

The Relationship Between Infant-Directed Prosody and Indices of Lexical Acquisition at 15 Months of Age

By Karla Gilbride
Swarthmore College

Abstract

The prosodic features of infant-directed speech are described, and several accounts of potential facilitative effects of such distinct prosodic features for the infant's socioemotional and/or linguistic development are discussed. One of these accounts--that infant-directed prosody facilitates word learning by exaggerating the perceptual salience of novel words in the discourse--is then evaluated through an analysis of two pairs of audio samples taken from the corpora of the Child-Directed Speech Project of the Language Science Research Group based at Washington University at St. Louis. The samples selected were recordings of naturalistic interactions between two mothers and their infants when the infants were nine-and-a-half and 15 months of age. The global prosodic characteristics of the mothers' speech in both the nine-and-a-half- and 15-month samples were assessed and compared to the level of productive language demonstrated by the infants at 15 months, an analysis which yielded no significant effects. In addition, based on previous reports that mothers highlight new and/or semantically focused words in speech to infants by placing such words on utterance-final pitch peaks, the pitch peaks of 40 maternal utterances in each 15-month sample were analyzed to determine how concentrated they were at the ends of utterances and how often they signaled emphatic stress of semantically focused information. Pitch peaks were found to coincide with emphatically stressed words much more often if they occurred utterance-finally than if they occurred elsewhere. The hypothesis that infants would display heightened sensitivity to pitch-peaked and utterance-final words was also evaluated by calculating how many of the infants' words in the 15-month sample had been previously spoken by their mothers in isolation or utterance-final position or on either focally stressed or non-focally stressed pitch peaks. Strong correlations were found to exist with all input variables except for non-focally stressed pitch peaks, although several of the infants' words had not been previously spoken by the mothers at all. These findings are then analyzed with respect to their implications for the relationship between the prosodic characteristics of infant-directed speech and the process of lexical acquisition. Limits on the generalizability of these results and suggestions for future research are also discussed.¹

¹ I would like to thank Kari Swingle in the Swarthmore linguistics department and Ed Kako in the Swarthmore psychology department for reading earlier drafts of this thesis and providing valuable feedback and guidance. I am also extremely grateful to Sean Crist, also of the Swarthmore linguistics department, for introducing me to the PRAAT program and showing me how to use it to conduct instrumental analyses of my data. Finally, the experimental portion of this paper would not have been possible without Michael Brent of Washington University at St. Louis, who pointed me to his project's website in response to a request for digital audio files of young children and their parents that I posted on the info-childes mailing

1. Introduction

Advanced knowledge of linguistics is not required to recognize the fact that when most adults talk to infants or young children, their speech is markedly different from the speech they address to other adults. Some of the characteristics of this specialized speech register, referred to as infant-directed speech or Motherese, are syntactic, semantic or phonological in nature. For example, speech addressed to infants tends to consist of shorter utterances and morphologically simpler, often monosyllabic, words as compared to speech directed to adults. It also contains more repetitions and expansions (Cooper and Aslin 1990), possesses clearly differentiated vowels which overlap minimally in their formants (Fernald 1991), and is less likely to undergo phonological reduction of consonants than adult-directed speech (Bernstein-Ratner 1984b).

The vast majority of the properties that set infant-directed speech apart from other registers, however, have to do with prosody, a relatively sparsely studied branch of linguistics pertaining to the suprasegmental characteristics of language. These suprasegmental characteristics include fundamental frequency or pitch, amplitude, (subjectively experienced as loudness), rhythmic properties such as tempo and segmental durations, and relative prominence, often described as stress, of the various elements of an utterance.

Speech directed to infants has been found to differ consistently on several prosodic dimensions from speech directed to adults. Moreover, researchers have noted strikingly similar prosodic features in Motherese as it is spoken by men and women, and the register has been documented in at least nine different languages to date (Werker, Lloyd and Pegg 1996, Fernald, Taeschner, Dunn, Papousek, deBoysson-Bardies and Fukui 1989, Eisen and Fernald 1991). This widespread evidence for a distinct infant-directed speech register begs the question of why so many people from so many different linguistic backgrounds should alter their speech in precisely the same way when addressing infants. Are we biologically predisposed as humans to adopt certain canonical prosodic patterns whenever we interact with an infant, and if so, what adaptive function might such canonical patterns serve?

This paper will explore the potential functions of infant-directed prosody,

specifically its possible social and emotional uses, its experimentally demonstrated ability to attract and retain infants' attention, and the role it has been suggested to play in helping children to segment the speech stream into either syntactic constituents or lexical units. After reviewing the existing literature concerning the theoretical bases for and evidence supporting each of these claims regarding the developmental functions served by infant-directed prosody, the paper will describe a study that I conducted in order to assess the predictive power of one of these claims, namely, that the prosodic characteristics of infant-directed speech facilitate lexical acquisition. This claim was assessed by analyzing sound files made from recordings of naturalistic interactions between two mothers and their infants when the children were nine-and-a-half and 15 months of age. Certain overall prosodic tendencies in the mothers' speech at both infant ages were isolated and compared against the infants' productive language at 15 months to determine whether any broad correlation could be discerned between the magnitude of specific prosodic variables in the input and the rate of lexical acquisition. A more fine-grained analysis was also conducted on the usefulness of pitch peaks within maternal utterances, especially those occurring in utterance-final position, as cues to novel or semantically focused information and the degree to which infants were more likely to produce words previously stressed by their mothers than to utter words which their mothers had used in a nonstressed position or had not previously spoken at all. This longitudinal analysis is intended, albeit in a very preliminary and limited way, to address the question of whether the correlation between infant-directed prosody and word learning is not merely theoretically plausible but objectively demonstrable.

2. Infant-Directed Speech as a Distinct Register

2.1 Prosodic features

Over the past 30 years, a number of studies have used both instrumental acoustic measurements and subjective judgments to isolate several intonational and durational characteristics which seem to typify infant-directed speech or IDS. On the intonational side, utterances addressed to infants tend to be spoken at a higher average pitch or fundamental frequency, (f_0), contain more pitch modulations, and traverse a greater total

f_0 range than adult-directed utterances. For example, while a female speaker saying “good morning” to another adult might begin the utterance at an f_0 value of around 220 hertz, (subsequently abbreviated hz), and rise to a peak of 235 hz on the second syllable “mor” before dropping down and leveling off at approximately 210 hz on the final syllable “ning”, the same woman addressing the same greeting to a six-month-old infant would tend to start the utterance somewhere around 300 hz, climb all the way to 350 hz on “mor” and then drop down steadily through the rest of the utterance so that the cessation of voicing at the end of “ning” would occur at an f_0 value of 200 hz. Pitch modulations such as these are especially likely to occur at the ends of ID utterances, with utterance-final words typically marked by either rising or rising-falling pitch contours. In fact, in contrast to adult-directed speech, where a process called declination causes pitch to drop steadily throughout most declarative utterances, one study found that 39% of English ID utterances and 45% of Japanese ID utterances actually ended in rising pitch (Ryan 1991).

In terms of durational factors, IDS tends to be spoken at a slower rate and to contain pauses which are both more frequent and significantly longer than those occurring in adult-directed speech or ADS. In fact, according to Fernald and McRoberts, (1996), the pauses between ID utterances are on average longer than the utterances themselves. In addition, utterance-final words, which, as mentioned previously, are marked by large f_0 excursions, also undergo significant vowel lengthening, with Bernstein-Ratner (1984) reporting that mothers lengthen vowels preceding utterance-final voiced consonants twice as much in IDS as in ADS. Morgan (1986) extended these findings by showing that in stories read to toddlers, words that occurred at the ends of syntactic phrases underwent greater vowel lengthening and traversed wider f_0 ranges than words in other sentential positions.

2.2 Universality of ID prosody

The prosodic characteristics described above, most notably elevated pitch, increased pausing, and widened f_0 range as compared with adult-directed speech, have been documented in speech to children learning a variety of languages, including English, French, German, Italian, Japanese, and Spanish (Werker et al 1996, Fernald et al 1989). These features have also been observed in speech directed to infants learning

Mandarin(Grieser and Kuhl 1988), Cantonese(Werker and McLeod 1993), and Xhosa, a language of southern Africa(Eisen and Fernald 1991). The evidence from the latter three languages is especially interesting because they are all tonal languages where pitch can influence semantic content, and it is possible that maintaining the exaggerated pitch modulations of IDS in these languages could compromise intelligibility.

Several researchers have challenged the claim that these attributes of ID prosody are linguistically universal. Bernstein-Ratner and Pye, (1984), for example, found that mothers in a Quiche Mayan speech community in Guatemala spoke with slightly lower pitch to their infants than to other adults, and Schieffelin (1979) reported that the Kiluli of New Guinea do not speak to their infants at all during the first year or so of their lives. The apparent contradiction suggested by this data may be nullified, however, by taking into account that the mean absolute f_0 of the speech directed to the Quiche Mayan infants was actually higher than that observed in many samples of English IDS and that the relatively lower pitch of Quiche Mayan IDS than ADS may be attributable to the fact that heightened pitch is used to signal deference and respect in Quiche Mayan(Fernald 1991). Furthermore, while Kiluli adults do not speak to infants per se, they often “speak for” them in exaggeratedly high-pitched voices(Fernald 1991). Thus, in both of these linguistic communities infants are frequently exposed to relatively high-pitched speech and so may be able to reap whatever sorts of benefits a high-pitched speech register affords to preverbal infants even if that register is not addressed to them in the same way that it is in German or Mandarin.

