192 sentences during the long familiarization over a period of 608 seconds. Since there were six word classes, with two tokens in each class, there were a total of 64 unique sentences with each kind of familiarization prosody. The 192 sentences represent each of the 64 2-clause familiarization sentences played three times in a pseudo-randomized order. The infants in the 1-clause group heard 218 sentences over the same time period (608 seconds). During the second "2-minute" familiarization phase, the infants were exposed to an additional two minutes of familiarization items (38 sentences in the 2-clause condition and 43 in the 1-clause condition). The 1-clause group heard more sentences during both familiarization phases because the 1-clause sentences were shorter than their 2-clause counterparts. Exposing the 1-clause infants to additional familiarization items gives them extra time to learn and works against the hypothesis that they will not discriminate between the consistent and inconsistent test items.

While familiarization prosody was a between-participants variable, test item type was a within-participants variable. All of the infants were tested on items with 2-clause prosody, of the forms DEF, ABC (e.g., *tɑz hɑs vɑt, bʌp dɪv ɹʌd*) and EFA, BCD (e.g., *hɑs vɑt bʌp, dɪv ɹʌd tɑz*). The consistent items (DEF, ABC) had the same clause-like constituents that the 2-clause infants heard during familiarization, although the clauses appeared in the opposite order and had a new prosodic contour (intermediate vs. final clause). Conversely, the inconsistent items (EFA, BCD) chunked the sentence into different, non-constituent units.

Both types of test items had the same transitional probability between the word classes, since one new transition was introduced (A → F) for both the consistent and inconsistent test items. Both types of test items also had new utterance-initial and utterance-final word types.

3.1.4 Task

After the initial long familiarization period, the infants were tested using a modified version of the Head-Turn Preference Procedure (e.g., Kemler Nelson et al., 1995). The infant was seated in a high chair in a soundproof booth, with the parent seated behind and to the side. The parent was given headphones to wear, so that he or 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-minute familiarization phase, the entire 2-minute 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. The experimenter watched the infant over a closed circuit television and recorded the direction of looking by entering it into a computer program.

During the test phase, the infants heard 12 test trials, divided into three blocks, each consisting of the same four trials (two consistent and two inconsistent 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 two seconds.

3.1.5 Hypotheses

If the infants in the 2-clause group treat the prosodically-marked clause-like strings as cohesive syntactic constituents, they should discriminate between the grammatical movement of the consistent test items and the ungrammatical reordering of the inconsistent test items. However, it is also possible that the infants in the 2-clause group would discriminate between the consistent and inconsistent items because the 2-clause familiarization sentences share more boundary words (utterance- and clause-initial and utterance-and clause-final words) with the consistent test items than the inconsistent test items. In particular, children have been shown to be biased to attend to the beginnings of sentences (Newport, Gleitman, & Gleitman, 1977). The sentence-initial words for the 2-clause familiarization items (class A in **A**BC, DEF) are clause-initial in the consistent test items (DEF, **A**BC), and the sentence-initial words for the consistent test items (class D in **D**EF, ABC) are clause-initial in the 2-clause familiarization items (ABC, **D**EF). In contrast, the inconsistent test items (EFA, BCD) have different clause-initial words (E and B). (The same argument could be made regarding utterance-final and clause-final words, which are also in perceptually-prominent positions.) Thus, discrimination between consistent and inconsistent test items by the 2-clause group could simply be due to recognition of perceptually-prominent boundary words.

The 1-clause familiarization group was included to control for this alternative explanation. The 1-clause familiarization items start with a word from class A and end with a word from class F (**A**BCDE**F**), so they share two boundary words with the consistent items (versus four boundary words shared between the 2-clause familiarization and consistent test items), while they do not share any boundary words with the inconsistent test items. If infants are merely attending to boundary words, those in the 1-clause group should also discriminate between the consistent and inconsistent test sentences (albeit perhaps to a smaller extent than the 2-clause group). On the other hand, if infants from the 1-clause group do not prefer one type of test item of the other, it would suggest that any discrimination seen in the 2-clause group is not driven by perceptually prominent boundary words, but instead reflects extraction of clause-like constituents from the speech stream.

3.1.6 Results

Each infant's looking time was recorded for each trial, and looking times shorter than two seconds were excluded from analysis, as is standard in the field. The mean overall looking time was 6.79 seconds. Nine of the 12 infants in the 2-clause condition and five of the 12 in the 1-clause condition listened longer to the consistent test items.

The data for this and subsequent experiments were analyzed using R (R Core Team, 2012) and the R packages *lme4* (Bates, Maechler, & Bolker, 2012), *languageR* (Baayen, 2011), and *LMERConvenienceFunctions* (Tremblay & Ransijn, 2012). Linear mixed effects modeling was used 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 (consistent versus inconsistent) was a within-participants variable. Since pilot testing revealed effects of test order, test block (1st, 2nd, 3rd) was also considered as a within-participants factor. Every effort was taken to include a maximal random effects structure (c.f. Barr, Levy, Scheepers, & Tily, under review). Test condition and test block were included as random effects grouped by subjects. However, given the small number of items, item was not used as a grouping factor for random effects.

Looking time was the dependent variable, and looking times were log transformed in order to improve the normality of the residuals. Likelihood ratio tests (LRTs) comparing a full model to a minimally-reduced model were used to assess the significance of the fixed effects, so results are reported based on the chi-squared (χ 2) statistic.

For experiment 1, there was a significant interaction of familiarization by test condition (χ 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 listening longer to the consistent test items. The effect of test condition was not significant for the 1-clause condition ($\chi^2(1) = .23$, $p = .63$).

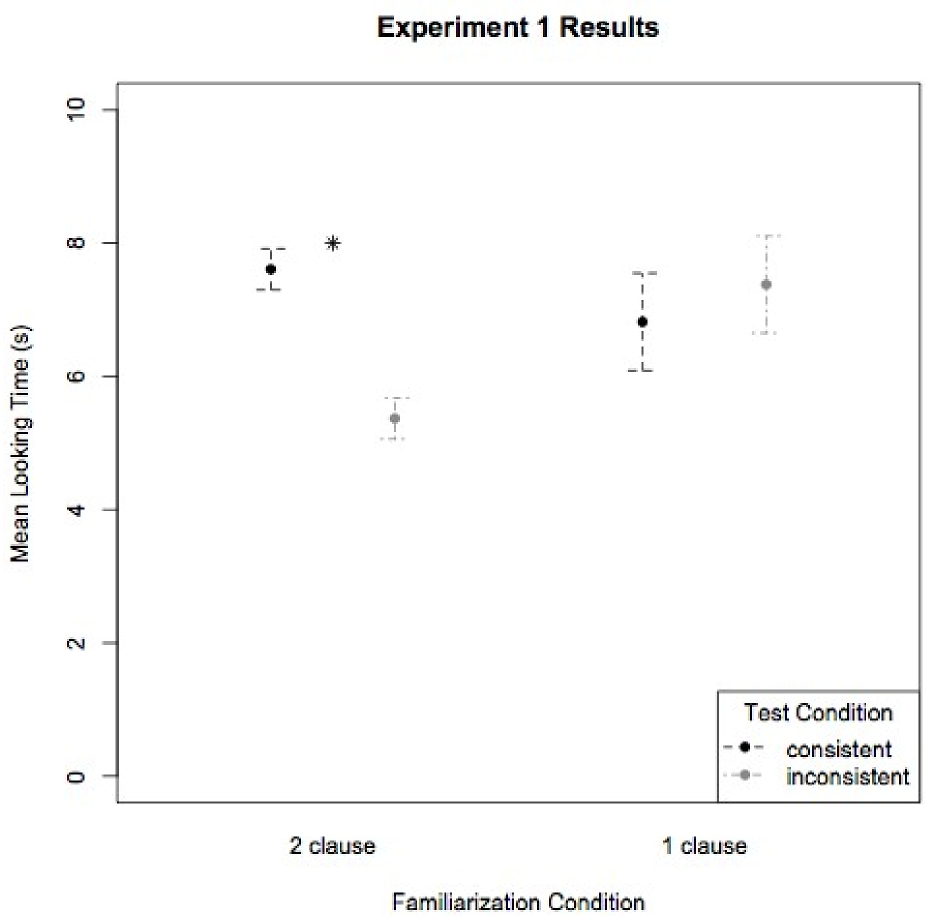

FIGURE 3.3 Error bars represent standard error of the mean difference. An asterisk represents a significant difference between the adjacent conditions. While the data were log transformed for the statistical tests reported above, this graph shows the raw data.

3.1.7 Discussion

This pattern of results supports the prosodic bootstrapping hypothesis. More specifically, it provides support for an intermediate version of prosodic bootstrapping in which infants can infer the syntactic constituency of clauses based on prosody. In previous studies (e.g., Nazzi, Kemler Nelson, Jusczyk, & Jusczyk, 2000; Soderstrom, Kemler Nelson, & Jusczyk, 2005) infants displayed enhanced memory for strings of words that were initially presented as well-formed prosodic sequences. The present experiment extends this prior work in two important ways.

First, unlike in previous studies, the infants in the present experiment had to generalize beyond the familiarization acoustics to recognize the moved constituents. In Nazzi et al. (2000), Experiment 1, the familiarization sequences (e.g., “Leafy vegetables taste so good.”) were excerpted directly from the longer test passage (e.g. “John doesn't know what rabbits eat. Leafy vegetables taste so good. They don't cost much either.”), so the infants did not have to generalize beyond identical acoustic tokens to recognize the familiarization sequence at test. In Experiments 2 and 3, the familiarization items were spliced from a different passage (e.g., “Guess what kind of leafy things taste so good? Leafy vegetables taste so good.”), but the well-formed items were complete sentences at both familiarization and test, so it is unlikely that the acoustic features were markedly different.

Soderstrom and her colleagues (2005) also did not use identical tokens for the familiarization and test items, but, like Nazzi et al. (2000, Experiments 2 and 3), the target well-formed (“clause-coincident”) sequences had complete sentence prosody in both familiarization and test, making it likely that the acoustic features were highly similar. Likewise, the ill-formed (“clause-straddling”) sequences introduced a sentence boundary between the same pair of words for both the familiarization and test items. In fact, the authors found that infants preferred to listen to test passages in which the prosodic features of the target sequence matched the familiarization prosody. For example, an infant who heard the sequence “rabbits eat leafy vegetables” as a well-formed prosodic contour during familiarization (“Rabbits eat leafy vegetables.”) would be more likely to recognize that sequence when it was presented as a well-formed sentence in the test passage. Likewise, an infant who heard the same string as a prosodically ill-formed grouping at familiarization (“...rabbits eat. Leafy vegetables...”) would listen longer to a passage that included the sequence as an ill-formed prosodic grouping at test.

In the present experiment, on the other hand, the individual clauses from the 2-clause familiarization items (ABC and DEF) had markedly different acoustics at test. The clause ABC, for example, had initial clause intonation during familiarization and final clause prosody at test. Figure 3.4 demonstrates this point. Given that the acoustics were different at familiarization and test, the fact that the infants in the 2-clause group recognized prosodically-grouped strings at test suggests that they were able to abstract away from the input acoustics. This type of abstraction is important for syntax acquisition, since it more closely mimics the child's normal input: infants are very rarely presented with identical acoustic tokens in their daily life (except for in frequently watched movies, etc.) and certain learning mechanisms, such as cross-sentential comparisons, require the learner to recognize a constituent when it appears in a slightly different context.

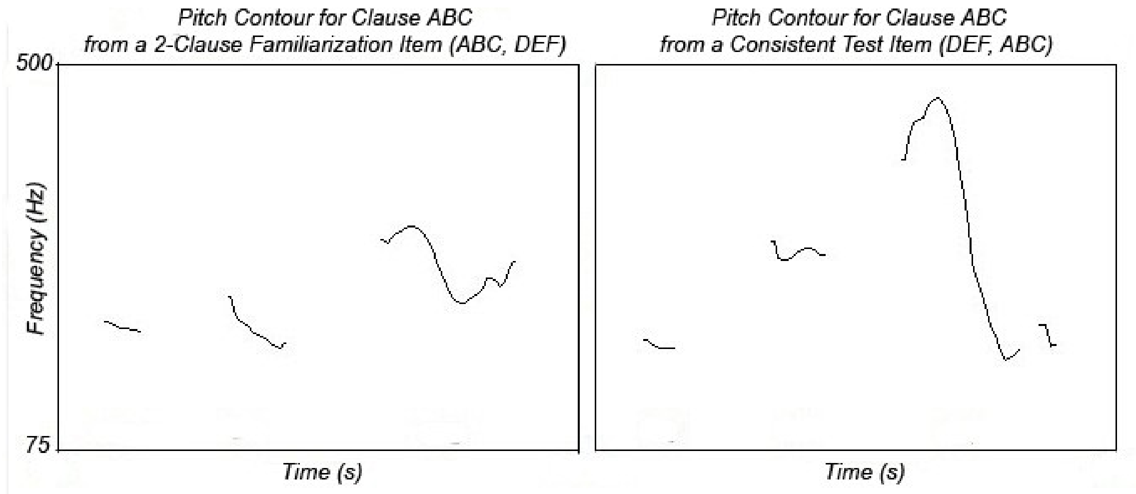

FIGURE 3.4 Example ABC clause: *bʌp dɪv kæg*. Figure captured from Praat (Boersma & Weenink, 2008).

Not only was the prosodic contour different at test, but the words themselves were variable, since each word class contained two tokens. Although sentences were not withheld during familiarization to specifically test for generalization at the word level, the particular combinations of A-B-C-D-E-F used at test were only heard three times each during the entire twelve minutes of familiarization, and were not heard at all during the 2-minute familiarization period that immediately preceded testing. There were 64 unique ABC(,)DEF combinations, making it unlikely that the infants memorized each familiarization sentence. Instead, it is probable that participants were acquiring the underlying structure of the sentence as well as the clausal groupings; however, the former type of generalization was not directly tested.

The second way in which the present experiment goes beyond prior work is that the results speak more directly to prosody as a cue for syntax acquisition. While remembering groups of words (Nazzi et al., 2000) and extracting sentences from the speech stream (Soderstrom et al., 2005) are important aspects of language processing, these previous studies only speak to the most basic form of prosodic bootstrapping: grouping the speech stream into units for further analysis. The current study, on the other hand, suggests that infants can use prosody for the more syntactically-sophisticated task of finding clausal constituents. The infants not only used prosody to extract clause-like units from the speech stream, they then recognized them when they moved to a new part of the sentence. This suggests that infants treat prosodically-grouped words as cohesive, moveable units. Since constituents can be defined as cohesive groups of words, and since movement is one way in which constituents demonstrate this cohesiveness, we can infer that the infants in the 2-clause group were learning something akin to syntactic constituency.

Section 3.1.5, above, raised the possibility that discrimination in the 2-clause group could be due to the infants' recognition of perceptually-prominent boundary words, and not an effect of syntactic learning. However, the lack of discrimination seen in the 1-clause group undermines this alternative explanation. Recall that the 1-clause familiarization items share two boundary words (A and F) with the consistent test items and no boundary words with the inconsistent test items. While the infants in the 2-clause condition listened an average of 2.24 seconds longer to the consistent test items, the infants in the 1-clause condition listened a non-significant average of .55 seconds longer to the *inconsistent* items. Therefore, there is no reason to suppose that infants in the 2-clause group are relying on lower-level heuristics based on perceptually-prominent syllables to recognize groups of words that constitute grammatical units across different syntactic and acoustic contexts.