Even if some exceptions do exist, the fact remains that characteristic ID prosody has been found to occur in the overwhelming majority of languages in which speech to infants has been formally studied. Many of these same studies have also revealed that ID prosody is used by both men and women from a range of socioeconomic backgrounds(Fernald et al 1989). IDS also appears to be a relatively persistent fixture throughout the first years of life, as evidenced by Stern, Spieker, Barnett and Mackain, (1983), who studied the speech of American mothers to their infants between birth and two years of age and reported that characteristic ID prosody remained prevalent throughout this period, although the magnitude of certain parameters did change over the course of the infants’ development and the most exaggerated use of all aspects of ID

prosody was found to occur when infants were four months of age.

These findings about the widespread regularities of ID prosody raise the question of whether adults are predisposed to speak in a certain way to young children, and, in turn, whether such a predisposition implies that this speech register is in some way adaptive or helpful for the developing child. Additionally, evidence that the prosodic features of IDS vary with the age of the addressee suggests that this specialized speech may serve different functions at different stages of development and that caregivers may adjust certain parameters of their speech over time, either consciously or unconsciously, in order to highlight those functions of IDS which are currently most beneficial and relevant to the child.

3. Possible Functions of ID Prosody

Three major classes of potential functions of ID prosody have been posited in the psychological and linguistic literature: social/emotional, attentional, and linguistic. These potential functions are by no means mutually exclusive, and in fact most researchers believe that multiple functions coexist and overlap in their applicability during development, with social/emotional functions predominating during early infancy and linguistic functions gaining importance towards the end of the first year of life. Though the experimental section of this paper will focus on a particular type of linguistic function which has been postulated for ID prosody, the main premises of and evidence for each of the three classes of potential functions will first be reviewed in order to give a sense of the range of current theories on the subject.

3.1 Social/Emotional functions

The class of social/emotional functions that have been claimed for ID prosody can be divided into two distinct theories.

3.1.1 Affective preference for ID prosody

The first of these social/emotional theories simply asserts that ID prosody is more pleasant for infants to listen to than AD prosody and generates more positive emotional responses in young children. Support for this theory has come from a study conducted in

1989 by Werker and McLeod which found that both four-to-five-month-old and seven-to-nine-month-old infants shown a videotape of a woman speaking to another adult with either ID or AD prosody responded with more positive affect to the tape containing ID prosody. The magnitude of this effect was somewhat larger among the four-to-five-month-old subjects, suggesting that this affective bias may be more pronounced in early infancy.

According to this view, ID prosody is adaptive because it enables caregivers to elicit positive emotional responses from the infants with whom they interact. In addition, the register could also promote bonding and attachment by causing infants to associate their caregivers with the auditory stimulus that brings them so much pleasure.

Successful attachment and bonding requires a commitment by both parties, however, and a follow-up study by Werker and McLeod (1989) has advanced the notion that the prosodic characteristics of IDS may indirectly enhance adults' emotional responsiveness to infants as well. In the study, adult observers were asked to watch silent videotapes of the four-to-five-month-old infants participating in the study described above as they listened to either IDS or ADS recordings. The adults then filled out questionnaires about the infants they had watched. Interestingly, they rated the infants who had been listening to the IDS recordings as more pleasant, friendly, fun, likable and cuddly than the infants who had been listening to the recordings of ADS, although these observers were unaware of the type of speech being heard by the infants. Thus, by triggering behaviors that adults find appealing and that cause them to react with positive affect themselves, the affective preference of young infants for ID prosody may initiate a positive feedback loop in which both adult and infant act in ways which further the other's enjoyment of the interaction, thus ensuring that such interactions will continue to be sought out and maintained.

3.1.2 Appropriate emotional responses to specific melodic contours

The other theory regarding ID prosody's possible social/emotional functions is more complex. It holds that within the general framework of IDS, caregivers use different melodic contours to convey different emotional states or to communicate different types of messages and that infants are sensitive to these distinct intonational patterns. For

example, studies by Stern, Spieker and Mackain (1982) and Papousek (1987) have found that English, German, and Chinese mothers whose infants are not visually attending to them and who wish to elicit eye contact from them tend to address them with relatively short utterances, even by IDS standards, spoken with rising pitch. These attentionals or attention bids, as they are called in the literature, often consist of monosyllabic words, such as “look” or “hey” in English, and they also often make use of the infant’s name. In contrast, when caregivers and infants are already engaged in an interaction which the caregivers simply wish to sustain, they typically speak with a gradual rising-falling pitch contour, as when an adult holding an infant smiles at him and says, “Hi there!”, with the pitch peak occurring on the nucleus of the diphthong in “hi” and a slow decline in both f_0 and intensity taking place throughout the word “there”. The same bell-shaped melodic contour is also used in a number of languages to express approval to infants (Fernald 1992), such as when a father, seeing that his nine-month-old has successfully stacked the rings in a ring toy, exclaims, “Good job!”, with the pitch peak and subsequent fall coming on the elongated vowel of “job”. Prohibitions and statements of disapproval such as “no!”, on the other hand, tend to be spoken to infants at lower mean f_0 and to have narrower f_0 ranges than either attentionals or approvals and to have a relatively short, sharp falling pitch contour (Fernald 1992). Finally, comfort vocalizations addressed to infants such as, “it’s all right”, also tend to occur at relatively low f_0 and to occupy narrow pitch ranges, but they are on average longer in duration and lower in intensity than prohibitions. In addition, while comfort vocalizations, like prohibitions, often exhibit a falling melodic contour, they fall less abruptly and have a smoother, more legato quality (Fernald 1991).

Not only are different types of infant-directed vocalizations prosodically distinct, but their prosodic characteristics have also been shown to be sufficient cues, at least for mature listeners, to their emotional tone or communicative intent. In a 1989 study by Anne Fernald, for example, adults who were exposed to natural samples of IDS and ADS which had been low-pass filtered to remove all segmental information and who were then asked to identify each vocalization they heard as either an attention bid, an approval, a prohibition or an expression of comfort were much more accurate in classifying the ID than the AD vocalizations, regardless of their amount of prior experience with infants.

Although it is difficult to tell whether preverbal infants are also aware of the meanings signaled by these distinct melodic contours, some evidence does exist that they respond to these contours with differential and appropriate affect. Hagelan and Luka (1987) found, for instance, that ten-week-old infants presented with their mothers' facial and vocal expressions of happiness, sadness and anger responded with matching facial affect in the happiness and anger conditions. Even more compelling is the finding by Fernald (1993) that five-month-old English-learning infants exposed to audio samples of unfamiliar women uttering approvals and prohibitions in either German, Italian, Japanese or English ID prosody or in English AD prosody responded with more positive affect to approvals and more negative affect to prohibitions in all conditions except for Japanese ID prosody and English AD prosody. The failure of these infants to respond differentially to the Japanese ID stimuli was explained by the fact that while Japanese approvals and prohibitions displayed the same melodic contours as the ID stimuli in the other languages, they possessed less overall pitch variability, a smoother amplitude envelope and a narrower f_0 range, thus causing them to fall short of some threshold of affective discriminability perhaps set in these subjects through habituation to the more exaggerated pitch excursions of English IDs (Fernald 1993;672).

Charles Darwin and other proponents of the nativist view of emotion would argue that infants respond with differential affect to vocal and facial expressions of different emotions because they are biologically hardwired to do so, in other words, that their "innate knowledge of expression" causes them to respond with "an instinctive sympathy" (Darwin 1872/1965;358). Alternatively, infants could simply find the rising pitch contours of attentionals to be intrinsically arousing, the rise-fall contours of approvals and attention maintenance utterances to be intrinsically pleasing, the sharp, falling contours of prohibitions to be intrinsically aversive, and the more gradually falling contours of comfort vocalizations to be intrinsically soothing. Support for this view comes from research showing that auditory stimuli, whether or not they are derived from human speech, which have the gradual rise time in intensity characteristic of ID approvals and attentionals elicit eye-opening and orientation responses in young infants while signals which have the more abrupt rise time in intensity characteristic of ID prohibitions elicit eye-closing and withdrawal (Kearsley 1973). Finally, signals which are

continuous at relatively low frequency and low intensity, like the comfort vocalizations of IDS, elicit heart rate deceleration and seem to have a soothing effect on young infants (Bench 1969). According to this account, infants initially respond to the approvals, comforts, and other types of speech addressed to them with the same affective biases that they bring to all auditory stimuli and only later, with additional experience and knowledge of language, come to associate the distinctive melodic contours of these utterance types with their different communicative messages. Whatever causes infants to respond differentially to different melodic contours in speech, the fact that they do so means that caregivers can use the distinctive prosodic features of ID attentionals, approvals, prohibitions, and comforts to regulate the affective responses of preverbal infants in situationally appropriate ways, thus constructing the foundations of a system of social interaction.

3.2 Attentional functions

A distinct though related class of theories regarding the possible functions of ID prosody emphasizes its ability to engage the attention of young infants. Numerous studies have demonstrated that infants will look longer at a stimulus in the operant head turn² or visual fixation³ procedures when doing so produces a recording of IDS than when looking results in a recording of ADS being played (Fernald 1985, Cooper and Aslin 1990, Werker et al 1996). These effects have been found in infants at one, two, five and a half, and seven and a half months of age and even in neonates tested at a mean age of 52 hours, ruling out the possibility that the longer looking times to IDS are a result of experience unless that experience is prenatal (Cooper and Aslin 1990). Given that the auditory damping caused by the abdominal wall particularly attenuates higher-frequency

²In this procedure, infants can control which of two auditory stimuli they hear by turning their heads towards one side of the room or the other. Usually, infants are seated in a sound-proof booth with a loudspeaker mounted on either side of them, and a head turn towards the loudspeaker on one side causes one recording to emanate from that loudspeaker whereas a head turn towards the speaker on the infant's other side causes the first recording to cease and a different recording to begin playing from the second loudspeaker. A significantly longer looking time to one side or the other is taken to indicate that the infant prefers the sound associated with a head turn to that side.

³This procedure is used in place of the operant head turn procedure in newborns and younger infants who lack sufficient motor control to make full head turns. Instead, one of two recordings is played based on which region of a checkerboard display an infant focuses his/her eyes on, and significantly longer looking times to parts of the checkerboard which produce recordings of one sound or the other are taken as evidence for the infant's preference for that sound.

components of sound like those that predominate in IDS while allowing lower-frequency components to pass through, however, and given the fact that pregnant women and others near them whose speech a fetus might be able to hear are unlikely to use ID prosody in their speech, it seems highly improbable that infants would develop a preference for characteristic ID prosody prenatally (Cooper and Aslin 1989).

In an attempt to determine which, if any, single prosodic feature of IDS is most responsible for the disproportionate attention paid to it by infants, Fernald and Kuhl (1987) generated three types of sine wave analogs of IDS and ADS by measuring the values of certain acoustic properties in natural ID and AD speech samples as they varied over time and then replicating these values in the sine wave analogs. One group of sine wave stimuli varied in accordance with the natural speech samples from which they were derived in terms of their fundamental frequency and speech rhythm but had their amplitude held constant. In a second group, amplitude and speech rhythm varied as in the natural speech samples but f_0 was held at a constant level. Finally, the third set of stimuli differed in accordance with the natural speech samples in their rhythmic characteristics but had both their amplitude and f_0 held constant. Four-month-old infants were then tested on pairs of ID and AD stimuli of each of these three types in an operant head turn procedure, and it was found that the only condition in which they looked significantly longer to either side was the one in which they were exposed to ID and AD “speech” that varied in f_0 and rhythmic characteristics with amplitude held constant. In this condition, the infants looked significantly longer at the side of the booth which produced recordings of the ID stimuli than at the panel associated with the stimuli derived from the ADS samples. These findings suggest that the wide variations in pitch associated with IDS may be the most attention-eliciting aspect of the register.

While many researchers have equated these longer looking times to the f_0 modulations of IDS with an infant preference for this type of speech, Anne Fernald offered an alternative account in an analysis of the different stimulus criteria that elicited positive affect and significantly longer looking times in her 1993 study of approvals and prohibitions. She explained that “infants attended more to stimuli with greater f_0 modulations regardless of whether they preferred these vocalizations in the same sense of liking them better. Wide-range vocal signals may simply be more compelling, just as

sirens are compelling without necessarily being pleasant to listen to”(Fernald 1993;670).

Regardless of whether the characteristic pitch contours of IDS are attractive or merely compelling, the fact remains that infants attend more to IDS than to ADS when given a choice. This fact leads almost inevitably to another question, however, the question of whether this elicitation of infant attention is an end in itself and thus fully explains the widespread use of ID prosody or whether the attention-eliciting function of IDS instead serves to promote some secondary purpose of the register. Such a secondary purpose could be social in nature, for by focusing the infant’s attention on the speaker as the source of the compelling vocalizations IDS may enhance social interaction and joint exploration of the environment. The secondary purpose of IDS could also be pragmatic, for in focusing attention on speech itself IDS could help the infant to recognize that speech is an important facet of social interactions and to begin to learn about how conversation as a collaborative enterprise works. For instance, Snow (1977) has argued that the high incidence of both pauses and interrogatives, as well as other utterances ending in rising f_0 , in IDS contributes to a sort of conversational simulation in which caregivers mark conversational turns with rising intonation and leave spaces for the infants to respond even before they are able to do so verbally. Finally, the secondary purpose of ID prosody could be a linguistic one, for if the prosodic characteristics of IDS direct infants’ attention to some portions of the speech stream over others, then IDS may help infants to focus on especially salient portions of the speech stream such as the boundaries between syntactic constituents or lexical items and could thus play an active role in the language acquisition process. Psychologists and linguists have dubbed the theory that ID prosody acts in such a way to facilitate language learning the “prosodic bootstrapping hypothesis”, and the two major variants of this hypothesis will be the focus of the following two sections.

3.3 Linguistic functions: syntax acquisition

One of the principal challenges that infants face in beginning to learn any language concerns how to segment the continuous acoustic signal that is speech into discrete meaningful units. In order to become a competent language user, an infant actually needs to learn to divide the same speech stream into several types of linguistic

units of varying sizes. S/he must figure out how to divide it into distinct morphemes, each of which has a unique meaning that s/he must also come to understand. Simultaneously, s/he must determine how to divide it into syntactic constituents such as noun phrases, verb phrases, prepositional phrases, and entire clauses. This syntactic bracketing of speech is essential to the infant's later task of forming and testing hypotheses about the syntactic categories and grammatical rules of his/her language, and an implicit understanding of these categories and rules is in turn necessary for the achievement of adultlike language production and comprehension. For example, in order to utter a well-formed proposition about a dog biting a man or to understand the difference in meaning between the statements, "The dog bit the man" and "The man bit the dog", a child must first recognize that NP's and VP's are distinct syntactic constituents which have different semantic functions and which must occur within clauses in certain set orders. S/he must also figure out that NP's contained within VP's play a different thematic role in those clauses than NP's which stand alone. Before a child can make any of these generalizations, however, s/he must first be able to realize that a sentence like, "The dog bit the man", can be broken up into several segments in a hierarchical manner, with the primary boundary occurring between "dog" and "bit", a secondary boundary located between "bit" and "the", and tertiary boundaries occurring between all the other words. One version of the prosodic bootstrapping hypothesis holds that prosodic features could serve as cues to the locations of syntactic boundaries, such as those between clauses or between NP's and VP's, for example, thus providing prelinguistic infants with an initial method of bracketing speech in syntactically appropriate ways and affording them access to the raw data that they can use over time to derive generalizations about the grammatical structure of their language. This claim of the sensitivity of prosody to syntactic constituent boundaries is based on evidence that in ADS, syllables occurring at the ends of phrases or clauses tend to be longer than syllables occurring in other utterance positions. In addition, pauses are more likely to occur between syntactic phrases in ADS than within them, and the relative lengths of pauses within an utterance tend to reflect the hierarchical syntactic structure of that utterance (Fisher and Tokura 1996). Several researchers have found similar effects of equal or even greater magnitude to occur in IDS.

3.3.1 Evidence for clausal and sentential boundaries

Some of this evidence pertains to the correlation between certain prosodic features and the boundaries between utterances or clauses in IDS. For example, while nearly half of the pauses observed in some ADS samples occurred within sentences (Butterworth 1980), 96% of the pauses 260 milliseconds in length or longer recorded in the spontaneous speech of three English and three Japanese mothers to their 13- to 14-month-old infants occurred at the boundaries of grammatically distinct utterances, which usually consisted of single clauses (Fisher and Tokura 1996). Furthermore, 59% of the grammatically distinct utterances in the English sample and 69% of those in the Japanese sample were separated by pauses of 260 ms or longer (Fisher and Tokura 1996). Another study, conducted by Fernald and Simon in 1984, found that the proportion of pauses which could be used as cues to sentence boundaries in a sample of German IDS was .98.

Not only pauses but vowel lengthening and pitch modulation as well can act as markers of utterance boundaries. In the Fisher and Tokura (1996) sample of English and Japanese IDS discussed above, for instance, prepausal vowels were found to be twice as long, on average, and to cover twice as wide a pitch range in both languages as the same vowels occurring in other utterance positions. Since 96% of the pauses in this sample immediately preceded utterance boundaries, these lengthened and melodically varied vowels were thus overwhelmingly likely to occur in the final syllable of an utterance, which also tended to be the final syllable of a clause or sentence. Remarkable on their findings, Fisher and Tokura maintained that “acoustic changes marking major prosodic boundaries are very robust, detectable in two phonologically distinct languages amid all the quirks and variability of casual spontaneous speech” (349).