A second alternative explanation is that the prosodic boundary induced grouping by the 2-clause participants, but that the participants did not realize that the prosodically-grouped words were constituents. Recall that Gout et al. (2004) found that 10- to 13-month-olds were more likely to turn their head in response to a disyllabic sequence at test if it originally occurred within a PhP (vs straddling a PhP boundary) during familiarization; Shukla, Nespor, & Mehler (2007) found comparable results with adults (see Sections 2.2.2 and 2.2.3 for more details). These studies suggest that prosodic phrasing constrains the domain over which infants and adults search for words and are evidence for the grouping version of the prosodic bootstrapping hypothesis. As in these previous studies, it may be the case that infants in the 2-clause condition simply used prosody to delimit the domain over which they learned sequences of legal syllables in the familiarization grammar. This is conceptually the same as saying that the 2-clause group learned that the longer finite-state grammar (ABCDEF) was composed of two mini-grammars (ABC and DEF), but not that the mini-grammars were in a meaningful relationship to each other. The present results are consistent with this grouping interpretation, which is discussed more fully in Section 5.1.1. The question then comes down to whether or not the infants recognized that the clauses were independently-manipulable, cohesive units that combined to form a longer structure.

It is not possible to directly address this question from the present experiment. However, it is likely that the infants were aware that the clauses grouped together to form sentences. Infants display sophisticated knowledge of sentential prosody from an early age (see Snow & Balog, 2002 for an excellent discussion). Many infants go through a jargon intonation stage at around 12-14 months, in which they babble with adult-like intonational contours (e.g., Elbers, 1982), and infants even display knowledge of the intonation patterns of their ambient language (Halle, de Boysson-Bardies, & Vihman, 1991). Additionally, multi-clause sentences are common in input to infants during the first year of life (Soderstrom et al., 2008). It is therefore probable that the infants in the present experiment were familiar with sentential prosody, even in multi-clause utterances. Thus, they likely recognized the sentences as such, and the preference of the 2-clause group for the consistent test items suggests that they used prosody to break the sentences into groups of words that they treated like moveable syntactic constituents.

At the very least, then, the experiment presented here provides the first evidence that infants can use prosody not only to group speech, but also to infer that prosodically-grouped chunks do not have to occur in the same relation to each other within a sentence and that the chunks can occur with different prosodic features. Whether this implies that the learner recognizes that the chunks are syntactic constituents, the ability to use prosody to locate cohesive, independent chunks of speech certainly represents a major step towards learning of constituency. In Experiment 2, I test adult learners on a similar task to investigate the extent to which adult learners can make use of prosodic information in syntax acquisition.

3.2 Experiment 2: Adults

3.2.1 Participants

Forty-eight undergraduates enrolled in linguistics classes at the University of Arizona (20 male and 28 female) participated in the study for extra credit. All participants were native monolingual English speakers with normal or corrected-to-normal vision and hearing. An additional six participants were excluded due to experimenter or equipment error (n = 2) and because their mean grammaticality rating scores were more than two standard deviations from the overall mean (n = 4).

3.2.2 Materials, design, and task

The experiment was designed to be as similar as possible to Experiment 1. Half of the adults were familiarized on sentences with 1-clause prosody and half on sentences with 2-clause prosody. Participants were told that they would be listening to sentences from a language called "Labese" and that they would later be asked to make judgments about new sentences. The participants then listened to the entire 1-clause or 2-clause 10-minute "long" familiarization audio file from Experiment 1. During this familiarization period, the participants were given white paper and colored pencils and they were told that they could doodle on the paper as they listened, but that they should not take notes about what they were hearing. Participants were asked to doodle so that they would remain attentive throughout the experiment and to mimic the quiet play phase of the infant experiment.

After the familiarization phase, the participants were tested on individual consistent and inconsistent test sentences. The test sentences were drawn from the set of test sentences that were played to the infants in Experiment 1. While a test trial for the infants in Experiment 1 involved consistent or inconsistent items played continuously until the infant's attention waned, a test trial for the adults in Experiment 2 consisted of a single test sentence. After each item, the participant was instructed to rate whether the sentence was a good sentence of Labese on a scale of 1-4. A rating of 1 corresponded to "not a possible sentence in Labese" and a rating of 4 corresponded to "a good sentence in Labese." The entire experiment was run using SuperLab Pro 4 experimental software.

3.2.3 Hypotheses

If adults, like infants, are able to use prosody to locate clause-like constituents in the speech stream, participants in the 2-clause group should treat the prosodically-marked clauses as cohesive syntactic constituents, and they should rate the consistent test items significantly higher than the inconsistent items. Participants in the 1-clause group, on the other hand, should rate both the consistent and inconsistent items as equally grammatical sentences of Labese, since their input stimuli did not contain internal prosodic cues to clausal constituent boundaries.

On the other hand, unlike the infants in Experiment 1, it may be that adults rely on heavily or exclusively on boundary words when making their judgments. If this is the case, the adults in both conditions should rate the consistent test items as more grammatical: word classes A and F are clause-initial and clause-final in both familiarization conditions, as well as in the consistent test items. Thus, if adults are basing their judgments on boundary words, both the 1-clause and 2-clause groups should prefer the consistent test items. The 2-clause group may prefer the consistent items to a larger degree, since words C and D are also clause-final and clause-initial (respectively) in both the 2-clause familiarization and consistent test items.

3.2.4 Results

The data were analyzed using linear mixed effects modeling, with familiarization condition (1-clause versus 2-clause) as a between-participants variable and test condition (consistent versus inconsistent) as a within-participants variable. Grammaticality rating scores were the dependent variable. As discussed in Section 3.1.6, the random effects structure was maximally-specified. Test condition was included as a random effect grouped by subject. Test condition was also included as a random effect by items, since each sentence occurred as a consistent item (i.e., with a grammatical movement) for half of participants and as an inconsistent item (with ungrammatical movement) for the other half of participants.

There was a significant familiarization condition by test condition interaction ($\chi^2(1) = 17.66$, $p < .001$), as well as a significant main effect of test condition ($\chi^2(1) = 31.04$, $p < .001$), with participants rating the consistent items as more grammatical than the inconsistent items (2.93 versus 2.31 mean grammaticality rating score).

In order to understand this pattern of results, subsequent tests were conducted to look at the effect of test condition for each familiarization condition individually. There was a significant effect of test condition for both the 2-clause condition ($\chi^2(1) = 26.74$, $p < .001$) and the 1-clause condition ($\chi^2(1) = 11.90$, $p < .001$).

While participants in both conditions rated the consistent test items as significantly more grammatical than the inconsistent test items, the significant overall interaction effect indicates that the participants from the 2-clause group differentiated significantly more between the consistent and inconsistent test items. This can be clearly seen in Figure 3.5.

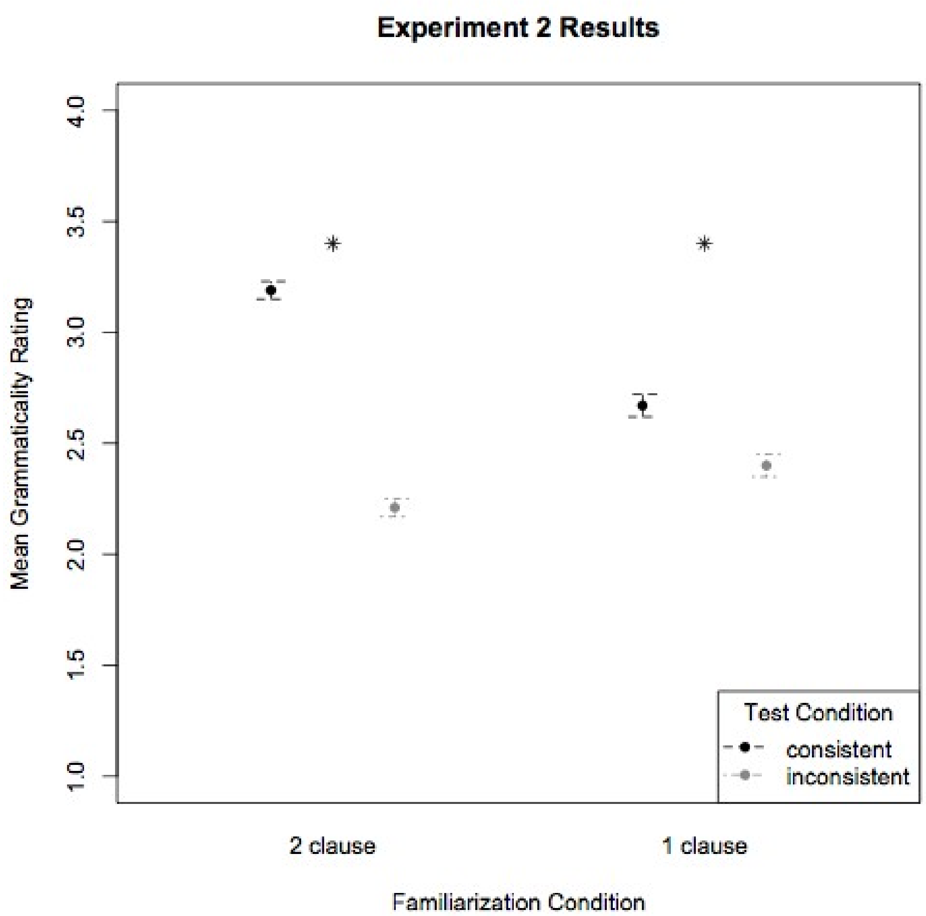

FIGURE 3.5 Error bars represent standard error of the mean difference. An asterisk represents a significant difference between the adjacent conditions.

3.2.5 Discussion

Both the 1-clause and 2-clause groups rated the consistent items as significantly more grammatical than the inconsistent items. However, there are two important effects to consider in these data: the significant effect in the 1-clause group and the significant familiarization by test condition interaction.

Since the 1-clause familiarization sentences did not contain prosodic cues to an internal clause boundary, the significantly higher grammaticality ratings for the consistent items support the hypothesis that the 1-clause group, at least, relied on boundary words for making judgments. This is in contrast to the infants in the 1-clause group who showed no hint of discriminating between the consistent and inconsistent test items.

On the other hand, the results from the 2-clause group are more difficult to interpret. The significant familiarization condition by test condition interaction indicates that the mean difference score between consistent and inconsistent items was significantly greater for the 2-clause group than the 1-clause group. It may be the case that the enhanced performance of the 2-clause group is simply due to the fact that the 2-clause familiarization items share more clause-initial and/or clause-final items with the consistent test items than the 1-clause familiarization items do. Alternatively, the adults in the 2-clause group may be behaving like the infants in Experiment 1 and using prosodic information to extract groups of words that they treat like cohesive, moveable constituent-like units. This question will be further explored in Section 5.1.

3.3 Interim Conclusions

In Experiments 1 and 2, 19-month-olds and adults were tested on their ability to use prosody to locate clause-like constituents in an artificial grammar. The infant data provide the first support for the constituents version of the prosodic bootstrapping hypothesis, suggesting that infants can use prosodic cues to locate groups of words that they interpret as having similar properties (cohesiveness, moveability, ability to occur with multiple prosodic contours) as syntactic constituents. The adult data are less clear. It may be the adults can use prosody to locate constituent-like groupings, but it is also possible that adult participants relied on perceptually-prominent clause-initial and/or clause-final words to make grammaticality judgments about moved constituents.

In these first two experiments, the participants were monolingual English-acquirers or speakers, and the prosodic features of the experimental stimuli were based on English. In Chapter 4, three additional experiments are presented, asking whether infants and adults can use non-native prosody to locate syntactic constituents. These experiments explore the mechanism underlying prosodic bootstrapping, and ask whether the general acoustic prominence of prosodic cues or acquired knowledge of language-specific prosody drives prosodic bootstrapping effects. In Experiment 3, Japanese-acquiring infants were tested on stimuli with English prosody, and in Experiment 4, English-acquiring infants were tested on stimuli with Japanese prosody. Finally, in Experiment 5, American English-speaking adults were tested on sentences with Japanese prosody, in order to examine age-related effects and to clarify the results from Experiment 2.

CHAPTER 4

THE TRANS-LINGUISTIC ROBUSTNESS OF PROSODIC CUES IN SYNTAX ACQUISITION

This chapter presents three experiments examining the cross-linguistic use of clause-level prosody for locating groups of words that learners treat like cohesive syntactic constituents. These experiments address the role of ambient language experience in shaping the prosodic bootstrapping of syntax. While prosody is acoustically salient to mammals, languages differ in their manifestations of prosodic cues, and even young learners have demonstrated sensitivity to these language-specific prosodic patterns (Friederici, Friedrich, & Christophe, 2007; Halle, de Boysson-Bardies, & Vihman, 1991). Thus, if learners can use unfamiliar prosody to bootstrap knowledge of constituent-like groupings, it will suggest that the general acoustic salience of prosodic cues is sufficient to induce prosodic bootstrapping effects. On the other hand, if learners cannot use non-native prosody to locate syntactic constituents, it would suggest that prior experience with language-specific correlations between prosodic cues and syntactic boundaries is a necessary prerequisite for the prosodic bootstrapping of constituent-like units.

In Experiment 3, Japanese infants are tested on sentences with English prosody, and in Experiment 4, American English-acquiring infants are tested on sentences with Japanese prosody. Finally, Experiment 5 examines adult use of non-native prosody by testing American monolingual adults on stimuli with Japanese prosody. Before presenting the experiments, this chapter starts with a comparison of English and Japanese prosody.

4.1 A Comparison of English and Japanese Prosody

Japanese and English differ along a number of prosodic dimensions that affect the manifestation of clause-level (Intonational Phrase [IP]) prosodic cues. As discussed in Section 2.1.3, English clause boundaries, particularly in IDS, are marked with pauses, final-syllable lengthening, and pitch resets. This section will compare English and Japanese use of these cues, as well as differences in linguistic rhythm between the two languages.

There has been much work attesting to the reliability of pauses at IP/clause boundaries in English, particularly in English IDS (Cruttenden, 1986; Broen, 1972). However, the higher levels of the prosodic hierarchy have largely been ignored in work on Japanese until quite recently (cf. Selkirk, 2009): Kawahara & Shinya (2008) recorded four young women from the Tokyo area and found that the speakers consistently placed pauses in clause-final position. The stimuli were recorded in a controlled laboratory environment and were limited to coordinated clauses and gap constructions, so it is not clear how generalizable this result is, particularly to casual speech or IDS. However, Cruttenden (1986) has argued that pausing at clause boundaries may be a universal trait and Fernald et al. (1989) found that IDS is characterized by longer pauses than ADS in all six languages they studied (including English and Japanese). Overall, then, there is no reason to assume that Japanese and English differ in their use of pausing at clause boundaries.

Conversely, English and Japanese do differ in their use of final-syllable lengthening. English has a relatively strong degree of final-syllable lengthening, particularly in IDS (Cooper & Paccia-Cooper, 1980; Soderstrom, Blossom, Foygel, & Morgan, 2008). Studies of laboratory speech have found final lengthening in certain Japanese structures (Venditti & Yamashita, 1994), but others have produced mixed results, including evidence of sentence-final shortening in read sentences (Takeda, Sagisaka, & Kuwabara, 1989). Others suggest that final-syllable lengthening occurs with question, but not continuation, intonation in Japanese (Nishinuma, 1979). Even 18-month-old Japanese infants show less final-syllable lengthening than French infants in 2-word utterances (Halle, et al., 1991) (French, like English, has relatively extreme final-syllable lengthening).

Fisher and Tokura (1996a) analyzed spontaneous speech addressed to 14-month-old infants by American English and Japanese mothers for evidence of durational boundary cues. They found that American English-speaking mothers produced utterance-final vowels an average of 2.41 times longer than non-utterance-final vowels, while Japanese mothers only lengthened utterance-final vowels an average of 1.75 times non-utterance-final vowels. The American mothers also produced durational cues at some major phrase boundaries, whereas Japanese mothers were more likely to mark major phrase boundaries with pitch changes. This difference in the use of durational cues may be due, in part, to the fact that Japanese has contrastive vowel length. For example, *tsuki*

([tsuki]) means “moon” and *tsuuki* ([tsu:ki]) means “airflow.” This provides an additional challenge for the listener when interpreting duration cues at prosodic boundaries and may help explain the lesser (compared to English) degree of final-syllable lengthening seen in Japanese (cf. Lehiste & Fox, 1993).

Pitch cues are also very different in English and Japanese. While they have a similar degree of pitch excursions in pre-pausal positions (Fisher & Tokura, 1996a), the two languages differ in the types of intonational contours that occur and in how pitch information is used, particularly in regards to the marking of linguistic prominence. In English, a stressed syllable and associated unstressed syllables group together to form a prosodic foot, and the stressed syllable can optionally be marked with a pitch accent. Stress is lexically determined and is manifested through some combination of vowel quality, increased intensity, increased duration, and higher pitch (Beckman, 1986). English pitch accents, on the other hand, play a role in marking focus, as shown in examples (1) and (2) (pitch accent is marked in all caps).

(1) John is running.

(2) John is RUNning, not walking.

Unlike English, Japanese does not have stress. Instead, high-falling pitch accents are lexically determined and used to mark the heads of accentual phrases (Venditti, Jun, & Beckman, 1996). Accentual phrases are a proposed level of the prosodic hierarchy that falls somewhere between the level of the clitic group and phonological phrase (see Beckman & Pierrehumbert, 1986a). Pitch accent is contrastive: *kami* with accent on the first syllable means 'god,' and *kami* with accent on the second syllable means 'paper' (from Beckman & Pierrehumbert, 1986a, p. 256). This is comparable to the contrastive nature of stress in English *record* (N) vs *record* (V). However, while pitch accent can change the meaning of an utterance in English (see examples (3) and (4)), it does not change the meaning of a word.

(3) John is RUNNING.

(4) JOHN is running.

Given that pitch accents are lexically-determined in Japanese, Japanese speakers have less freedom to manipulate pitch for other purposes. In English, the pitch contour of an entire utterance can be manipulated to express things like uncertainty and impatience ((5a) and (5b), respectively), but this is less possible in Japanese (Beckman & Pierrehumbert, 1986b). In fact, while Japanese IDS is characterized by a larger overall pitch range as compared to Japanese ADS, American parents expand their pitch range significantly more in IDS than Japanese parents do (Fernald et al., 1989).

(5) a) He should buy the red one?

(6) b) He should buy the RED one!!

A final key difference between English and Japanese prosody is at the level of linguistic rhythm. English is considered a syllable-timed language: stress and other linguistic and psychological phenomena rely on syllables (e.g., Sato, 1993), while Japanese is a mora-timed language. (A mora is any phonetic segment that occurs in the syllable rhyme; for example, *ku*("nine") has one mora and *kun* (a suffix attached to boys' names) has two, even though they are both one syllable). Although the underlying mechanism is not well-understood, there is a large literature attempting to get at the roots of the intuition that mora-timed and stress-timed languages are perceptually distinct. Even newborns are sensitive to this rhythmic difference. Newborn French infants are able to discriminate between English and Japanese (a stress-timed and a mora-timed language), but not between English and Dutch (both stress-timed languages) (Nazzi, Bertoncini, & Mehler, 1998). Similarly, cotton-top tamarins and rats can discriminate between Dutch and Japanese speech (Ramus, Hauser, Miller, Morris, & Mehler, 2000; Toro, Trobalon, Sebastian-Galles, 2003), suggesting that the difference between stress-timed and mora-timed languages is based on general acoustic properties of the languages.

One proposed rhythmic difference between Japanese and English occurs at the level of syntactic phrases. English sentences are typically Subject-Verb-Object (SVO, e.g., "Mary kissed John"), and Japanese sentences are SOV, with the object preceding the verb (the equivalent of "Mary John kissed."). Based on the iambic/trochaic law – a perceptual phenomenon whereby elements of alternating duration are perceived as having the most prominent (longer) element last, and elements of alternating length intensity are perceived as having the more prominent (louder) element first – and the complement Law – the tendency for the complement to be more perceptually prominent than its head (Nespor & Vogel, 2007) – Nespor et al. (2008) argue that English and Japanese have different phrasal prosody. Verbs precede objects (their compliments) in English, so the ends of English predicates tend to be marked with duration cues. This has been confirmed by studies of final-syllable lengthening in English. In Japanese, objects precede verbs (the complement precedes the head), so the beginnings of Japanese predicates tend to be louder, with a subsequent decrease in amplitude. Indeed, Fisher and

Tokura (1996a) found evidence of final-syllable amplitude decrease in Japanese, further supporting the argument that English and Japanese prosody differ at the higher levels of the prosodic hierarchy.

Overall, there are numerous prosodic differences between English and Japanese, and infants have demonstrated sensitivity to certain language-specific prosodic features in early perception (Friederici et al., 2007) and production (Halle et al., 1991). Comparative work by Fisher and Tokura (1996a) and Nespor et al. (2008) suggests that English and Japanese have different prosodic templates for clauses and other major syntactic groupings. Therefore, if acquired knowledge of clause-level prosodic cues in the learner's native language drives the prosodic bootstrapping effect seen in Experiments 1 and 2, Japanese infants should fail to use English prosody to locate clausal constituents. On the other hand, the general acoustic prominence of pausing, final-syllable lengthening, and pitch resets may be sufficient to induce learning, regardless of the exact manifestation of these cues in the target language. If this is the case, Japanese infants should behave like the American infants in Experiment 1: those in the 2-clause group should discriminate between test sentences in which the clauses are consistent with those heard during familiarization versus sentences in which non-clausal units have been moved.

4.2 Experiment 3: Japanese Infants Listening to English Prosody

4.2.1 Participants

Twenty-four Japanese-acquiring 19-month-olds from the Tokyo area in Japan were tested (18 male, 6 female). Nineteen-month-olds were chosen to be consistent with the age of the American infants tested in Experiment 1. The mean age was 574 days, ranging from 553 days to 590 days. The inclusion criteria were the same as Experiment 1, and the infants had no more than ten hours a week of foreign language exposure, including time spent listening to American popular music. Data from an additional 19 infants were either not collected or excluded due to equipment error (n = 1), extreme shyness and crying during the long familiarization phase (n = 3), fussiness or crying during the 2-minute familiarization or testing phases (n = 14), and for standing with head out of camera range during testing (n = 1).

4.2.2 Materials, Design, and Task.

Materials, design, and task were identical to Experiment 1, insofar as was possible given that they were conducted in different labs. Note that while the stimuli were the same across experiments, the prosody and phonology were based on English and were non-native to the Japanese-acquiring infants in several ways. First, all syllables were CVC, with a variety of consonants in coda position. In Japanese, coda consonants can only be nasals or geminates (with the onset of the next syllable in a multi-syllabic word). Second, the phonetic inventory used in the stimuli words included several phonemes that do not occur in Japanese. The consonants /f/ and /v/ only occur rarely and exclusively in loanwords; /ɹ/ and /l/ do not occur at all. Phonemically, Japanese only has the tense vowels (/i, e, ɯ, o, a/), so the lax vowels used in most of the stimuli words (/ɪ, ɛ, æ, ʌ/ and the tense vowel /ɑ/) were not canonical native vowels. However, since the tense/lax dimension is not distinctive in Japanese, Japanese speakers exercise a certain amount of flexibility in their pronunciation, so the vowel sounds in the stimuli words should not be unfamiliar to the infants.

The task was the same as in Experiment 1. The experiment began with a ten minute "long" familiarization period, during which the infant, his or her parent, and an experimenter played quietly with toys while 1-clause (ABCDEF) or 2-clause (ABC, DEF) familiarization sentences played in the background. The infants were then tested using a modified version of the Head-Turn Preference Procedure in a separate room. During this phase of the experiment, there were an additional two minutes of familiarization, followed by 12 test trials, divided into three blocks, as described in Section 3.1.4. All of the infants were tested on items with 2-clause prosody that were either consistent (DEF, ABC) or inconsistent (EFA, BCD) with the clauses that the 2-clause infants heard during familiarization.

There were two small changes to the experimental set-up. First, while the American infants were seated in a high chair for testing, the Japanese infants sat on their parents' laps. Second, the attention-getting lights were different colors than in Experiment 1. There was a green light in front of the infant and two orange lights directly below speakers, one on each side of center (compared to an amber light in the front and red lights on the sides in Experiment 1).

4.2.3 Hypotheses

If the infants in the 2-clause group can use non-native English prosody to locate groups of words that they treat like syntactic constituents, they should behave like the American infants in Experiment 1 and discriminate between the consistent and inconsistent test items, since the consistent items involve a grammatical movement operation, while the inconsistent items involve an ungrammatical reordering of non-constituent units. The infants in the 1-clause group should also behave like their American counterparts and fail to discriminate between the consistent and inconsistent test items. This pattern of results would support the general acoustic hypothesis, which argues that prosody is useful in syntax acquisition because it is acoustically salient.

As in Experiments 1 and 2, the 1-clause group serves as a control for the possibility that learners preferentially attend to perceptually-salient boundary words (word classes A, C, D, and F in the 2-clause condition, A and F in the 1-clause condition). The 1-clause familiarization sentences share two boundary words with the consistent test items and zero with the inconsistent test items, while the 2-clause familiarization sentences share four boundary words with the consistent items and none with the inconsistent items. Therefore, if the infants base their listening preferences on boundary words, both the 1-clause and 2-clause groups should discriminate between the consistent and inconsistent items, with the 2-clause group perhaps discriminating to a larger extent.

Finally, it is also possible that neither the 2-clause nor the 1-clause infants will discriminate between the test item types. English and Japanese have numerous differences in clause-level prosody, so the English-based stimuli will be unfamiliar to the Japanese infants. This result would support the language-specific hypothesis, which argues that acquired knowledge about the target language's prosody is necessary before prosodic bootstrapping can occur.

4.2.4 Results

After looking times shorter than two seconds were excluded, the mean overall looking time was 7.43 seconds. Eight of the 12 infants in the 2-clause condition and five of the 12 infants in the 1-clause condition listened longer to the inconsistent test items.

The data were analyzed using mixed effects modeling, as discussed in Section 3.1.6. Familiarization condition and test condition were included as fixed effects (between and within subjects, respectively). Test condition was included as a random effect grouped by subjects. Looking time was the dependent variable, and looking times were log transformed to increase the normality of the residuals.

There was a significant interaction of familiarization by test condition ($\chi^2(1) = 8.65$, $p = .0033$). The effect of test condition was significant for the 2-clause condition ($\chi^2(1) = 4.66$, $p = .031$), with infants listening longer to the inconsistent test items. The effect of test condition was not significant for the infants in the 1-clause condition ($\chi^2(1) = 3.22$, $p = .073$), although there was a non-significant trend towards a preference for the consistent test items.

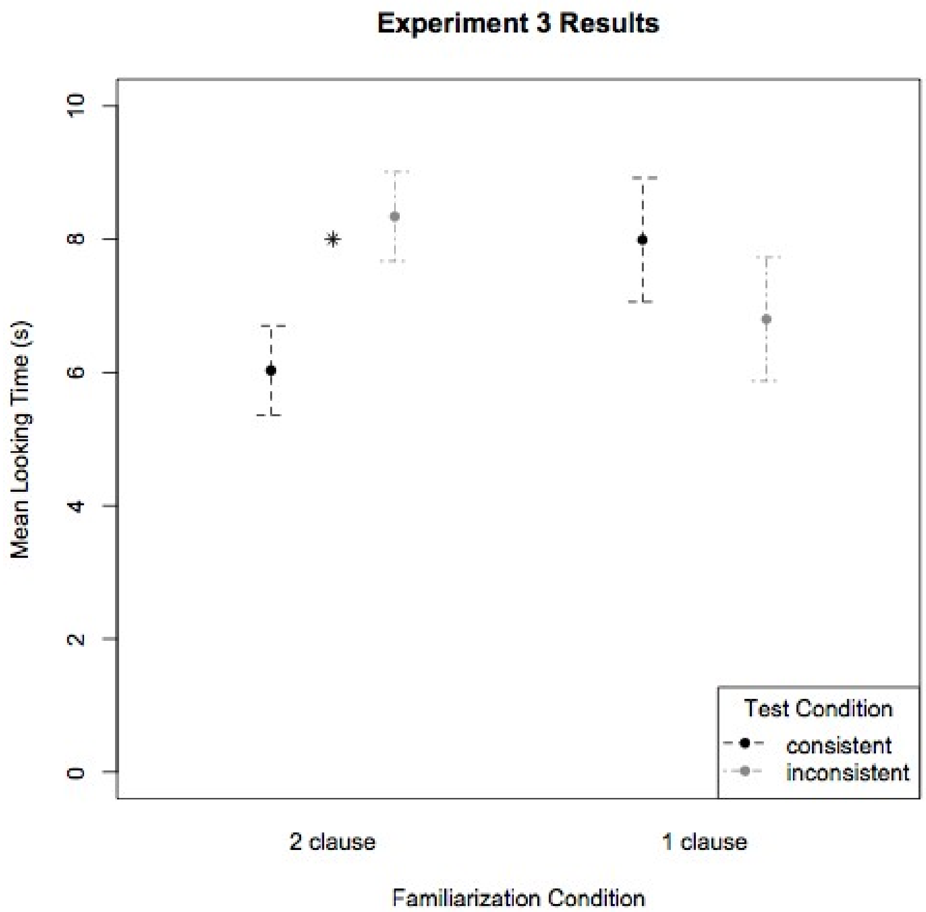

FIGURE 4.1 Error bars represent standard error of the mean difference. An asterisk represents a significant difference between the adjacent conditions. While the data were log transformed for the statistical tests reported above, this graph shows the raw data.