3.3.2 Evidence for phrasal boundaries

There is also some evidence to suggest that in certain cases, prosodic variables can be used to identify the boundaries between smaller syntactic constituents such as NP's and VP's. For example, Fisher and Tokura (1996) found that when at least two syllables preceded an internal phrase boundary between an NP and VP, the syllable immediately preceding the boundary was reliably longer than the surrounding syllables in

the English sample and reliably lower in pitch in the Japanese sample. In the English sentence, “The boy caught the ball”, spoken to a 14-month-old infant, for instance, the syllable “boy” would tend to be significantly longer than either the preceding syllable “the” or the following syllable “caught”, with most of the elongation probably occurring in the nucleus of the diphthong. In addition, nine-month-old infants have demonstrated some sensitivity to prosodic markers of phrase boundaries, for in an operant head turn procedure infants at this age looked longer in the direction which produced recordings of sentences with pauses inserted at the NP/VP boundary, known as coincident stimuli, than in the direction which elicited noncoincident stimuli, or sentences with pauses inserted within the NP or VP (Hirsh-Pasek, Kemler-Nelson, Jusczyk, Cassidy, Druss and Kennedy 1987). Furthermore, this same preference was later found to persist even when the stimuli were low-pass filtered, ruling out the possibility that the infants were using emerging segmental knowledge to distinguish between the two types of sentences (Jusczyk et al 1992), yet this preference was only found for stimuli derived from IDS, with coincident and noncoincident ADS stimuli eliciting no differential responses from these infants (Kemler-Nelson, Hirsh-Pasek, Jusczyk and Cassidy (1989). These findings may not be as powerful or as generalizable to normal language acquisition as they first appear, however, as shall be discussed in the next section.

3.3.3 Limitations of the prosodic bootstrapping account for syntax acquisition

There are several significant problems with this account of prosody as an aid to syntax acquisition. For one thing, while pitch movement, vowel lengthening and pausing are consistent signals of the presence of an utterance boundary in IDS, the utterances that they divide are often not the sorts of canonical clauses that syntacticians typically analyze. In a sample of 100 ID utterances spoken by five American English mothers, for instance, more than 40% of the utterances were subclausal fragments such as, “A doggie!” or “Goes woof woof”, attentionals such as, “hey!”, or stock expressions such as, “Thank you very much”, many of which were even subphrasal (Fernald 1989). It is unclear how much or what type of useful syntactic information infants would be gaining if they were to use prosody to parse the speech stream into these sorts of utterances and then look for grammatical patterns among the resulting units. For that matter, even when

ID utterances are complete, well-formed clauses, Newport, Gleitman and Gleitman (1977) have pointed out that typical IDS contains a greater variety of sentence types than ADS, including imperatives and both “wh” and yes-no questions, and that all of these constructions are more syntactically complex than the declaratives which predominate in ADS. The presence of such varied and complex syntactic information in IDS would be apt to prolong and complicate the child’s task of forming generalizations about grammatical structure, even if prosodic characteristics could be reliably used to separate and identify syntactic constituents in the speech stream. Thus Newport, Gleitman and Gleitman’s observation could lead one to conclude that the achievement of certain pragmatic and communicative goals made possible by a large number of sentence types is a higher priority in speech to infants than syntactic simplicity and transparency.

Additional problems arise when the specific ability of prosody to signal internal phrase boundaries is considered. Despite Fisher and Tokura’s finding reported in the previous section regarding the lengthening or pitch lowering of the syllables immediately preceding NP/VP boundaries, the sorts of sentences for which this trend was found, that is, those whose subjects were at least two syllables in length, were the exception rather than the rule in their sample. In fact, most sentences in Japanese did not contain subject-predicate boundaries at all because of the frequency of phrase deletion in Japanese IDS, while in the English sample fully 84% of subject NP’s were pronominal. When discernible prosodic boundaries occur at all within sentences containing pronominal subjects, they tend to directly follow the main verb rather than preceding it, thus creating a mismatch between prosodic and syntactic structure.

Some experimental evidence may have documented the perceptual ambiguity resulting from this prosodic-syntactic mismatch. When 9-month-old infants were tested in an operant head turn procedure involving the sorts of coincident and noncoincident stimuli discussed in the preceding section, they demonstrated a longer looking time for utterances with pauses inserted at internal phrase boundaries when those utterances were declarative sentences with lexical subjects or inverted yes-no questions with pronominal subjects, but they demonstrated no looking preference at all when they were exposed to different pause placements in declarative sentences with pronominal subjects (Gerkin et al 1994).

These findings suggest that using prosody to locate the boundaries between subclausal syntactic constituents would be a much more difficult task for infants than proponents of the syntactic version of the prosodic bootstrapping hypothesis might have hoped. Indeed, the entire process of attempting to infer syntactic structure from the prosodic characteristics of IDS would be fraught with ambiguities and potential miscues, and this has led Gleitman, Newport and Gleitman (1984) to discredit the hypothesis. In that article, they claimed that innate biases of infant perception and attention rather than modifications in the input account for the speed and facility with which young children learn language and argued that infants themselves “are the prime movers of the acquisition process”(70).

3.4 Linguistic functions: lexical acquisition

Since all of these criticisms focus on the syntactic properties of ID utterances and the relationship between prosodic and syntactic boundaries, none of them preclude the possibility that prosodic properties of the input play a role in an aspect of language learning so basic as to be unaffected by any of these syntactic considerations: the development of a lexicon. The task of separating the speech stream into discrete lexical items is far from straightforward for infants, as adults who try to isolate the boundaries between words in an unfamiliar language can quickly infer, yet success in this task is an essential prerequisite for the first mappings of meaning onto sound which occur when infants begin to associate certain discrete combinations of sounds with objects in their environment. Later, in combination with their growing awareness of the syntactic attributes of their language, infants also use lexical knowledge to sort out which words belong to which syntactic categories and to reach a comprehensive understanding of the syntax-semantics interface. The lexical version of the prosodic bootstrapping hypothesis asserts that by using properties like temporal patterning, amplitude, and pitch movement to highlight certain focused or stressed words in an utterance, ID prosody may help infants to pick those words out from the surrounding speech stream and identify them as distinct lexical units, in effect guiding the infants’ first steps down the path of ever-accelerating lexical acquisition.

3.4.1 Focal stress

The phenomenon of focal or emphatic stress has been well-documented in the psychological and linguistic literature (Bolinger 1961, Halliday 1967, Chafe 1974, Aslin, Woodward, Lamendola and Bever 1996). This type of stress does not necessarily occur on words occupying those syntactic positions typically associated with primary sentential stress but is instead motivated by the semantic need to accent information which is novel with respect to the prior discourse. For example, in the question, “Where were you a year ago?”, primary sentential stress would usually fall on the word “were”. In a situation where two people had been discussing what various mutual friends were doing last year and one of them wanted to shift the focus of discussion to his conversational partner, however, he might produce this same sentence with emphatic stress on the word “you”, while if the same two people had been discussing their activities over the past three months and one of them wished to find out about his interlocutor’s more distant past, he could ask the same question while focally stressing the word “year”.

Focal or emphatic stress in ADS is generally marked by an increase in duration and amplitude and a rising or rising-falling f_0 contour on the word or words in question. While focally stressed words in speech to adults are usually merely new to the current discourse, words which are treated in a similar way in speech to infants may be new in the more absolute sense of having never been heard before or of having been heard but not yet learned. In light of this fact, and of the other prosodic differences distinguishing ID and AD speech, several researchers have sought to determine how focal stress is realized prosodically in IDS. Their results indicate that a combination of utterance position, utterance length, amplitude, and f_0 are used to mark new or focused words in speech to infants and that where the correlates of focal stress in IDS are analogous to those in ADS, the magnitude of these effects in IDS is considerably greater.

3.4.2 Temporal characteristics of focal stress in IDS

A number of studies have found that novel words tend overwhelmingly to be presented to infants in utterance-final position, even when constraints are placed on the experimental procedure in an attempt to discourage this tendency. In one such study, Woodward and Aslin (1989) had 19 English-speaking mothers try to teach three novel

nouns to their 12-month-olds. They found that the target words occurred utterance-finally in 89% of the multiword utterances that the mothers produced. Three years later, they conducted a variation on the same experiment in which they asked seven Turkish-speaking mothers to teach their 12-month-olds three novel nouns while instructing 20 English-speaking mothers to expose their infants of the same age to three novel transitive verbs. Since Turkish is a verb-final language and transitive verbs cannot occur utterance-finally in English, both groups would have had to violate grammaticality constraints in order to place the novel words in utterance-final position. This is precisely what these mothers did, with the English-speaking participants deleting subjects in one out of three of their utterances and deleting objects in one out of every six. Among the Turkish-speaking subjects, target nouns were equally split between utterance-medial and utterance-final position, where either utterance-initial or utterance-medial position would have been grammatical. Moreover, noun-final utterances were found to occur more than twice as often in this sample of Turkish IDS than in a sample of Turkish ADS (Woodward and Aslin 1992).