Closer examination of the data reveals that the trend towards a familiarity effect seen in the 1-clause group was largely caused by an odd pattern of looking times for the first test trial. As seen in Figure 4.2, infants in the 1-clause group whose first test trial was a consistent item listened much longer than average to that first test trial: a mean of 17.00s, which is more than 2.5 standard deviations above the overall mean looking time of 7.35s and over twice as long as the mean looking time for the first test trial (8.20s).

Conversely, the infants in the 1-clause group whose first test trial was an inconsistent item had unusually short looking times – an average of 4.87s. This is reflected in the significant interaction between test condition and test trial type (first or non-first) seen in the 1-clause group (χ^2 (2) = 10.15, p = .0063).

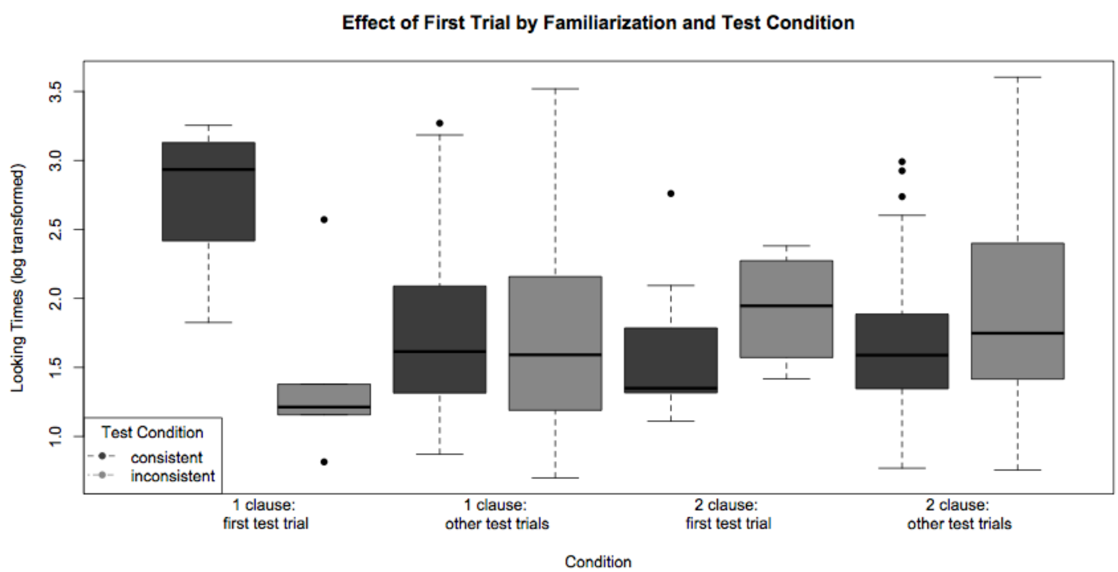

FIGURE 4.2 Box plot for Experiment 3 data, demonstrating the unusually high looking times of the infants in the 1-clause condition whose first test item was a consistent item.

It is difficult to interpret this first trial effect. The 1-clause participants were exposed to a change from 1-clause to 2-clause prosody at test, which could account for unusually long looking times; however, this interpretation is challenged by the short looking times from the 1-clause infants whose first test trial was an inconsistent item, since they also went from 1-clause prosody at familiarization to 2-clause prosody at test. Another possibility is that the infants from the 1-clause group with a consistent first test trial recognized that the clause-initial words from familiarization were still clause-initial in the consistent test items and were therefore showing a true familiarity effect. This would suggest that the infants in the 1-clause group would show an overall preference for the consistent test items, however, which is clearly not the case, (see Figure 4.2). A third, more likely, possibility is that the effect is due to random variation. There were only five participants from the 1-clause condition with a consistent item for their first test trial, and only four of them had an above-average listening time for that trial. Given that the 1-clause group did not show a statistically-significant familiarity effect overall, it would be unwise to conclude from these four extreme data points that the 1-clause group truly had a familiarity preference.

To check the validity of the results without the potentially biasing influence of the first test trial, the data were reanalyzed with the first test trial omitted for all participants. As was the case when the first test trial was included, there was a significant interaction between familiarization and test condition ($\chi^2(1) = 5.48$, $p = .019$), and a significant effect of test condition for participants in the 2-clause group ($\chi^2(1) = 4.54$, $p = .033$). However, there is no longer a trend towards an effect of test condition in the 1-clause condition ($\chi^2(1) = 1.15$, $p = .28$).

4.2.5 Discussion

Both with and without the inclusion of the first test trial, the results support the general acoustic version of the prosodic bootstrapping hypothesis and suggest that experience with the target language is not necessary for prosody to be an effective cue to constituent-like groupings. Despite the fact that the English-based stimuli are phonologically very different from their ambient language and should violate their clause-level prosodic expectations, Japanese-acquiring 19-month-olds in the 2-clause group were able to use prosodic information to (1) chunk the speech stream into smaller units and (2) treat those units like syntactic constituents by recognizing them when they obey a syntactic rule and occur with a new acoustic contour (initial versus final clause prosody). Infants in the 1-clause group did not significantly discriminate between consistent and inconsistent test items, suggesting that the infants in the 2-clause group did not base their looking time preferences on perceptually-salient clause-initial and/or clause-final words. Instead, it suggests that prosodic cues are sufficiently acoustically robust that they can be used to facilitate the acquisition of syntactic constituents without prior language-specific knowledge of the interactions between prosody and syntax in the target language.

An interesting pattern of results emerges when comparing the performance of the Japanese and American infants in Experiments 1 and 3. In both cases, only the infants in the 2-clause group discriminated, but the direction of preference was different, depending on the infant's native language. The English-acquiring infants had a familiarity preference for items that were consistent with the clauses they heard during familiarization, while the Japanese-acquiring infants preferred the novel, inconsistent test items.

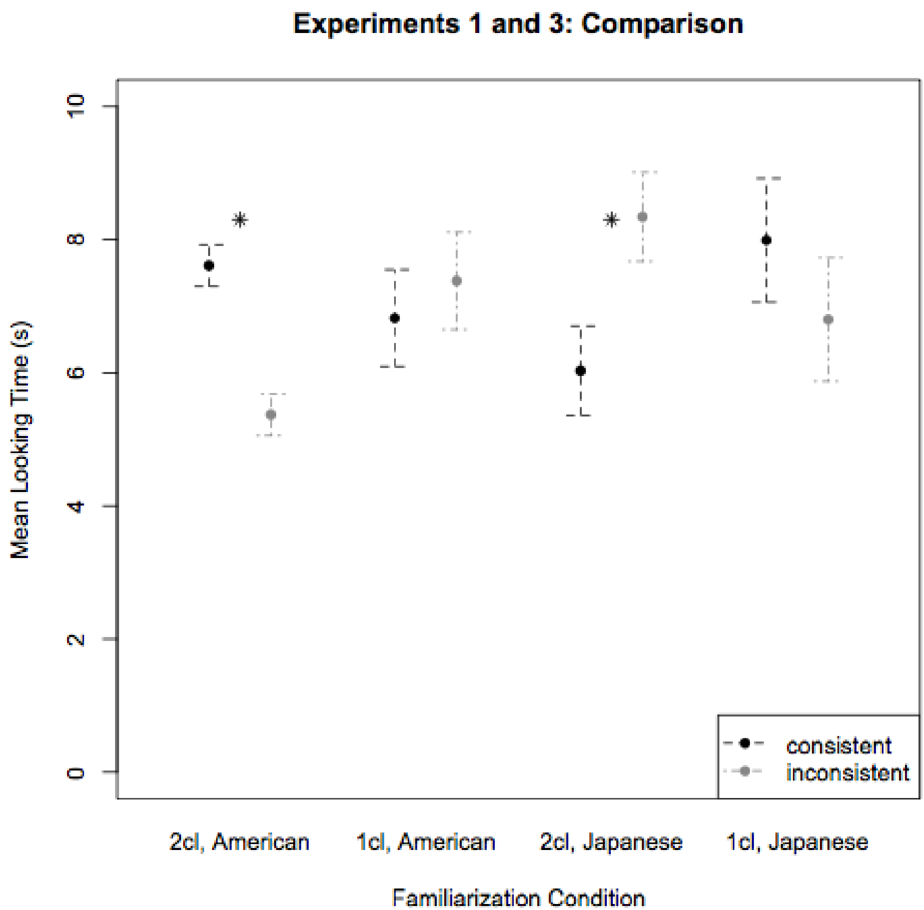

FIGURE 4.3 Error bars represent standard error of the mean difference. An asterisk represents a significant difference between the adjacent conditions.

Given conventional wisdom about the meaning of familiarity and novelty preferences, this is a rather surprising pattern of results. Infants tend to prefer familiar things (e.g., their mother's face instead of the face of a stranger; their native language instead of an unfamiliar language), but they eventually habituate to familiar stimuli and begin to prefer more novel stimuli. The easier the learning task, the more quickly the infant habituates, and the more likely she is to display a novelty preference (Hunter & Ames, 1988; Houston-Price & Nakai, 2004). So, at first glance, it appears that the

Japanese infants found the task of grouping the speech stream into clausal constituents *easier* than English-acquiring infants, even though the stimuli were prosodically and phonologically English-like.

There are at least two explanations that can account for this pattern of results. First, the Japanese-acquiring infants may have learned something different than the English-acquiring infants, due to differences in how native and non-native linguistic input are processed. Since the stimuli are quite different from Japanese in terms of phonemic inventory and syllable shape, the Japanese infants may have abstracted away from the phonological details of the individual stimuli words, focusing instead on the overall contour and the most salient words – those that occur in clause-initial and/or clause-final position. However, the lack of discrimination in the 1-clause condition suggests that this is not the case.

A more likely explanation is that the discrepancy in direction of preference is simply due to cultural differences. It has been noted that Japanese mothers spend more time rocking and carrying their infants compared to American mothers (Caudill & Weinstein, 1969), and this was observed by the experimenters during the initial long familiarization phase. In Experiment 1, every American infant played with toys during the entire 10 minute period, while the majority of Japanese infants in Experiment 3 spent a portion – or all – of that time on their mothers' laps, watching the experimenter play with the toys. Since the Japanese infants were, on average, less actively engaged with the toys during this period, they may have spent more time listening to the familiarization stimuli, such that by the testing period they preferred to listen to the more novel, inconsistent test sentences. Experiment 4, in which English-acquiring infants are tested on stimuli with Japanese-like prosody and phonology, will help solidify the interpretation of the direction of preference seen in Experiments 1 and 3 by replicating the experiment with American English-acquiring infants listening to sentences with Japanese-like prosody.

4.3 Experiment 4: American Infants Listening to Japanese Prosody

4.3.1 Participants

Twenty-four English-acquiring 19-month-olds from the Tucson area were tested (14 male, 10 female). The infants were between 563 and 604 days at time of test, with a mean of 583 days. The birth weight and birth term exclusion criteria from Experiments 1 and 3 were applied, and no infant had more than ten hours a week of foreign language exposure. Data from an additional 17 infants were not included due to experimenter or parental error (n = 2) and fussiness (n = 13). Two participants were excluded because they had more than three looking times below two seconds.

4.3.2 Materials

The materials were designed to be as similar as possible to the English-based materials used in Experiments 1-3, but the words and sentences were phonemically, phonotactically, and prosodically Japanese. The stimuli sentences each had six words with either 1-clause (ABCDEF) or 2-clause (e.g., ABC, DEF) prosody. The prosodic features are discussed in Section 4.3.2.1.

As with the English stimuli, each word class had two tokens, but instead of CVC words (which are rare in Japanese), each word had a long vowel, a diphthong, or a nasal coda consonant. Thus, the English-based and Japanese-based words were all bi-moraic in a way that complied with the phonotactic features of the target language. Care was taken that the stimuli words were non-occurring in both Japanese and English. The words are listed in Table 4.1.

Word Class	**A**	**B**	**C**	**D**	**E**	**F**
	gi:	ʥa:	ɸɯ:	ʤai	po:	ne:
	ʤeη	ɾoi	sɯη	toη	jɯη	doi

TABLE 4.1 Japanese stimuli words (in IPA)

The sentences were recorded in Tokyo, Japan by a young adult native Japanese speaker from Aichi Prefecture with an accent similar to the Tokyo dialect. The speaker was instructed to speak in a happy, IDS register. The 1-clause sentences were modeled on (7) and the 2-clause sentences were modeled on (8). Note that while some of the words have more than one syllable, they are all bi-moraic. The recordings were made using a Marantz PD 670/F1B recording device and a Audio-Technica ES935 unidirectional microphone in a sound attenuated booth.

(7)	Mō	ii	ame	huru	e-o	mita
	already	good	rain	fell	picture-ACC	looked

“I've already seen the good picture of falling rain.” -- OR --

“I've already seen the picture of good falling rain.”

(8)	Ame	huru	naka,	ii	e-o	mita
	rain	fell	while	good	picture-ACC	looked

"While it rained, I looked at a good picture."

The recorded sentences were partially-standardized using the criteria and methods discussed in Section 3.1.2. Unlike the English stimuli, the consistent and inconsistent test items were of comparable duration, so no overall duration adjustments were made.

4.3.2.1 Acoustic Analysis of the Prosodic Features

The stimuli sentences were analyzed as described in Section 3.1.2.1 for clause-level prosodic cues: pausing, final-syllable lengthening, and pitch resets. Five matched sentences were analyzed for both the 1-clause and 2-clause items. Unless otherwise noted, the significance tests reported below were paired-samples t-tests.

The average duration sentence duration was 2948ms for the 2-clause items and 2072ms for the 1-clause items. This 876ms difference is statistically significant ($t(4) = 18.18$ $p < .001$). The inter-clause pause of 526ms between words C and D for the 2-clause items partially accounts for this difference. There were no measurable pauses in the 1-clause sentences, making pausing a reliable cue to clause boundaries in the experimental stimuli.

To test for final-syllable lengthening in the English stimuli, vowel durations were measured and compared. However, Japanese has long vowels and moraic nasal consonants, so syllable rhyme durations were measured instead. The durations for each word class were compared with paired-samples t-tests, and results are presented in Table 4.2.

Word Class	2-Clause Mean	1-Clause Mean	t-value (df = 4)	Significant?
A (gi:, ʤeη)	310	283	3.73	p = .020
B (ʥa:, ɾoi)	262	239	6.73	p = .0025
C (ɸɯ:, sɯη)	420	291	6.47	p = .0029
pause	508	no pause	n/a	n/a
D (ʤai, toη)	247	243	0.00	no
E (po:, jɯη)	271	271	0.77	no
F (ne:, doi)	477	415	1.9	no

TABLE 4.2 Rhyme duration measurements (in milliseconds)

Unlike the English stimuli, in which duration only differed across familiarization condition for word class C, there were significant differences between the 1-clause and 2-clause items for word classes A, B, and C, implying that the first clause of the 2-clause items was overall produced more slowly than the same words in their 1-clause counterparts. More important than the effect of familiarization condition on the duration of each word class is the comparison of clause-final versus non-clause-final words. An unpaired t-test was conducted comparing the durations of clause-final and non-clause-final rhymes (A, B, D, and E versus C and F for the 2-clause items; A, B, C, D, and E versus F for the 1-clause items). The clause-final rhyme duration was significantly longer than the non-clause-final rhyme duration ($t(16.57) = 8.65$, $p < .001$)[5]. Thus, the Japanese stimuli contain reliable final-syllable lengthening at clause boundaries.

While there is a simple effect of clause-final versus non-clause final position on duration for both English (reported in Section 3.1.2.1) and Japanese, the interaction of stimulus language (English versus Japanese) and clause-final versus non-clause final position is also significant ($F(1, 114) = 11.04$, $p < .01$), with the English stimuli having

[5] See footnote 2 from Chapter 3 for an explanation of the non-integer degrees of freedom reported here.

relatively more lengthening in clause-final positions than the Japanese stimuli. For example, the vowels in word class C were an average of 2.5 times longer in the 2-clause familiarization items (where the word is pre-pausal) than in the 1-clause familiarization items (where it is not) for the English items, but only 1.4 times longer in the Japanese stimuli. This is in line with previous findings (e.g., Fisher & Tokura, 1996) that American English speakers lengthen final syllables to a greater degree than Japanese speakers.

Pitch patterns in Japanese clauses are more constrained than in English. As discussed in Section 4.1, high-falling pitch accents are lexically determined and mark the head of an accentual phrase. Japanese also has consistent declination throughout the sentence, with pitch resets at clause boundaries (Beckman & Pierrehumbert, 1986a). The stimuli shown in Figures 4.4 and 4.5 clearly reflect these patterns. The single clause in Figure 4.4 has one pitch accent followed by declining pitch through the rest of the sentence. The two clauses in Figure 4.5 both exhibit declination, and there is a pitch reset across the clause boundary. Pitch cues, therefore, are only consistent with an internal clause boundary for the 2-clause items.

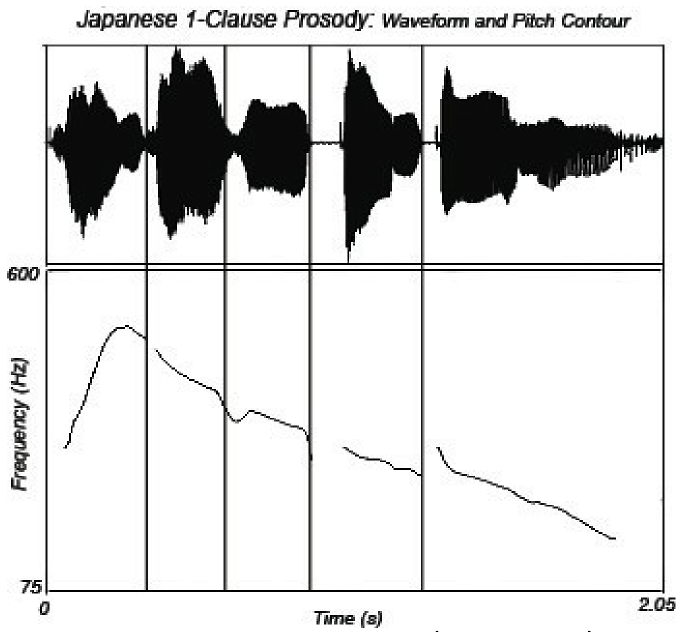

FIGURE 4.4 Japanese 1-clause prosody. Example sentence: ʤɛŋ ʥ*a: ɸu: ʤai po: ne:*. The pitch contour trails off at the end because the speaker was using creaky voice. Figure captured from Praat (Boersma & Weenink, 2008).

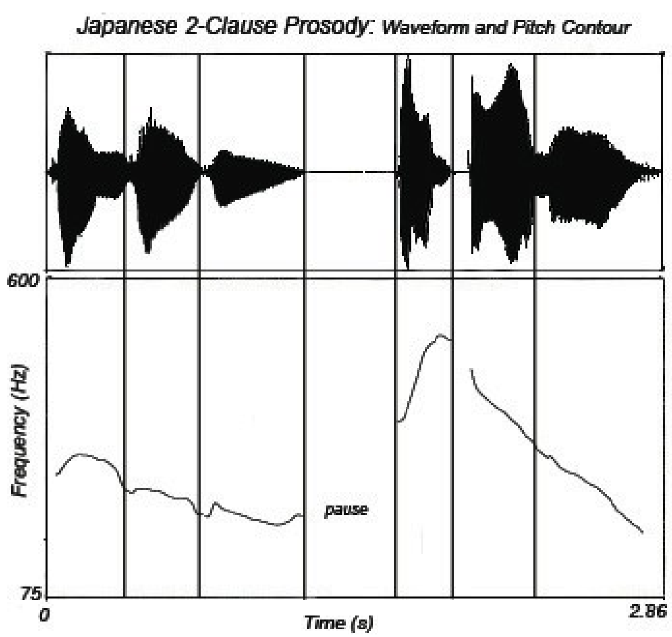

FIGURE 4.5 Japanese 2-clause prosody. Example sentence: ʤɛŋ ʥ*a: ɸu:, ʤai po: ne:*. Figures captured from Praat (Boersma & Weenink, 2008).

Like the English stimuli, the Japanese items have reliable pausing and final-syllable lengthening at the internal clause boundary. While the intonational contours differ dramatically across the two stimuli sets, the Japanese 2-clause sentences have a pitch reset and declination patterns consistent with a clause boundary between words C and D. Therefore, the prosodic features of these stimuli should be sufficient to support prosodic bootstrapping of clause-level units if participants are able to engage in prosodic bootstrapping in a language with unfamiliar prosody.

4.3.3 Design and Task

The design and task were identical to Experiments 1 and 3. Infants were familiarized with either 1-clause (ABCDEF) or 2-clause (ABC, DEF) sentences over two phases: a "long" familiarization phase, followed by a "2-minute" familiarization phase. Familiarization was immediately followed by consistent and inconsistent test trials, using the Head-Turn Preference Procedure. The consistent items represent a grammatical movement of the clausal constituents from the 2-clause familiarization items (DEF, ABC), and the inconsistent items contain an ungrammatical reordering of non-constituent units (EFA, BCD).

Since the Japanese sentences had a slightly longer duration than their English counterparts, the Japanese "long" and "2-minute" familiarization audio files were longer than the English audio files in order to keep the total number of sentences as consistent as possible. Table 4.3 lists the overall duration of and total number of sentences in each familiarization audio file. Recall that the 2-clause sentences are longer than the 1-clause sentences, so the total duration of the audio files was matched within each language's stimuli (e.g., the 2-minute sound files were 120s for both the English 1-clause and 2-clause files, and 150s for both the Japanese 1-clause and 2-clause files) . Since the 1-clause group is predicted not to discriminate, having the 1-clause participants listen to more sentences gives them extra learning time and works against the hypothesis. Also, note that the total number of sentences was matched across the English and Japanese 2-clause files, in order to give the participants the same number of familiarization sentences in the crucial condition.

		English-Based Stimuli		**Japanese-Based Stimuli**	
		1 clause	**2 clause**	**1 clause**	**2 clause**
Short File	**duration**	*120s*	*120s*	150s	150s
	# of sentences	43	**38**	49	**38**
Long File	**duration**	*608s*	*608s*	757s	757s
	# of sentences	218	**192**	244	**192**

TABLE 4.3 Comparison of total duration and number of sentences in each familiarization file. Total duration was matched across the 1-clause and 2-clause conditions for each language's stimuli. Total number of sentences was matched across the English and Japanese stimuli for the 2-clause condition.

4.3.4 Hypotheses

If the English-acquiring infants in the 2-clause group can use Japanese prosody to facilitate learning of constituency, despite the unfamiliar intonational contour and lesser degree of final-syllable lengthening, it will provide additional support for the general acoustic version of the prosodic bootstrapping hypothesis. It will suggest that extensive prior knowledge of the correlations between prosodic and syntactic boundaries in the target language is not a prerequisite to prosodic bootstrapping. Importantly, this would also be evidence that the results from Experiments 1 and 3 generalize to languages with different prosodic properties than English.

In the event that participants in the 2-clause group do show significant discrimination, it will also be instructive to examine the direction of their looking time preference. If participants in the 2-clause group show a familiarity preference for the consistent items with the Japanese-based stimuli (as they did with English-based stimuli), it will suggest that the difference in direction of preference seen in Experiments 1 and 3 is due to cultural factors. Alternatively, if English-acquiring infants in the 2-clause group show a novelty preference, it will suggest that the novelty preference of the Japanese-acquiring infants listening to English prosody was due to a difference in how infants interpret native and non-native linguistic stimuli.

On the other hand, it is also possible that the infants will not show evidence of learning with the Japanese stimuli. Compared to Japanese, English has strong final-syllable lengthening and very exaggerated pitch contours in IDS (e.g. Fisher & Tokura, 1996a). It is therefore possible Japanese infants can use English prosody for locating constituent-like groups of words, while English-acquiring infants can't use Japanese prosody for the same task.

4.3.5 Results

Once looking times shorter than two seconds were excluded, the overall mean looking time was 7.02 seconds. Nine of the 12 infants from the 2-clause condition and seven of the 12 infants from the 1-clause group showed a familiarity preference for the consistent test items.

The data were analyzed using mixed effects modeling, as described in Section 3.1.6. Familiarization condition was included as a between-subjects fixed effect and test condition was included as a within-subjects fixed effect and a random effect by subjects. Looking times (the dependent variable) were log normalized to comply with the requirement that residuals be randomly distributed.

There was a significant interaction of familiarization by test condition (χ^2 (1) = 4.18, p = .041). The infants in the 2-clause condition listened significantly longer to the consistent test items (χ^2 (1) = 5.22, p = .022), while the 1-clause participants did not discriminate between the consistent and inconsistent test items (χ^2 (1) = .06, p = .81).

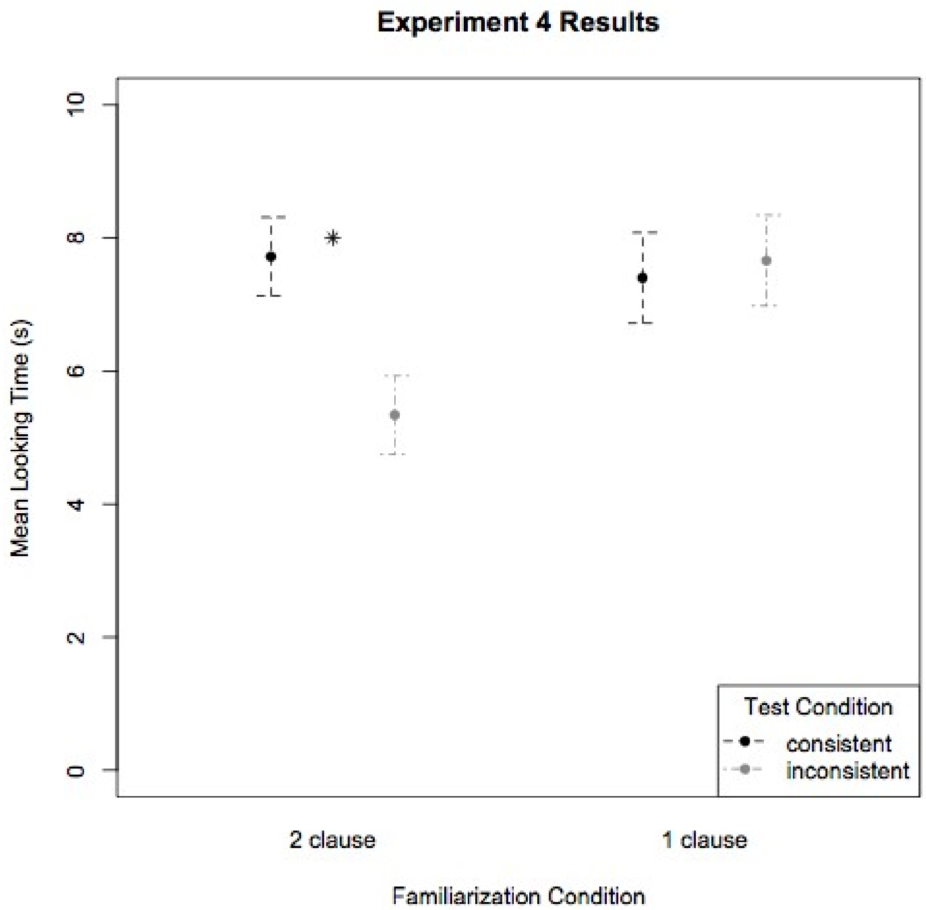

FIGURE 4.6 Error bars represent standard error of the mean difference. An asterisk represents a significant difference between the adjacent conditions. While the data were log transformed for the statistical tests reported above, this graph shows the raw data.

4.3.6 Discussion

Results support the general acoustic version of prosodic bootstrapping, bolstering the findings of Experiment 3 and further suggesting that infants do not require extensive prior experience with the prosodic features of the target language to use prosody as a cue to constituent-like units.

While Experiments 1 and 3 tested American and Japanese infants on English-based stimuli, the stimuli in the present experiment were based on Japanese, which differs markedly in the manifestation of clause-level prosodic cues, particularly intonation. Despite the unfamiliar nature of the prosodic information, English-acquiring 19-month-olds produced looking time patterns that were nearly identical to the English-acquiring participants listening to English-based stimuli in Experiment 1. This is clearly reflected in Figure 4.7. There is, therefore, no indication that prior experience with the prosodic features of the familiarization language played any role in the prosodic bootstrapping effects seen in American infants in Experiments 1 and 4.

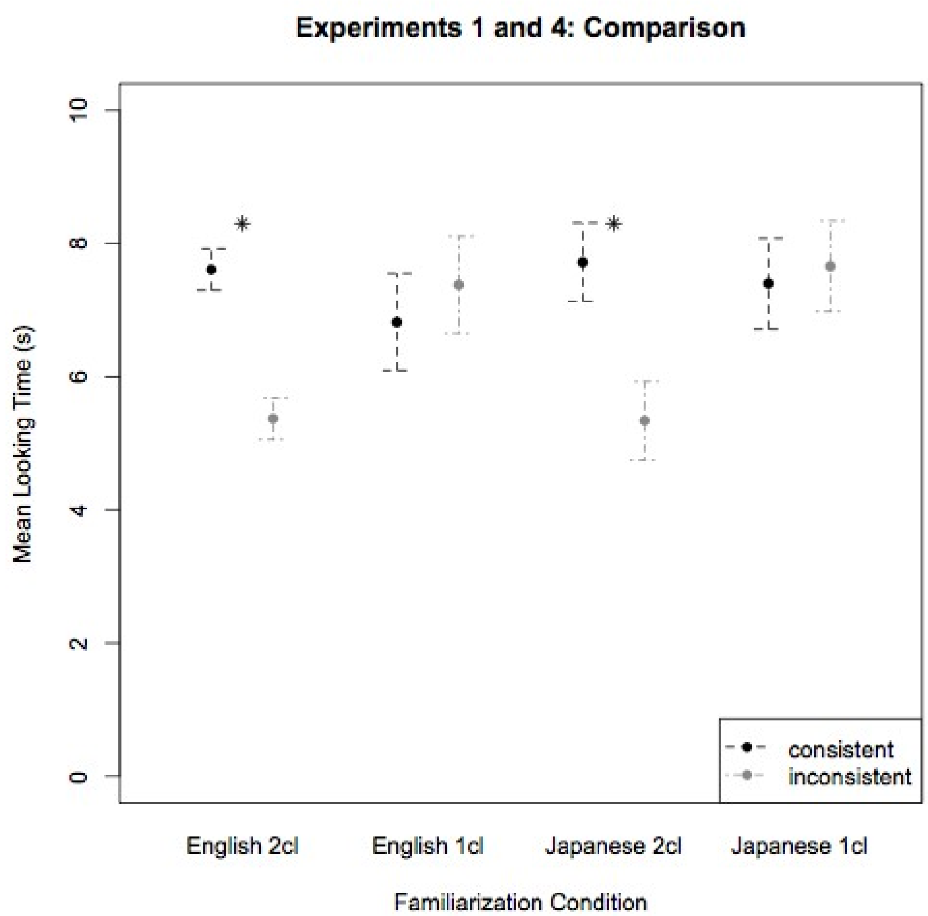

FIGURE 4.7 Error bars represent standard error of the mean difference. An asterisk represents a significant difference between the adjacent conditions.