A similar preference for placing focused words in utterance-final position was found in a 1991 study by Fernald and Mazzie. In this study, 18 female subjects were presented with a picture book and asked to describe the events depicted in it both to their 14-month-old infants and to an adult experimenter. The book showed a child getting dressed in several stages, and each page featured a new item of clothing highlighted in a bright color. When the mothers pointed out these items of clothing, treated by Fernald and Mazzie as the target words, for the first time to their infants, they did so by placing the target words in utterance-final position 75% of the time. In contrast, when describing the same book to the adult experimenter, the first uses of the target words occurred utterance-finally only 53% of the time, a statistically significant difference (Fernald and Mazzie 1991).

Focal stress in IDS also tends to be correlated with the length of the utterance in which the stressed word appears. A 1996 study by Fisher and Tokura in which mothers described the action in a puppet show to their 14-month-old infants and to another adult revealed that the first mentions of new target words in the ID sample were more likely to occur within shorter utterances as compared either with the first mentions of target words

in the AD sample or second mentions of the same target words in the ID sample. This result prompted Fisher and Tokura to theorize that “shorter utterances provide simpler, less distracting carrier phrases for target words”(358).

3.4.3 Amplitude as a cue to ID focal stress

Increased amplitude has also been cited as a feature of focal stress in IDS, with Messer (1981) noting that in a corpus of speech gathered by asking mothers to show toys to their 14-month-olds, the labels of toys had a .47 probability of occurring on the amplitude peaks of the utterances in which they appeared. This study has been criticized, however, for not comparing the magnitude of this tendency to mark focused words with increased amplitude in IDS with that of the similar tendency observed in ADS(Fernald and Mazzie 1991). It may indeed be the case that amplitude is an equally robust cue to focal stress in both registers, for an amplitude analysis of the data from the puppet show experiment described in the previous section revealed that the first mentions of target words were significantly louder than their second mentions in both the ID and AD samples(Fisher and Tokura 1996). Simply because amplitude increases on focused words are not greater in IDS than in ADS, however, does not mean that the amplitude increases which do occur cannot help in drawing infants’ attention to the focused portion of an utterance.

3.4.4 Pitch as a correlate of emphatic stress

Finally, raised f_0 and increased f_0 movement, while correlated with emphatic stress in both IDS and ADS, have been found to be substantially more prevalent and extreme in IDS. In Fernald and Mazzie’s 1991 study involving picture book descriptions, for instance, target words were significantly more likely than any of the surrounding words to occur on the f_0 peaks of utterances in the ID sample, even when they were mentioned for the second time, while this trend was not nearly as pronounced in the AD sample. In addition, even when interrogatives were eliminated from consideration, 52% of the target words in the IDS sample occurred on utterance-final pitch peaks, more than twice as many as in a comparable sample of ADS(Fernald and Mazzie 1991). For that matter, some researchers have even speculated that the high proportion of interrogatives

in IDS, rather than serving a pragmatic or communicative function as posited by Snow (1977), instead reflects the use of a grammatical vehicle to further increase the number of focused words that combine the perceptually salient features of utterance-final position and relatively high pitch (Fernald and Mazzie 1991).

As well as occupying f_0 peaks, stressed words in speech to infants are also especially likely to occur in regions of pitch movement, with 68% of the ID utterances in Woodward and Aslin's 1989 study of English noun teaching techniques containing target words situated on either rising or falling pitch contours. Finally, Fernald and Mazzie (1991) found that 68% of first mentions of target words in their ID sample both occurred on f_0 peaks and were judged by linguistics graduate students to be the most stressed word in the utterance while only 28% of the first mentions of target words in the AD sample were both situated on pitch peaks and subjectively labeled as stressed, indicating that f_0 prominence is used more exclusively as a marker of stress in IDS while the cues to focal stress in ADS are more varied and complex.

3.4.5 Innate processing biases harnessed by prosodic bootstrapping

Taken together, these results seem to suggest that the prosodic features of IDS which are most attention-eliciting for infants are often concentrated around those lexical items which are new to the discourse or which the speaker wishes to treat as loci of semantic focus, a finding which may well have significant implications for word learning. Fernald and Mazzie have in fact suggested that "by positioning focused noun labels on exaggerated f_0 peaks at the ends of utterances, mothers speaking to infants may intuitively exploit the perceptual and attentional listening biases that make certain sounds much easier than others to detect, discriminate and remember"(218).

One of these perceptual and attentional biases may be the much-studied recency effect in memory which causes the last element in a series to be remembered more successfully than earlier elements. Another is documented in a finding by Watson (1976) that adults are better at discriminating and identifying relatively high-frequency tones occurring later in a tonal sequence than lower-frequency tones occurring earlier in the same sequence. In addition, infants have been shown to demonstrate a preferential orientation response to tones having a higher f_0 (Kearsley, cited in Bernstein-Ratner and

Pye 1984), and wide f_0 modulation has been shown to be the most attention-eliciting feature of ID prosody (Fernald and Kuhl 1987).

After making a case for the potential facilitative effects of ID prosody on word segmentation and learning, however, Fernald and Mazzie hasten to add that “the observation that some characteristics of infant-directed speech seem ‘well-suited’ for facilitating a developmental task such as lexical acquisition merely suggests hypotheses about function without providing evidence one way or another” (219). We therefore next turn our attention to some empirical tests of this particular variant of the prosodic bootstrapping hypothesis, both the few that have already been conducted and described in the literature and the one that I attempted to carry out by analyzing sound files of naturalistic interactions between two mothers and their nine-and-a-half- and 15-month-old infants.

4. Evidence for a Role of Prosody in Lexical Acquisition

4.1 Evidence from previous studies

Some preliminary evidence that prosodic bootstrapping does play a role in speech segmentation and vocabulary acquisition comes from a 1985 study by Karzon in which one- to four-month-old infants were tested in the high amplitude sucking procedure⁴ on their ability to discriminate two three-syllable nonsense words differing in only one of their segments, “marana” and “malana”. He found that the infants only responded differentially to the two stimuli when the medial syllable containing the differing segments was highlighted with ID prosody. Another study with older, (15-month-old), infants involved showing subjects two slides of familiar objects such as a ball or a shoe and then playing a recording of an adult saying things like, “Look at the ball! See the ball?” with either characteristic ID or AD intonation. The children looked reliably longer at the slide being described when the accompanying vocalization used ID prosody but not when it used AD prosody (Fernald and McRoberts 1991).

⁴ In the high amplitude sucking procedure, young infants are given a pacifier to suck on which is attached to a machine that measures the frequency and intensity of the infant’s sucks. The infant is then exposed to an auditory stimulus which is repeated until the infant habituates and the sucking rate decreases to baseline. Next, a new stimulus is introduced, and a change in the sucking rate in response to this new stimulus is

Previously published findings become even sparser when we move beyond perception and comprehension and look for correlations between prosodic properties of the input and later linguistic production by young children. Slobin (1982) has made the interesting and suggestive observation that children acquire inflectional morphemes at earlier ages in languages where these morphemes are encoded in isolable stressed syllables, and Echols and Newport (1992) found that infants at the one- and two-word stage were less likely to omit or mispronounce words which had previously been spoken to them by adults when the adult tokens of those words were either stressed or utterance-final.

A thorough and decisive analysis of the facilitative effect, or, on a stronger view of the prosodic bootstrapping hypothesis, the necessity, of ID prosody to any aspect of language acquisition would, however, seem to require one of two elaborate and time-intensive experimental designs. The experimenter would either need to also be a primary caregiver of an infant so that s/he could consciously manipulate his/her own speech to that infant in an effort to track the effects of this manipulation on later linguistic development, or an exhaustive longitudinal study would need to be conducted in which caregiver-child dyads where caregivers differed significantly in the prosodic characteristics of their IDS could be followed over time to determine whether these differences in input were correlated with any reliable differences in the infants' emerging linguistic capabilities. The difficulty of finding subjects who both satisfy these criteria and are willing to commit to such intrusive and time consuming experiments, the many methodological and ethical problems that these types of studies present, and the large number of subjects that would have to be studied in order to achieve any sort of meaningful or reliable results would make this a problematic area of research to say the least.

A more plausible potential source of data on this intriguing question already exists within the public domain, however, in the form of the recordings of spontaneous speech between parents and other adults and young children available on CHILDES and other online child language databases. In the analysis which follows, I made use of one such database, namely, a corpus of digital audio files and language transcripts of mothers

taken as evidence that the infant is able to discriminate between the pre-shift and post-shift sounds.

interacting with their infants that was compiled for the Child-Directed Speech Project of the Language Science Research Group of Washington University at St. Louis and made available over the Internet. Using this previously published data, I was able to construct my own longitudinal study across a six-month period in which I looked for correlations between the prosodic features of two mothers' IDS and the productive language development of their infants and, more specifically, at the reliability of f_0 peaks as cues to loci of emphatic stress in the mothers' speech and the prevalence of such previously focally stressed words in the speech of their infants.