The similarity between behavior in Experiments 1 and 4 is particularly interesting given the extremely different intonational patterns in the English and Japanese stimuli. Recall that Seidl (2007), in a replication of Nazzi et al's (2000) experiment on the influence of prosody on memory, found that pitch was the only prosodic cue that was necessary to obtain memory effects in American 6-month-olds – pausing and final-syllable lengthening combined were not sufficient. The present findings suggest that infants are flexible in their interpretation of what constitutes constituent-like intonational contours. While this type of flexibility would clearly be beneficial to young infants who may be born into any language community, the fact that it continues beyond the first year and a half of life suggests that infants remain open-minded about their prosodic expectations for clauses at least into the beginnings of the syntactically-important two-word stage.

Finally, Experiment 4 supports a cultural interpretation for the discrepancy in direction of preference seen in Experiment 1 versus Experiment 3. Recall that the American infants in the 2-clause group of Experiment 1 showed a familiarity preference for the consistent test items, while their Japanese counterparts had a significant novelty preference. The familiarity preference displayed by the American infants in the 2-clause group of the current experiment suggests that the discrepancy in preference was not due to a difference in how infants interpret native and non-native prosody, but rather to the fact that the Japanese infants were less engaged with the toys and more engaged with listening to the familiarization stimuli during the 10-minute long familiarization period.

In Experiment 5, I ask whether adults are similarly able to use the general acoustic manifestations of prosody to parse speech into constituent-like units or if the flexibility seen at 19 months disappears by adulthood.

4.4 Experiment 5: American Adults Listening to Japanese Prosody

4.4.1 Participants

Forty-eight undergraduates from the University of Arizona linguistics classes (15 male and 33 female) participated in the study for extra credit. The exclusion criteria were the same as for Experiment 2. An additional six participants were excluded due to failure to understand the experimental task (n = 3) and because their mean grammaticality rating scores were more than two standard deviations from the overall mean (n = 3).

4.4.2 Materials, design, and task

The experiment was identical to Experiment 2 in design and task; the only difference is that Experiment 2 used English-based stimuli and Experiment 5 used Japanese-based stimuli. Participants were familiarized with either the Japanese 1-clause or Japanese 2-clause long familiarization audio file, and were then asked to rate consistent and inconsistent test sentences on a scale of 1-4, based on how well each sentence complied with the familiarization grammar.

4.4.3 Hypotheses

If adults are able to use non-native prosody to locate clause-like constituents, participants in the 2-clause group should recognize the prosodically-marked clauses from familiarization when they have moved to a new position in the utterance and therefore rate the consistent items as significantly better than the inconsistent items. Participants in the 1-clause group should rate both types of test items as equally grammatical, since the 1-clause familiarization items do not contain internal prosodic boundaries. This would suggest that adults, like infants, are able to capitalize on the general acoustic manifestations of prosodic cues that occur at major constituent boundaries.

Alternatively, it may be the case the participants in both the 1-clause and 2-clause groups rate the consistent items as significantly more grammatical than the inconsistent items. As in Experiment 2, they may do so to different extents, with the 2-clause participants differentiating more between the test item types. This pattern of results would suggest that prominent words at the beginnings and/or endings of clauses play a role in adults' judgments, since the 1-clause group otherwise would have no reason to prefer the consistent test items over the inconsistent items. However, the stronger effect seen in the 2-clause group may also suggest that adults are not solely relying on boundary words, but they are also using prosody to learn which words are grouped together to form cohesive, moveable constituents. On the other hand, if there is no interaction between familiarization and test condition, such that the 1-clause and 2-clause participants both prefer the consistent items to the same extent, it would suggest that the participants are basing their judgments entirely on perceptually prominent boundary words words, rating the consistent items better because they share clause-initial and clause-final words with sentences from both familiarization conditions.

Finally, adults in both conditions may fail to discriminate between the consistent and inconsistent items, suggesting that adults are not able to use prosody that violates their acquired prosodic expectations for the task of chunking the speech stream into moveable constituent-like units. This would support the language-specific version of prosodic bootstrapping for adults, and would present a contrast to infants, who make flexible use of acoustically salient information at prosodic boundaries to facilitate syntax learning.

4.4.4 Results

The data were analyzed with linear mixed effects modeling as described in Section 3.2.4. Familiarization condition (1-clause versus 2-clause) was a between-participants variable and test condition (consistent versus inconsistent) was a within-participants variable. The dependent variable was grammaticality rating scores. Test condition was included as a random factor by subjects and by items.

The interaction of familiarization and test condition was not significant ($\chi^2(1) = .76$, $p = .38$), indicating that the 1-clause and 2-clause groups did not differ in their grammaticality judgments of the consistent and inconsistent items. The main effect of test condition was significant ($\chi^2(1) = 51.06$, $p < .001$), with participants judging the consistent items as significantly better than the inconsistent items.

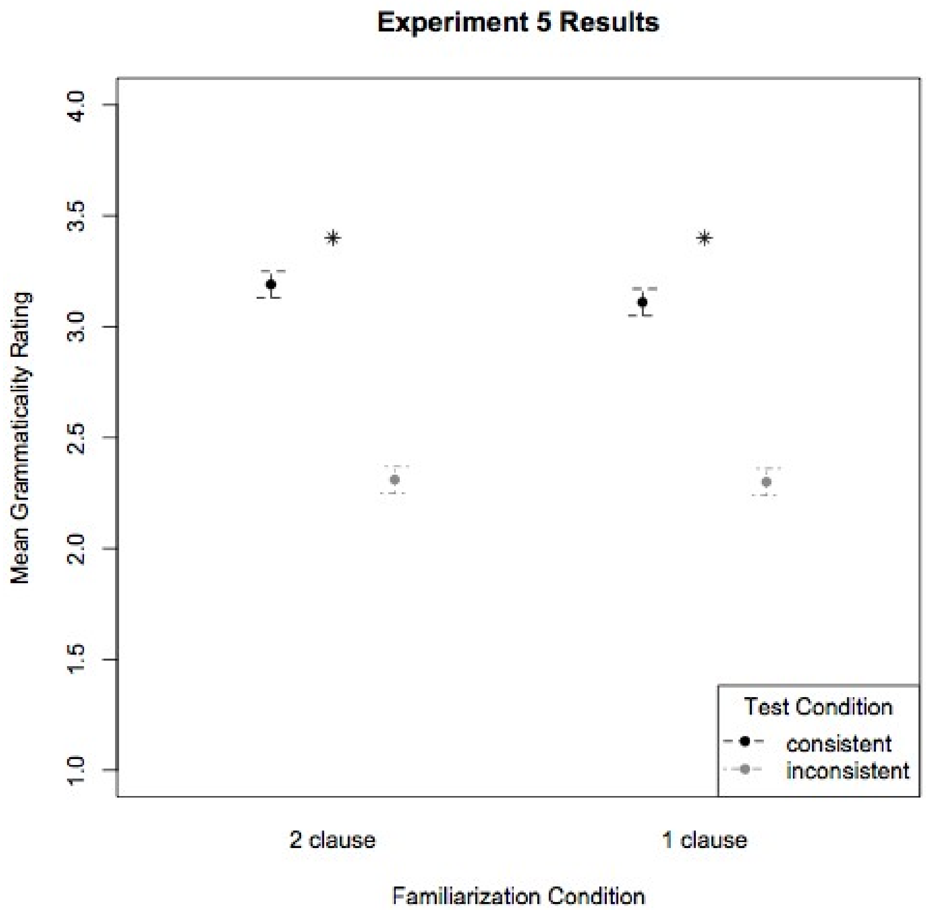

FIGURE 4.8 Error bars represent standard error of the mean difference. An asterisk represents a significant difference between the adjacent conditions. Although the interaction of familiarization by test condition was not significant, the effect of test condition was computed separately for each familiarization for purposes of visual comparison. The effect of test condition was significant in both the 2-clause ($\chi^2(1) = 28.98$, $p < .001$) and 1-clause conditions ($\chi^2(1) = 23.19$, $p < .001$).

The data were also considered with respect to data from Experiment 2. Familiarization language (English [Experiment 2] versus Japanese [Experiment 5]) and familiarization condition (1-clause versus 2-clause) were between-participants factors, and test condition (consistent versus inconsistent) was a within-participants factor. Test condition was included as a random effect grouped by both subjects and items. Grammaticality rating scores were the dependent variable. There was a significant 3-way interaction (χ^2 (4) = 9.71, $p = .046$), indicating that the overall pattern of results differed depending on familiarization language. This is supported by the significant familiarization condition by test condition interaction for the English stimuli (χ^2 (1) = 17.66, $p < .001$), but the lack of a familiarization condition by test condition interaction for the Japanese stimuli (χ^2 (1) = .76, $p = .38$).

To determine whether the 3-way interaction trend was driven by the 2-clause or 1-clause conditions, the familiarization language by test condition interaction was calculated for each familiarization condition separately. There was no significant interaction between familiarization language and test condition for the 2-clause groups (χ^2 (2) = .86, $p = .65$), but the interaction was significant for the 1-clause groups (χ^2 (2) = 10.62, $p = .0049$), with participants in the 1-clause Japanese group having a significantly larger discrepancy between their ratings of consistent and inconsistent test sentences than the 1-clause English group. This suggests that the different pattern of results across Experiments 2 and 5 is driven by how participants in the 1-clause condition treat the experimental stimuli.

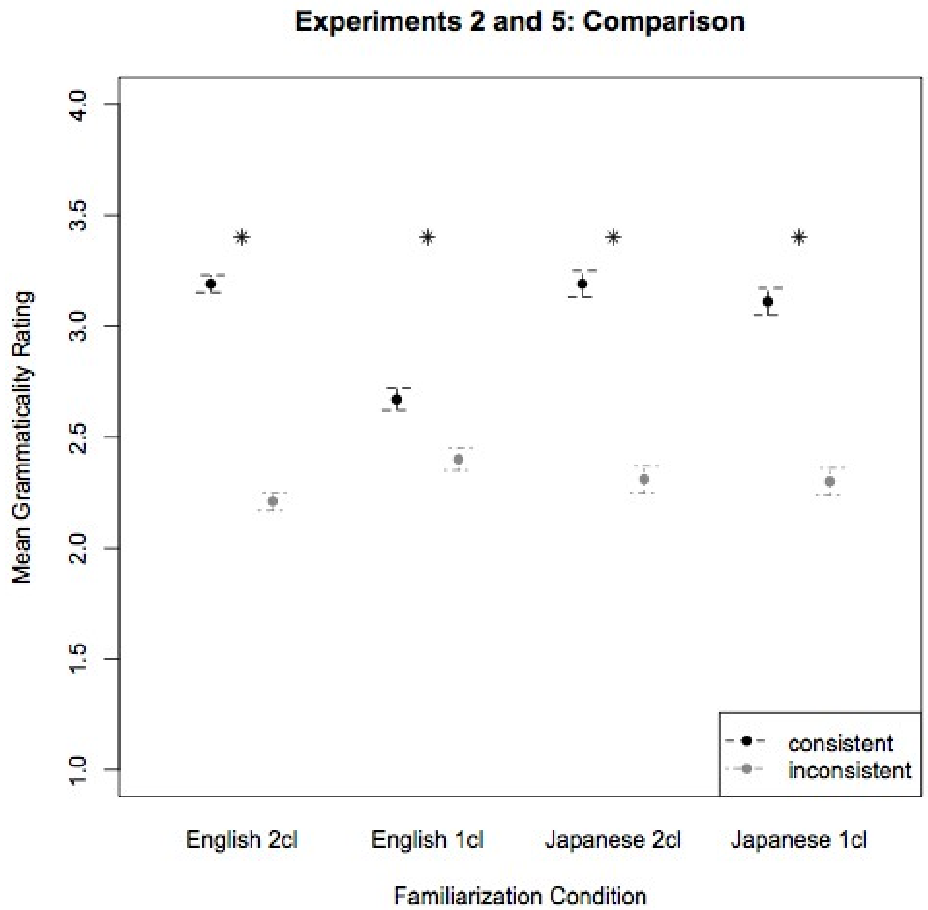

FIGURE 4.9 Error bars represent standard error of the mean difference. An asterisk represents a significant difference between the adjacent conditions. Note that while there was a significant effect of test condition for each familiarization condition, the familiarization by test condition interaction was only significant in the Experiment 2 (English-based stimuli).

4.4.5 Discussion

The lack of interaction between familiarization and test condition supports the language-specific version of the prosodic bootstrapping hypothesis for adult learners. Both the 2-clause and 1-clause groups rated the consistent test items as significantly more grammatical than the inconsistent items. While it is possible that the 2-clause group showed this preference because they were learning the constituent structure of the familiarization items, the fact that their responses were not significantly different from the responses of the 1-clause group suggests that this was not the case. Instead, the most likely explanation is that participants in both groups based their judgments on words A and F, which occurred in perceptually-prominent clause boundary positions for sentences from both familiarization conditions as well as the consistent test condition.

A possible alternative explanation is that the Japanese stimuli do not contain sufficient prosodic information pointing to the internal clause boundary: as previously discussed, Japanese IDS is less prosodically-exaggerated than American English IDS. However, American infants were equally adept at learning with the Japanese stimuli in Experiment 4 as they were with the English stimuli in Experiment 1, suggesting that the this was not the case.

Taken together, then, the experiments presented in this chapter point to a difference in the role of prosody in syntax acquisition for infant and adult learners. While English-acquiring infants can use both native and non-native prosody to parse speech into cohesive, moveable, constituent-like units, adults are only able to do so with familiar, native prosody. This is in contrast to prior studies that have found that adults are able to use unfamiliar prosody in grouping tasks. For example, American adults rate Korean sentences with a pause at the subject-predicate boundary more highly than sentences in which the pause occurs inside a constituent or word (Pilon, 1981). Similarly, Italian adults can use synthesized Japanese-like IP boundaries to constrain word segmentation (Shukla, Nespor, & Mehler, 2007) – another grouping task. The present experiment asked adults to go beyond mere grouping of the speech stream to recognize that prosodically-grouped words can form syntactic constituents. While adults succeeded with familiar native prosody in Experiment 2, their failure to learn the constituent structure of the Japanese stimuli sentences suggests that bootstrapping with non-native prosody breaks down in adults for the more syntactically-sophisticated task of parsing speech into groups of words into constituents.