4.2 Longitudinal analysis: methods

4.2.1 Subjects

My initial plan for this longitudinal analysis was to find a corpus containing speech samples of children and their caregivers beginning when the children were 10 or 12 months of age, a period when children produce few if any words of their own but may begin acquiring a receptive lexicon, i.e., a list of words whose meanings they understand. I reasoned that this would be a plausible stage in children's linguistic development for them to begin attending to and benefiting from the facilitative linguistic effects of ID prosody, if any such effects could be demonstrated. I then wanted to track both the prosodic input these children were receiving and the lexical acquisition they had achieved by 18 months of age, when children typically possess a productive vocabulary of between 50 and 75 words but significant individual differences in linguistic competence can be observed. Unfortunately, however, numerous Internet searches and e-mail correspondence with researchers in the child language field failed to unearth any digitized audio data of longitudinal studies spanning this particular age range, and all corpora containing speech samples of 18-month-olds did not begin until these children were several months past their first birthdays. Since I was concerned that the global prosodic features of the maternal speech in these corpora might be influenced by the infants' considerable linguistic knowledge, even in the earliest recordings, I decided to use data from a younger cohort of subjects obtained by Washington University at St. Louis' Language Science Research Group.

This group conducted a longitudinal study of language use in 15 infants and their mothers by providing mothers with taperecorders and asking them to record themselves interacting naturally with their infants for one hour every two weeks when the infants were between approximately nine and 15 months of age. The four audio files and linked transcripts that I used in this analysis were downloaded from the group's website at <http://lsrg.cs.wustl.edu/>. I was somewhat concerned that the relatively young age of these children at the time of the last recordings and their correspondingly small productive lexicons would make it difficult to reach any meaningful conclusions about the relationship between prosodic characteristics of the input and the infants' own linguistic behavior. Thus I chose two mother-child dyads from the corpus based on the single criterion that the infants produced at least ten distinct words or phrases during the last recording session included for that dyad. The first of the two subjects that I ultimately selected was a dyad that the LSRG website refers to as C1, a Caucasian female and her mother who were recorded on 13 occasions when the infant was between 9 months 3 days and 15 months 4 days old. This dyad will be labeled M-C1 in the analysis now being described, and the mother in this dyad will be referenced as M1 while her infant will be referenced as C1. The second dyad studied, an African-American male and his mother recorded on 11 occasions when the infant was between 9 months 25 days and 15 months 13 days old, was identified on the LSRG site as W3. For the purposes of this analysis, this dyad will be referred to as M-C2, and the mother and child in this dyad will be labeled M2 and c2, respectively. Two speech samples were analyzed for each dyad; for M-C1, the second recording, made at 9 months 17 days of infant age, and the last recording, made at 15 months 4 days of infant age, were used, while for M-C2, the first and last recordings, made at 9 months 25 days and 15 months 13 days, respectively, were subjected to analysis.

4.2.2 Prosodic analysis of maternal speech

Because elevated f_0 and wide pitch modulations have been identified as the most salient and attention-eliciting aspects of ID prosody for young children, (e.g., Fernald and Kuhl 1987), and because properties of the speech signal related to frequency lend themselves most readily to instrumental analysis, the prosodic analysis of the mothers'

speech in these samples was restricted primarily to three intonational dimensions, although impressionistic observations of temporal characteristics like utterance length and vowel elongation were also made. For each mother, I chose 20 child-directed utterances at random from each of the two speech samples and used PRAAT, a pitch tracking and acoustic analysis program, to create a spectrogram for each utterance. I then used a function included in the PRAAT software to calculate the mean f_0 of each utterance and averaged these values across the sampled utterances to determine the average pitch levels used by each mother in speaking to their infants at nine-and-a-half and 15 months of age. Next, the amount of pitch modulation in each of the four groups of utterances was assessed subjectively both by listening to the audio files and inspecting the spectrograms. Finally, in light of the fact, noted by Fernald and Mazzie (1991) and Woodward and Aslin (1989, 1992), that infants may have innate perceptual and memory biases favoring relatively recent, high-frequency information and that mothers may exploit these innate biases by highlighting novel and semantically focused words on utterance-final pitch peaks in their IDS, I inspected the spectrograms in PRAAT to determine what proportion of the sampled utterances from each recording contained f_0 peaks in utterance-final position. Thus, I analyzed a total of 80 maternal utterances in order to generate three distinct prosodic measures for each of the two mothers at each of the two infant ages: mean f_0 , amount of pitch modulation, and proportion of utterances ending on f_0 peaks.

4.2.3 Analysis of infants' productive language

The two infants' productive language capabilities at 15 months of age were assessed by counting both the total number of utterances and the number of distinct lexical items the infants produced during the hour-long 15-month sample. Multiword phrases such as "thank you" were counted as single lexical items, since these types of stock expressions are typically treated as unanalyzable units by children until they are substantially older than the children in this study.

4.2.4 Analysis of focal stress

A number of studies described in section 3.4.4 above have demonstrated that focally stressed words in speech to infants are significantly more likely than focally

stressed words in ADS to occur on the f_0 peaks of utterances, especially at the ends of those utterances. While these findings are interesting and certainly suggest an area for further research, there are two major problems with using this evidence to infer that infants use utterance pitch peaks to help segment the speech stream and pick out focally stressed words in fluent spontaneous speech. For one thing, all of the studies in question either involved explicit word teaching tasks or required parents to describe scenes to their infants that experimenters had deliberately “stacked” with new characters and actions in an effort to elicit frequent focal stress. Thus an open empirical question remains as to whether focal stress is nearly as prevalent in IDS in naturalistic settings as it was found to be in these experiments. A related problem with these results is encapsulated in a concept which Fernald and McRoberts, in their 1996 review of the prosodic bootstrapping literature, call “cue reliability”. They point out that while saying that mothers place 75% of focally stressed words on utterance-final pitch peaks in speech to infants may constitute convincing proof of the prevalence of this prosodic marking strategy for language researchers and lay adults, this probabilistic tendency will mean nothing to language-learning infants if 75% of the pitch peaks they hear are not cues to focal stress at all but stem from other phenomena such as sentential stress or the canonical melodic contours of IDS attentionals or approvals. In other words, what is important for infants is not that a given semantic or syntactic feature is reliably signaled by the same prosodic cue but that that prosodic cue reliably signals the presence of the same semantic or syntactic feature. Thus, in order for infants to begin mentally linking those segments of the speech stream that occur on the pitch peaks of utterances with the objects and people in their environments that are being talked about, it must be the case that the words highlighted on utterance f_0 peaks also have a strong tendency to correspond to the objects of semantic focus at that point in the discourse.

To my knowledge, no published studies have yet addressed these issues of the prevalence of focal stress in naturalistic IDS and the cue reliability of pitch peaks as markers of focal stress. I therefore undertook an analysis of these questions in the LSRG data by extracting a segment of each 15-month sample comprising 40 intelligible, multiword maternal utterances.⁵ I used PRAAT to locate the f_0 peak of each of these

⁵ One-word maternal utterances were not included in the 40-utterance count because identifying the word

utterances and identified the syllable or syllables occupying that f_0 peak. (No f_0 peak extended for longer than two syllables.) In the language transcripts for these two samples, I then marked the word or words which occurred on f_0 peaks.⁶ There were only four cases in which two adjacent monosyllabic words jointly occupied an utterance's f_0 peak and only two occasions where words at two different points in the utterance reached the same peak f_0 level. In these six utterances both affected words were marked as occupying pitch peaks and were treated independently as potential candidates for focal stress.

Next, I examined each of the 86 resulting target words in the context in which it was spoken and determined whether it had received emphatic stress. Though these judgments were ultimately subjective, I employed the same three criteria consistently across all candidates in reaching my decisions. First, I looked for any mismatch between the word occupying the pitch peak and the word which would occupy this position if the utterance were spoken out of context with only sentential stress patterns assigned. Second, I determined whether the target word seemed to be the semantic topic of the utterance as it figured into the context of the discourse. Third, I tried to determine whether the mother seemed to be directing the infant's attention to the object, or requesting that s/he perform the action, signified by the target word. Using these criteria, I calculated the percentage of utterance pitch peaks in these transcript portions which corresponded to focally stressed words.⁷ Lastly, in light of research findings suggesting that focally stressed words in IDS are concentrated in utterance-final position, (e.g., Woodward and Aslin 1989, 1992), and that peak f_0 and utterance-final position are often combined in marking such words, (e.g., Fernald and Mazzie 1991), I calculated the

on the pitch peaks of these utterances would have been trivial and not instructive, but these utterances were not excluded from other analyses conducted on these transcript segments because I felt that doing so would have significantly disrupted the flow of the discourse.

⁶ In those cases where not all syllables of a multisyllabic word occurred on the utterance pitch peak, this fact was noted by rewriting only the pitch-peaked syllable(s) of the word in capital letters. The entire word, however, was treated as a potential candidate for focal stress.

⁷ While I realize that this analysis is only an incomplete measure of the prevalence of focal stress in spontaneous IDS since focally stressed words may have occurred elsewhere in these samples besides on f_0 peaks, the investigation of other prosodic correlates of focal stress and the subjection of these prosodic features to tests of cue reliability were beyond the scope of the current study. Such questions would make excellent topics for future research.

proportion of pitch-peaked words occurring in utterance-final position and the proportion of utterance-final, pitch-peaked words which were recipients of focal stress.