CHAPTER 5

CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

Prosody is a rich potential source of information for many major tasks facing young language learners, from locating word boundaries (Jusczyk, Houston, & Newsome, 1999) to determining a speaker's affect (Fernald, 1993) to constraining the domain over which learners look for words (Gout, Christophe, & Morgan, 2004). This dissertation addresses two questions related to the prosodic bootstrapping of syntax. First, can learners use prosody to parse speech into syntactic constituents (Section 5.5.1)? Second, does prosodic bootstrapping rely on the general acoustic salience of prosodic cues or does language-specific experience mediate learning (Section 5.1.2)? Infant and adult performance was compared with respect to both questions in order to examine how prosodic bootstrapping changes with age and increased native-language experience.

5.1 Summary and conclusions: Prosodic bootstrapping of constituency.

In Chapter 1, I introduced three versions of the prosodic bootstrapping hypothesis: grouping, constituents, and phrase structure. All three argue that infants use prosody to pull groups of words from the speech stream, but they differ in how infants treat those words. In the grouping version of prosodic bootstrapping, infants use prosody to segment a string of words into substrings that they hold in working memory for further analysis. Previous work reviewed in Sections 2.2.2 and 2.2.3 offers evidence that infants and adults are able to use prosody for grouping.

The experiments in Chapter 3 tested the constituents version of the prosodic bootstrapping hypothesis. According to this proposal, learners can not only use prosody to group speech into substrings, but they recognize that these substrings are syntactic constituents. In Experiments 1 and 2, learning of constituents was assessed by asking if infants and adults recognize prosodically-grouped words when they move to a new position in the sentence. This movement test was chosen because (1) constituents are cohesive groups of words that act together in sentences, (2) only constituents can move, and (3) moved constituents have a new prosodic contour. Thus, if a learner recognizes a prosodically-parsed group of words when it has moved, it suggests that she is treating the words like a syntactic constituent. It also suggests that she is not merely reacting to identity or near-identity of acoustics.

In Experiment 1, 19-month-olds whose familiarization sentences contained two prosodically-marked clauses listened significantly longer to test sentences that maintained the clause-like constituents from familiarization, providing evidence that infants treat prosodically-grouped words like syntactic constituents. Infants from the 1-clause familiarization group showed no preference for consistent or inconsistent test items. In Experiment 2, adults in both the 2-clause condition and the 1-clause control condition rated the consistent test sentences as more grammatical than inconsistent sentences in which a non-constituent string of words had moved. Since the 1-clause items share two boundary words (classes A and F) with the consistent test items, the significantly higher ratings for consistent items from the 1-clause participants suggests that adult performance was influenced by perceptually-salient words at boundaries. While this could well be a prosodic effect – words A and F occurred at prosodic boundaries – this does not demonstrate evidence of syntax learning.

However, there was also a significant familiarization by test condition interaction for the adults in Experiment 2. The 2-clause group discriminated more in their ratings for consistent versus inconsistent test items than the 1-clause group. This could be because the 2-clause items share more clause-initial and clause-final words with the consistent items (A, C, D, and F) than the 1-clause items do (A and F). However, given that this is also true for the 1- versus 2-clause stimuli in Experiment 5 and there was no familiarization by test condition interaction in that experiment, this explanation alone is not sufficient. A more likely possibility is that the 2-clause group in Experiment 2 relied on boundary words but simultaneously used prosody to chunk the sentences into constituent-like units. The higher grammaticality ratings for the consistent test items suggest that adults, like infants, recognized clause-like constituents when they had moved to a new position in the sentence.

5.1.1 Alternative interpretations and possible follow-up studies

An alternative interpretation raised in Section 3.1.7 merits further comment and experimentation. Instead of learning constituency, participants in Experiments 1 and 2 may have used prosody to constrain the domain over which they learned legal sequences of syllables (e.g., ABC is a legal sequence, but CDE is not, since it crosses a prosodic boundary in the 2-clause familiarization items). Under this interpretation, 2-clause participants may have learned that the grammar comprised two finite-state automata (FSAs), but not that the FSAs were in a meaningful relationship with each other – namely, that they grouped together to form a single sentence (Figures 5.1 and 5.2, respectively). If this is the case, than these experiments are evidence for the grouping version of the prosodic bootstrapping hypothesis, but do not speak to whether participants inferred that grouped words were syntactic constituents.

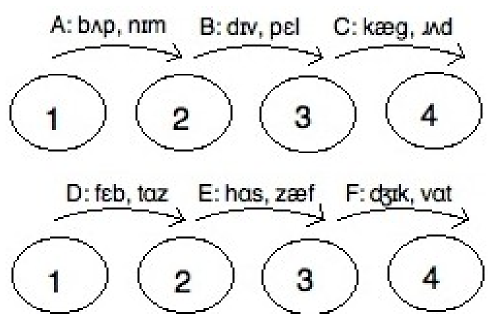

FIGURE 5.1 Two independent finite-state automata. State 1 is the start state and state 4 is the accepting state.

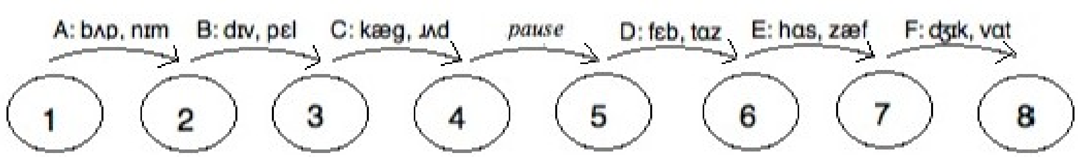

FIGURE 5.2 One finite-state automaton. State 1 is the start state and state 8 is the accepting state.

There are two separate issues at play in this interpretation. First, are learners acquiring a regular grammar (equivalent to FSAs), or are they learning a more natural-language-like context-free grammar? Second, do participants realize that the clause-like units (or clause-sized FSAs) combine to form sentences? In order to determine what kind of grammar learners are inferring, Experiments 1 and 2 could be replicated with one small change: completely withholding particular combinations of words until test. A potential 2-clause familiarization grammar is shown in Figure 5.3. This FSA does not generate the ABC clauses *bʌp dɪv kæg* and *nɪm pɛl ɹʌd* or the DEF clauses *fɛb hɑs* ʤɪ*k* and *t*ɑ*z zæf v*ɑ*t*. If these unattested combinations of words occur for the first time at test and learners in the 2-clause conditions still discriminate between consistent and inconsistent items (infants) or rate consistent items as more grammatical (adults), it would suggest that learners are generalizing beyond the familiarization input. In Experiments 1 and 2, participants generalized to new prosodic contours, but generalizing at the level of the word class would be stronger evidence that participants are learning a natural-language-like grammar.

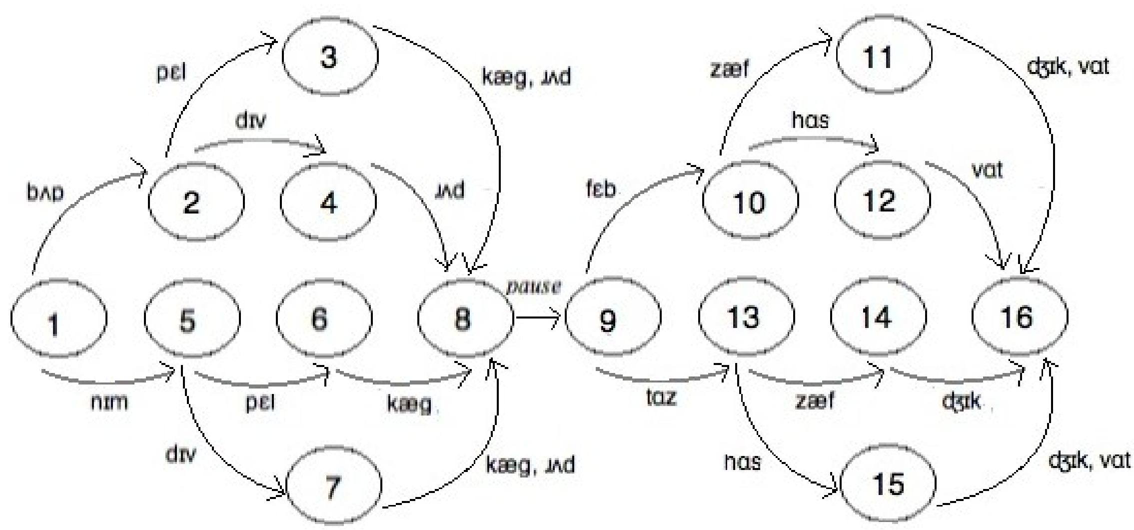

FIGURE 5.3 Possible alternative familiarization grammar. State 1 is the start state and state 16 is the accepting state.

Although this type of word class generalization was not explicitly tested in the current experiments, it is likely that learners did generalize to the word classes. Recall that there were 64 unique sentences with each type of familiarization prosody. Each sentence was heard only three times during the long familiarization period, and none of the test sentences occurred in the 2-minute familiarization period immediately preceding testing (note that adults only had a single, long familiarization period). Thus, while further experimentation would strengthen the claims made here, the current work certainly provides initial support for the constituents version of prosodic bootstrapping.

The second aspect of the alternative explanation discussed above – whether 2-clause participants interpret ABC and DEF as part of a larger, sentence-level grouping or as two separate FSAs or context-free grammars – could be addressed by asking if learners treat ABC and DEF as independent units. If participants interpret the 2-clause items as comprising two clauses in a meaningful (i.e. sentential) relationship to each other, they should treat test sentences in which the two clauses occur together with sentence prosody (Example 1) differently from test items in which the two clauses do not occur in relationship to each other (Example 2) (a backslash indicates a 1-second pause between stimuli items).

(1) ABC, DEF / DEF, ABC / DEF, ABC....

(2) ABC / DEF / DEF / ABC...

If infants discriminate between test stimuli like Examples 1 and 2, it would suggest that learners recognize that prosodically-marked clauses are cohesive units that operate as independent chunks within the larger domain of the sentence. This would support the constituents version of the prosodic bootstrapping hypothesis and would even hint that infants can use prosody to acquire embedded structures – clauses within sentences. A null result, in which 2-clause participants did not discriminate, would be difficult to interpret, since it could indicate that participants did not infer that the clauses ABC and DEF are in a meaningful relationship, or it could indicate that participants interpreted the clauses as possible complete sentences. For example, in (3), the clause "Colin likes to read the news" is a perfectly legal English sentence.

(3) During the election, Colin likes to read the news.

While the potential experiments outlined in this section would provide further evidence, this dissertation puts forth the first results suggesting that infants can infer the constituent status of a string of words solely from the prosodic features of the input. However, the strongest version of prosodic bootstrapping – phrase structure – has yet to be tested by the experiments presented here or elsewhere. According to this proposal, infants can not only infer that prosodically-grouped words are constituents, but also that those constituents are part of a hierarchically-organized phrase structure grammar. An early experiment by Morgan, Meier, and Newport (1987) provides initial evidence that adults may be able to use prosody as a cue to hierarchically-embedded structures, but the results are not clear. Langus, Marchetto, Bion, and Nespor (2012) found that adults were able to track two levels of prosodic cues (final-syllable lengthening as a cue to PhP boundaries and pitch declination/resets as a cue to IP boundaries), hinting that adults can use prosody to learn embedded structures. However, an alternative explanation presented in Section 2.2.3 suggests that the participants were not necessarily learning embedded structures (the test phase required them to choose between strings that introduced additional variability in non-boundary words versus strings that grouped phrases into sentences that did not occur during familiarization).

Therefore, while there is some initial support for the idea that adults can use prosody as a cue in the acquisition of hierarchically-organized phrase structure, more decisive tests are needed. For infants, there is, to date, no evidence that they can use prosody for anything beyond locating constituents. Future studies can address this question by asking if learners can use prosody to learn an embedded structure. For example, if a learner can use prosody to find subject and predicate constituents, and – at the same time – to find the clause that contains these constituents, it would be evidence that prosody is an effective cue to hierarchically-organized structures.

5.2 Summary and conclusions: Mechanisms underlying prosodic bootstrapping.

While the experiments in Chapter 3 provide evidence that prosody is useful in syntax acquisition, the mechanism underlying this type of prosodic bootstrapping effect is not well understood. In Chapter 1, I introduced two possibilities. First, according to the general acoustic hypothesis, prosodic bootstrapping occurs because pitch cues, pauses, and changes in duration are acoustically salient. The fact that newborns, cotton-top tamarins, and rats have all been found to discriminate between different languages based on their prosodic features evinces this acoustic salience. Conversely, the language-specific hypothesis argues that prosodic bootstrapping relies on acquired knowledge of the interactions between prosody and syntax in the target language. Pitch, for example, is manifested differently in Mandarin and English. In Mandarin, it conveys lexical information about the meaning of a word, whereas it conveys sentence-level information, for example about whether the sentence question or statement, in English.

The experiments in Chapter 4 tested the general acoustic and language-specific versions of the prosodic bootstrapping hypothesis by asking if infants and adults can use non-native prosodic information to locate groups of words that they treat like syntactic constituents, as they did with native prosody in Experiments 1 and 2. English-acquiring American infants and English-speaking adults were tested on sentences with Japanese-like prosody and Japanese infants were tested on sentences with English-like prosody. English and Japanese were chosen as the target languages because they differ along a number of prosodic dimensions that influence the manifestation of clause-level prosody. Therefore, listeners with English experience should have formed different expectations about what constitutes clause-like prosody than listeners with Japanese experience.

Japanese and American infants (Experiments 3 and 4, respectively) from the 2-clause familiarization groups showed significant discrimination – albeit with different directions of preference – between sentences in which prosodically-marked groups of words moved versus sentences in which strings of words that crossed a prosodic boundary moved. Coupled with the lack of discrimination in the 1-clause groups, these studies suggest that infants are willing to entertain non-native prosody as a useful cue in syntax acquisition. In Experiment 5, American English-speaking adults were asked to complete a similar learning task. While infants showed learning both with native and non-native prosody and adults showed learning with native prosody in Experiment 2, adults were not able to use non-native prosody for a syntactic task in Experiment 5, as shown by the lack of difference between the 2-clause and 1-clause (control) groups. This suggests that adults are less flexible in their ability to use non-native prosody for grouping speech into constituent-like units.

5.2.1 Alternative interpretations and possible follow-up studies

In the present experiments, the non-native stimuli were unfamiliar not only in their prosodic features, but also in the phonotactics and phonemes of the stimuli words. This was done in order maintain naturally-produced prosody. It is not possible for a speaker to produce sentences with Japanese prosody and English phonology or English prosody and Japanese phonology. Even a completely balanced Japanese-English bilingual would likely be unable to simultaneously access prosody from one language and phonology from another. For the infants this did not seem to have any effect. English-acquiring American infants (the only infants tested with both native and non-native stimuli) had nearly identical results in Experiment 1 (listening to English) and Experiment 4 (listening to Japanese).