4.2.5 Relationship of input variables to infants' lexical production

Finally, for any variant of the prosodic bootstrapping hypothesis to gain empirical support, it is not sufficient to show that a particular prosodic cue reliably signals a semantic feature such as focal stress. It must also be shown that infants actually use this cue to guide or facilitate their language acquisition process in some way. To assess the evidence for such facilitation in these samples, I compiled a list of the distinct lexical items produced by each 15-month-old during the same segment of the speech sample on which the focal stress analysis on pitch peaks in maternal utterances described in the previous section was conducted. Restricting my inspection to the same segment of the speech sample, I then determined how many of the child's words fit into each of the following four categories with respect to prior input: words not previously spoken by the mother at all, words previously spoken by the mother in a one-word utterance or in utterance-final position, words previously spoken by the mother on an f_0 peak, and words previously spoken by the mother on a pitch peak that was also judged to be a locus of emphatic stress. The first of these measures addressed the general question of the extent to which infants' early language production is contingent upon recent linguistic input, while the second and third targeted hypotheses regarding infants sensitivity to certain intonational and temporal characteristics of speech, regardless of their reliability as signals of semantic content. The fourth of these measures tested the stronger claim that infants preferentially attend to and repeat only those words which are both prosodically marked and semantically relevant to the discourse.

4.3 Results and discussion

4.3.1 Global measures of ID prosody and infants' productive language abilities

Instrumental analysis of a selection of utterances from the four speech samples revealed that M1 spoke to her infant at nine and a half months with a considerably (12 hz) higher mean f_0 than she used when the infant was 15 months old. Both of the mean

pitch levels for M1 were at least 25 hz higher than the mean f_0 of M2's speech, which in turn showed only a slight decrease of 4 hz between the nine-and-a-half- and 15-month samples. The mean f_0 values for all four speech samples are shown in table 1.

In terms of the amount of pitch modulation in the two mothers' speech, both auditory impressions of the four speech samples and inspection of the spectrograms for the 80 sampled utterances clearly indicated a great deal more overall variability in the pitch of M1's speech at both infant ages as compared to that of M2. The peaks in M1's utterances tended to occur at considerably higher f_0 levels than either the peaks in M2's speech or the surrounding syllables in M1's utterances, and M1 was also more likely than M2 to change the direction of her pitch contour more than once in a single utterance. As a result, the spectrograms for M1's utterances in both of the M-C1 samples had more spikes and dips than those of M2's utterances, and those spikes and dips were substantially larger in magnitude. Rather than reflecting truly distinct intonational patterns, however, some of these differences may simply stem from the fact that M2's ID utterances were significantly shorter on average than those of M1 and were far more likely to contain only one word, (23 of the 40 sampled as opposed to only two of the 40 sampled for M1). It may simply not be natural or motorically possible to produce the sorts of wide f_0 movements evident in M1's speech within the relatively short, predominantly one-word utterances characteristic of M2.

The analysis of the prevalence of utterance-final pitch peaks in these samples produced a somewhat more complex pattern of results. For both mothers, utterances in which the f_0 peak coincided with the cessation of voicing and which could thus be characterized as ending in an f_0 rise were quite rare, (2 in the M-C1 nine-and-a-half-month-old sample, 3 in the M-C1 15-month-old sample, 2 in the M-C2 nine-and-a-half-month-old sample, and 4 in the M-C2 15-month-old sample). However, when pitch peaks occurring anywhere in the final word of an utterance were counted, a majority of the utterances in each of the four samples were found to have their $f_{0\text{ peaks}}$ occurring during their final word(see table 1).⁸

⁸ The fact that over half of M2's utterances contained only one word and yet were counted as having their f_0 peaks occurring in their final word certainly represents a confound in this data and weakens the impact of the reported finding. I considered the alternative of discounting one-word utterances from this analysis as I did in the analysis of focal stress, (see section 4.2.4 above and section 4.3.2 below), but I decided that the

In many cases, disyllabic words occurring utterance-finally exhibited pitch peaks on their initial syllable followed by precipitous drops in f_0 during the final syllable, often so steep a decline that the second syllable ended up occupying the f_0 minimum of the utterance. A similar tendency, though usually smaller in magnitude, was apparent in some utterances ending in monosyllabic words, where the f_0 drop occurred during an elongated vowel or diphthong. These vowel elongations were especially prevalent in M2's speech, in which many of the one-word utterances, often the infant's name used as an attentional, were monosyllables pronounced with a canonical bell-shaped melodic contour where virtually all of the pitch modulation occurred during one elongated vowel.

The assignment of higher pitch to the initial syllables of disyllabic words might have been predicted based on the fact that the majority of English words have a trochaic stress pattern and elevated pitch is a common prosodic correlate of stress, as has been extensively discussed in this paper. However, the size of the pitch differentials between the syllables in these utterance-final IDS words far exceeds that which most other speech registers would use simply to signal trochaic stress. One possible interpretation of this effect is that the magnitude of the pitch decrease in the second syllable of these words, and in the codas of those monosyllabic words which underwent a similar process, may have provided a perceptual contrast which served to highlight the pitch peak at the word's onset, thus aiding the infant in parsing out and identifying that word as a discrete lexical unit. Further research would be needed to fully elaborate and empirically test this theory, but for the sake of the present analysis, the number of utterances in each sample with their pitch peaks located on final words was taken to be a better indicator of characteristic ID intonational patterns than the number of utterances with terminal rising f_0 , and so this measure was included in table 1.

Before turning to indices of the two infants' productive language, a few brief comments should be made about some aspects of the mothers' IDS that were not covered in the preceding analyses. For one thing, it is interesting to note that while M1's speech showed considerably more f_0 modulation than M2's, the sampled M2 utterances

intonational patterns used in these one-word utterances were interesting subjects of investigation and that the very prevalence of these utterances in M2's speech constituted an important fact about her IDS that should not have simply been eliminated from the discussion. The data in table 1 about utterance-final pitch peaks in M2's speech should, however, be viewed with some caution in light of this confound.

contained a greater variety of pitch contours. Thus while the rising-falling contour associated with ID approvals and attention maintenance utterances (see section 3.1.2 above) typified the majority of both mothers' utterances, M2, but not M1, also produced a substantial number of utterances with steeply falling pitch contours. These contours occurred primarily in prohibitions and imperatives but were also present in several more conversational utterances that seemed to be exhibiting the process of declination common in ADS. In addition, especially in the nine-and-a-half-month sample, several of M2's utterances had f_0 traces which were relatively flat. These contours tended to characterize words like "mama" and "dada" spoken in isolation in an attempt to elicit repetition from the infant. In general, M2 was much more likely than M1 to overtly model the words she wanted her infant to produce by speaking them in isolation or in short carrier phrases like, "Say mama." In contrast, M1 tended to place object labels and other seemingly targeted words at the ends of more complex utterances, such as, "Morgie, look at that TRIangle!", and to mark them with the sort of steeply rising-falling f_0 contours described in the preceding paragraphs. Unlike M2, M1 also repeated many of her infant's vocalizations at nine-and-a-half months and discernible words at 15 months, saying things like "GOOgoo!" in the earlier sample and "flower! That's right!" in the later one in response to utterances produced by her infant. It is unclear whether this infant imitation on the part of M1 or the deliberate word teaching undertaken by M2 were representative of these mothers' everyday interactions with their children, or whether one or both of these "strategies" arose because the mothers knew that they were being recorded and consequently became hypersensitive to and conscious of their infants' linguistic behavior.

Table 1: Global Correlates of ID Prosody and Infant Productive Language Capacity

Dyad	Infant age	Mean maternal f_0	Proportion of utterances with f_0 peak in final word ⁹	# of infant utterances containing intelligible words	# of distinct lexical items produced by infant ¹⁰
M-C1	9 months 17 days	305 hz	70%	0	0
M-C1	15 months 4 days	293 hz	60%	41	17
M-C2	9 months 25 days	266 hz	60%	2	1
M-C2	15 months 13 days	262 hz	65%	83	23

In terms of her mean f_0 level and amount of pitch modulation, both features which have been repeatedly cited as core elements of ID prosody, M1 seems to present a more prototypical picture of the register than does M2. When we look at the amount of productive language evidenced by the two infants, however, (see table 1 above), we find that C2 produced over twice as many utterances containing identifiable linguistic content in the 15-month sample and also articulated six more distinct lexical items than C1 in that sample. IN addition, while C1 produced no intelligible utterances during the nine-and-a-half-month sample, C2 produced two--both composed of the word "dada." Thus these data provide no evidence for a facilitative effect of global correlates of ID prosody on the rate of lexical acquisition, and if anything they seem to suggest a relationship running in the opposite direction. Of course, the tiny size of this sample both in terms of the number of subjects and the types of prosodic variables studied makes it impossible to draw any

⁹ These percentages were determined by calculating the number of utterances in each sample whose f_0 peaks occurred on their final word, dividing each of these totals by the 20 utterances analyzed for each sample, and multiplying the results by 100.

¹⁰ A complete listing of the lexical items produced by each infant and the frequency of each one in their speech can be found in Appendix A.

conclusive findings from this analysis. Additionally, it must be noted that the ages of these two infants were not exactly matched and that C2 was eight days older when the first of these samples was made and nine days older in the second recording, perhaps accounting for at least some of the disparity in their levels of productive language.