However, the performance of American English-speaking adults did differ based on whether they were listening to English or Japanese stimuli. It is therefore possible that the lack of familiarization by test condition interaction in Experiment 5 was due to a problem processing non-native phonotactics and phonemes rather than a problem processing non-native prosody. This could be teased apart by testing adult participants on sentences with Japanese prosody and English phonology and sentences with English prosody and Japanese phonology. While these stimuli could not be recorded naturally, they could be synthesized from the English and Japanese stimuli used for the experiments in this dissertation.

If adults in the 2-clause condition listening to Japanese prosody and phonology in Experiment 5 failed to show learning beyond the participants in the 1-clause control condition because of difficulties interpreting non-native prosody, they should be able to locate clausal constituents from stimuli with English prosody and Japanese phonology, but not from stimuli with Japanese prosody and English phonology. On the other hand, if the adults in Experiment 5 were distracted by the non-native phonotactics and phonemes in the Japanese stimuli, the opposite pattern of results should obtain: participants should be able to use prosody to locate constituent-like units if the stimuli have English phonology and Japanese prosody, but not the other way around. This experiment would provide stronger evidence about the role of experience in shaping how adults interpret prosodic information.

Together, the experiments in this dissertation provide the first evidence that learners can use prosody for a truly syntactic task – parsing speech into cohesive, moveable, acoustically-variable syntactic constituents. Infants can do this equally well with familiar, native prosody and unfamiliar, non-native prosody. Adults, on the other hand, depend largely perceptually-prominent boundary words and only show evidence of syntactic learning when listening to sentences with native prosody and phonology. The experiments outlined in this chapter illustrate the logical next steps in examining prosody's role in syntax acquisition.

REFERENCES

Aasland, W. A., & Baum, S. R. (2003). Temporal parameters as cues to phrasal boundaries: A comparison of processing by left-and right-hemisphere brain-damaged individuals. *Brain and Language*, *87*, 385-399.

Apple, W., Streeter, L.A., Krause, R.M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology, 37*, 715-727.

Baayen, R.H. (2011). languageR: Data sets and functions with "Analyzing linguistic data: A practical introduction to statistics." R package version 1.4.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (under review). Random-effects structure for confirmatory hypothesis testing: Keep it maximal. Manuscript under review. Retrieved October 14th, 2012 from www.psy.gla.ac.uk/docs/download.php?type=PUBLS&id=2301.

Bates, D., Maechler, M., & Bolker, B. (2012). lmer4: Linear mixed-effects models using S4 classes. R package version .999999-0.

Beckman, M.E. (1986). *Stress and non-stress accent (Netherlands Phonetics Archives No. 7)*. Dordrecht, Netherlands: Foris Publishing.

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. (1986a). Intonational structure in Japanese and English. *Phonology Yearbook, 3*, 255-309.

Beckman, M.E. & Pierrehumbert, J.B. (1986b). Japanese prosodic phrasing and intonation synthesis. In *Proceedings of the 24th ACL* (pp. 173-180). New York.

Bernstein Ratner, N. (1986). Durational cues which mark clause boundaries in mother-child speech. *Journal of Phonetics, 12*, 245-254.

Bion, R. A., Benavides-Varela, S., & Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech*, *54*, 123-140.

Blake, J. & de Boysson-Bardies, B. (1992). Patterns in babbling: A cross-linguistic study. *Journal of Child Language, 19*, 51-74.

Boersma, P. & Weenink, D. (2008). Praat: doing phonetics by computer (Version 5.0.35) [Computer program]. Retrieved October 16, 2008, from http://www.fon.hum.uva.nl/praat/.

Broen, P.A. (1972). The verbal environment of the language-learning child. *Monograph of American Speech and Hearing Association, 17*, 1-103.

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

Carnie, A. (2002). *Syntax: A generative introduction*. Cornwall, England: Blackwell Publishing.

Caudill, W. & Weinstein, H. (1969). Maternal care and infant behavior in Japan and America. *Psychiatry: Journal for the Study of Interpersonal Processes, 32*, 12-43.

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

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *The Journal of the Acoustical Society of America*, *95*, 1570.

Christophe, A., Mehler, J., & Sebastián-Gallés, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy*, *2*, 385-394.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language*. New York: Harcourt Brace Jovanovich.

Collins, B. & Mees, I.M. (1981). *The phonetics of English and Dutch*. Leiden, Netherlands: Brill.

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

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

Cutler, A. & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 113-121.

Eady, S.J., Cooper, W.E., Klouda, G.V., Mueller, P.R., & Lotts, D.W. (1986). Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments. *Language and Speech, 29*, 233-251.

Elbers, L. (1982). Operating principles in repetitive babbling: A cognitive continuity approach. *Cognition, 12*, 45-63.

Epstein, W. (1961). The influence of syntactical structure on learning. *The American Journal of Psychology, 74*, 80-85.

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development, 64*, 657-675.

Fernald, A. & McRoberts, G. (1996). Prosodic bootstrapping: A critical analysis of the argument and the evidence. In J.L. Morgan & K. Demuth (Eds.), *Signal to syntax* (pp. 365-388). Mahway, NJ: Erlbaum.

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.

Fisher, C. & Tokura, H. (1996a). Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. *Child Development, 67*, 3192-3218.

Fisher, C. & Tokura, H. (1996b). Prosody in speech to infants: Direct and indirect acoustic cues to syntactic structure. In J.L. Morgan & K. Demuth (Eds.), *Signal to syntax* (pp. 343-363). Mahway, NJ: Erlbaum.

Forster, K.I. (1966). Effect of syntactic structure on nonordered recall. *Journal of Verbal Learning and Verbal Behavior, 5*, 292-297.

Friederici, A.D., Friedrich, M., & Christophe, A. (2007). Brain responses in 4-month-old infants are already language specific. *Current Biology, 17*, 1208-1211.

Gambell, T. & Yang, C. (2006). *Word segmentation: Quick but not dirty*. Unpublished manuscript.

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

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

Gerken, LA. (1996). Prosodic structure in young children's language production. *Language, 72*, 683-712.

Gerken, L.A., Jusczyk, P.W., & Mandel, D.R. (1994). When prosody fails to cue syntactic structure: Nine-month-olds' sensitivity to phonological vs. syntactic phrases. C*ognition, 51*, 237-265.

Gerken, L.A., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language, 32*, 249-268.

Gleitman, L.R. & Wanner, E. (1982). Language acquisition: 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.

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.

Hayashi, A. & Mazuka, R. (2002). *Developmental change in infants' preferences for continuousness in Japanese speech*. Paper presented at the 13th Biennial International Conference on Infant Studies, Toronto.

Hayashi, A., Tamekawa, Y., Deguchi, T., & Kiritani, S. (1996). Developmental change in perception of clause boundaries by 6- and 10-month-old Japanese infants. *Spoken Language, 1996. ICSLP 96. Proceedings, Fourth International Conference on Spoken Language Processing, 13*, pp. 1565-1568. 3-6 Oct 1996.

Hayes, B. (1989). The prosodic hierarchy in meter. In P. Kiparsky and G. Youmans (Eds.), *Phonetics and Phonology, Vol 1: Rhythm and Meter*. San Diego: Academic Press, pp. 201-260.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

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.

Houston-Price, C. & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development, 13*, 341-348.

Hunter, M.A. & Ames, E.W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research, 5*, 69-95.

Johnson, E.K. & Seidl, A. (2008). Clause segmentation by 6-month-old infants: A crosslinguistic perspective. *Infancy, 13*, 445.

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

Joshi, A.K. & Levy, L.S. (1982). Phrase structure trees bear more fruit than you would have thought. *American Journal of Computational Linguistics, 8*, 1-11.

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. (1997). *The di~’covery of spoken language*. Cambridge, MA: MIT Press.

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

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

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

Kawahara, S. & Shinya, T. (2008). The intonation of gapping and coordination in Japanese: Evidence for intonational phrase and utterance. *Phonetica, 65*, 62-105.

Kemler Nelson, D.G., Hirsh-Pasek, K., Jusczyk, P.W., & Wright-Cassidy, K. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language, 16*, 55-68.

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.

Klatt, D. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America, 59*, 1208-21.

Ladd, D.R. (2009). *Intonational phonology* (2nd ed.). Cambridge, MA: Cambridge University Press.

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.

Lehiste, I. & Fox, R.A. (1992). Perception of prominence by Estonian and English listeners. *Language and Speech, 35*, 419-434.

Lehiste, I. & Fox, R.A. (1993). Influence of duration and amplitude on the perception of prominence by Swedish listeners. *Speech Communication, 13*, 149-154.

Levy, L.S. & Joshi, A.K. (1978). Skeletal structural descriptions. *Information and Control, 39*, 192-211.

Lust, B., & Chien, Y. C. (1984). The structure of coordination in first language acquisition of Mandarin Chinese: evidence for a universal. *Cognition*, *17*, 49-83.

Maclay, H. & Osgood, C.E. (1959). Hesitation phenomena in spontaneous English speech. *Word, 15*, 19-44.

Mampe. B., Friederici, A.D., Christophe. A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology, 19*, 1994-1997.

Mandel, D.R., Jusczyk, P.W., & Kemler Nelson, D.G. (1994). Does sentential prosody help infants organize and remember speech information? *Cognition, 53*, 155-180.

Mandel, D.R., Kemler Nelson, D.F., & Jusczyk, P.W. (1996). Infants remember the order of words in a spoken sentence. *Cognitive Development, 11*, 181-196.

Marantz, A. (1984). On the nature of grammatical relations. *Linguistic Inquiry Monographs Cambridge, MA.*, (10), 1-339.

Mazzoni, D. (2009). Audacity. [Computer program]. Retrieved March 28, 2009, from http://audacity.sourceforge.net.

Mazuka, R. (1996). Can a grammatical parameter be set before the first word? Prosodic contributions to early setting of a grammatical parameter. In J.L. Morgan & K. Demuth (Eds.), *Signal to syntax* (pp. 313-330). Mahway, NJ: Erlbaum.

McCawley, J. D. (1968). *The phonological component of a grammar of Japanese* (No. 2). The Hague: Mouton.

Miller, G.A. & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior, 2*, 217-228.

Millotte, S., Morgan, J.A., Margules, S., Bernal., S., Dutat, M., & Christophe, A. (2011).

Phrasal prosody constrains word segmentation in French 16-month-olds. *Journal of Portuguese Linguistics, 9*, 67-86.

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

Morgan, J.L. (1996). Prosody and the roots of parsing. *Language and Cognitive Processes, 11*, 69-106.

Morgan, J.L., Meier, R.P., & Newport, E.L. (1987). Structural packaging in the input to language learners: Contributions of prosodic and morphological 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.

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

Morgan, J. L., Swingley, D., & Miritai, K. (1993). Infants listen longer to speech with extraneous noises inserted at clause boundaries. Paper presented at the Biennial Meeting of the Society for Research in Child Development, New Orleans, LA.

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., Shukla, M., Van de Vijver, R., Avesani, C., Schraudolf, H., & Donati, C. (2008). Different phrasal prominence realizations in VO and OV languages. *Lingue E Linguaggio, VII.2*, 1-29.

Nespor, M. & Vogel, I. (2007). *Prosodic phonology* (2nd ed). Dordrecth, Holland: Forris Press.

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

Nishinuma, Y. (1979). *Un modèle d'analyse automatique de la prosodie : Accent et intonation en japonais*. Paris: Éditions du Centre National de la Recherche Scientifique.

Pannekamp, A., Toepel, U., Alter, K., Hahne, A., & Friederici, A. D. (2005). Prosody-driven sentence processing: An event-related brain potential study. *Journal of Cognitive Neuroscience*, *17*, 407-421.

Pannekamp, A., Weber, C., & Friederici, A. D. (2006). Prosodic processing at the sentence level in infants. *Neuroreport*, *17*, 675-678.

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

Pilon, R. (1981). Segmentation of speech in a foreign language. *Journal of Psycholinguistic Research*, *10*, 113-122.

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

R Core Team (2012). R: A language and environment for statistical computing. R. Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, *288*, 349-351.

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

Sato, Y. (1993). The durations of syllable-final nasals and the mora hypothesis in Japanese. *Phonetica, 50*, 44-67.

Schmitz, M. (2009). *The Perception of Clauses in 6- and 8-Month-Olds.* (Doctoral dissertation, University of Potsdam).

Schmitz, M., Hohle, B., & Weissenborn, J. (2003). *How pause length influences the perception of major syntactic boundaries in 6-month-old German infants.* Poster presented at the conference of Generative Approaches to Language Acquisition (GALA), Groningen.

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

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

Seidl, A. & Hollich, G. (2002). Infants' and toddlers' comprehension of subject and object wh-questions. *Proceedings of the Annual Boston University Conference on Language Acquisition, 26*, 597-607.

Selkirk, E. (1996). The prosodic structure of function words. In J.L. Morgan & K. Demuth (Eds.), *Signal to syntax* (pp. 187-213). Mahway, NJ: Erlbaum.

Selkirk, E. (2009). On clause and intonation phrase in Japanese: The syntactic grounding of prosodic constituent structure. *Gengo Kenkyu, 136*, 35-73.

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

Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, *72*, B11-B21.

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Neuroscience, 54*, 1-32.

Snow, D., & Balog, H. L. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua*, *112*, 1025-1058.

Soderstrom, M., Blossom, M., Foygel, R., & Morgan, J.L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language, 35*, 869-902.

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

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

Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and Language, 86*, 142-164.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2*, 191-196.

Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *The 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.

Takahashi, E. & Lidz, J. (2007). Beyond statistical learning in syntax. *Proceedings of Generative Approaches to Language Acquisition.*

Takeda, K., Sagisaka, Y., & Kuwabara, H. (1989). On sentence-level factors governing segmental duration in Japanese. *Journal of the Acoustical Society of America,86*, 2081-2096.

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

Toro, J. M., Trobalon, J. B., & Sebastián-Gallés, N. (2003). The use of prosodic cues in language discrimination tasks by rats. *Animal Cognition*, *6*, 131-136.

Tremblay, A. & Ransijn, J. (2012). LMERConvenienceFunctions. R package version 1.6.8.3.

Venditti, J.J., Jun, S.A., & Beckman, M.E. (1996). Prosodic cues to syntactic and other linguistic structures in Japanese, Korean, and English. In J.L. Morgan & K. Demuth (Eds.), *Signal to syntax* (pp. 287-311). Mahway, NJ: Erlbaum.

Venditti, J.J. & Yamashita, H. (1994). Prosodic information and processing of temporarily ambiguous constructions in Japanese. In *Proceedings of the 1994 International Conference on Spoken Language Processing (ICSLP), 3*:1147-1150. Yokohama, Japan.

Wakefield, J. A., Doughtie, E. B., & Yom, B. H. (1974). The identification of structural components of an unknown language. *Journal of Psycholinguistic Research*, *3*, 261-269.

Wexler, K., & Culicover, P. (1980) *Formal principles of language acquisition*. Cambridge: MIT Press.

Yoshida, K.A., Iverson, J.R., Patel, A.D., Mazuka, R., Nito, H., Gervain, J., & Werker, J.F. (2010). The development of perceptual grouping biases in infancy: A Japanese-English cross-linguistic study. *Cognition, 115*, 356-361.