Further research should certainly be undertaken to more systematically analyze the link between prosodic aspects of the input and lexical acquisition which this study began to explore, but any such research would need to take several potential contaminants and interacting variables into account. Firstly, future studies should involve a much larger corpus of speech samples from each mother than what was analyzed here and should begin analyzing maternal speech at earlier infant ages, so that the changes in ID prosody throughout each child's development could be more reliably tracked and so that the influences of the infants' own emerging linguistic capabilities on the mothers' speech would be minimized. Secondly, subsequent investigations should focus on one particular element of maternal prosody at a time, since in the present study a clear inverse relationship emerged between the two mothers levels and movements of f_0 and the proportion of short, one-word utterances in their speech. It may thus be the case that different mothers exploit different intonational and durational attributes to distinguish their ID and AD registers, and the presence of such individual differences provides a strong argument for studying prosodic variables independently in IDS. Thirdly, it is essential that future database analyses obtain samples of each mother's ADS for comparison purposes, as was not done in the current study, because the observed difference in mean f_0 between M1 and M2, for example, may simply have represented a global difference in the pitch of their speaking voices that would have shown up in their ADS as well. Rather than comparing the mean f_0 of speech addressed to infants by different mothers, then, future studies of this type should instead measure the difference in each mother's mean f_0 between her IDS and ADS samples and then compare these difference scores across all mothers in the study.

In addition, future researchers may want to subdivide their samples according to social class, ethnicity, and/or sex of child to determine both whether these demographic variables are related to differences in the magnitude or specific correlates of ID prosody and whether they interact in any way with the link between ID prosody and infant

language production. Finally, researchers conducting analyses similar to this one in the future might want to attain some measure, either in the form of parental reports or standardized tests like the Bailey Scale, of the temperaments, activity levels and cognitive abilities of the infants in their samples. This information could be used to look for any correlations between infants' temperamental and cognitive characteristics and the prosodic features of their mothers' speech as well as for the more obvious purpose of controlling for the effect of these individual differences on infants' linguistic productivity. These qualifications suggest that a definitive and universal conclusion about the global effect of ID prosody on lexical acquisition is unlikely to emerge even from more extensive studies than the one described above, but there are a number of more nuanced and complex questions about this relationship which can and should be asked and which will hopefully provide the basis for meaningful research in the child language field for some years to come.

4.3.2 Analysis of focal stress

Application of the three criteria for focal stress outlined in section 4.2.4 above to the words occupying the pitch peaks of 40 multiword maternal utterances in the two 15-month samples revealed that in M1's sampled utterances 19 of the 44 words located on f_0 peaks, or 43.2%, were judged to be focally stressed. Among the 40 M2 utterances studied, 16 of the 42 words occurring on pitch peaks, or 38.1%, were also recipients of emphatic stress.¹¹ Thus while focal stress was somewhat more reliably signaled by pitch prominence in M1's speech, both mothers used this prosodic variable as a cue to emphatic stress in less than half of their multiword utterances. This implies that simply focusing on utterance f_0 peaks would not be a particularly effective strategy for infants to use in their efforts to identify the words whose referents are being highlighted in the discourse.

The case for peak f_0 as an informative stress marker is somewhat strengthened, however, when the analysis is restricted to those pitch-peaked words which occurred utterance-finally. As can be seen in table 2, at least half of the words spoken on f_0 peaks

¹¹ The language transcripts corresponding to the portions of the 15-month samples from which these maternal utterances were taken, along with my notations of f_0 peaks and focal stress, are included in appendices 2 and 3.

by both mothers in these samples occurred in utterance-final position, and 50% or more of these relatively high-pitched final words were also judged to have been focally stressed. The large proportion of those words labeled as focally stressed which were produced utterance-finally, especially by M2, adds further support for the notion that pitch prominence and utterance position together are more predictive of emphatic stress than pitch prominence alone,¹² although this latter measure does not directly assess the cue reliability of this composite variable as it would be experienced by language-learning infants, (see section 4.2.4 above).

Table 2: Relationships Between F₀ Prominence, Utterance Position and Focal Stress in 15-Month Samples

- 1 = number of utterance-final and pitch-peaked words which were focally stressed
- 2 = percentage of utterance-final and pitch-peaked words which were focally stressed
- 3 = percentage of focally stressed words which were pitch-peaked an utterance-final

Speaker	Number of words on f ₀ peaks	# focally stressed	% focally stressed	# utterance-final	% utterance-final	1	2	3
M1	44	19	43.2%	22	50%	11	50%	57.9%
M2	42	16	38.1%	22	52.4%	12	54.5%	75%

F₀ prominence may provide more salient cues to focal stress than the preceding analysis would lead one to believe, however, because while every utterance necessarily contains a pitch peak unless it is spoken in a complete monotone, not all pitch peaks are equal in

¹² Since focal stress as a function of utterance position alone was not calculated in this study, there is no way of knowing whether the interaction of focal stress with the composite of pitch and utterance position discussed here is significantly larger than a simple focal stress-utterance position interaction would have been. Hopefully future research in the field will take up this question.

magnitude in terms of the f_0 distance between the peak and the syllables preceding and following it. Furthermore, impressionistic observations of the 80 maternal utterances analyzed above suggest that focally stressed words were far more likely to occur on relatively large pitch peaks than on those peaks which were smaller in magnitude. It may well be the case that infants perceive relatively minor pitch peaks merely as part of the overall melodic contour of the speech stream but respond with increased attention to peaks whose deviation from the surrounding f_0 baseline exceeds some threshold, treating them as discrete events or departures from the normal. On this view, it would only be necessary for f_0 peaks whose magnitude is sufficient to surpass this perceptual threshold to correlate significantly with focal stress in order for pitch prominence to constitute a reliable and useful semantic cue for infants. Future researchers might thus wish to construct a continuum of f_0 peak magnitude by comparing each peak to the f_0 of its surrounding syllables and then assess the prevalence of focal stress in words occupying various points along this continuum. Such a continuum-based measure would yield a more accurate picture than the binary analysis presented here of the prosodic variable of pitch prominence as it is actually experienced by the infant, and this more comprehensive view may in turn shed more light on the semantic and lexical information which different degrees of this variable could provide to language learners.

4.3.3 effects of input on words produced by infants

We end by considering the results of a more microcosmic analysis of the relationship between prosodic characteristics of the input and productive infant language than that described in section 4.3.1 above. Within the portion of the 15-month sample comprising the 40 multiword maternal utterances used in the focal stress analysis, (see the previous section), C1 produced five distinct lexical items while C2 produced ten. As table 3 shows, three of the words spoken by C1 had not been introduced into the discourse by her mother earlier in that sample segment.¹³ When M1 did produce a word

¹³ I restricted my search for earlier maternal uses of infant words to the transcript portions delimited by the 40 maternal utterances already analyzed for two reasons. First, doing this gave me easy access to information on whether particular maternal tokens were focally stressed. Second, I reasoned that maternal tokens occurring much earlier in the sample were apt to relate to a different pragmatic context than that in which the child used the word. In other words, I wished to concentrate on the prosodic elements of maternal utterances that seemed to have at least a potential influence on their children's later speech, and

that C1 later repeated, however, it always occurred in isolation or on an utterance-final pitch peak and was judged to have received focal stress. In contrast, all but one of the ten words produced by C2 during this portion of his 15-month sample were first used by his mother, many occurring multiple times in isolation or in utterance-final position. While pitch prominence was less commonly associated with these maternal tokens of infant words than position as the last or only word in an utterance, those tokens which did occupy f_0 peaks were also overwhelmingly likely to have been labeled as focally stressed.

Table 3: Proportion of Words Spoken by Infants at 15 Months Exhibiting Various Prosodic Features in Previous Maternal Tokens

- Uf/Ow = previously spoken in utterance-final position or as a one-word utterance
- Pp = previously spoken on an utterance f_0 peak
- Stressed = previously spoken on an utterance pitch peak which was judged to be a locus of focal stress

Dyad	# of infant words	% previously spoken by mother	% Uf/Ow	% pp	% stressed
M-C1	5	40%	40%	40%	40%
M-C2	10	90%	80%	50%	40%

In addition to the percentages displayed in table 3 above for each infant's total productive lexicon during this portion of the 15-month sample, the relatively few words spoken by infants in these segments, while limiting the analysis in some ways, also made it feasible to present a more detailed overview of the previous tokens of each infant word spoken by his/her mother, (see tables 4 and 5). These tables make it possible to assess both the frequencies and some of the temporal and intonational properties of various

significant distance between the maternal and child tokens of a word made such an influence appear unlikely. It should be acknowledged, however, that the 40-utterance segment does place somewhat

infant words as they appeared in prior maternal speech, as well as providing some raw data to test hypotheses about the prosodic features of the input and the influence of that input on word production for words belonging to different semantic and grammatical categories.

Table 4: Previous Maternal Tokens of Words Produced by C1 at 15 Months

- Uf/Ow = words previously spoken in utterance-final position or as one-word utterance
 - Pp = words previously spoken on an utterance f_0 peak
- Stressed = words previously spoken on an utterance pitch peak which was judged to be a locus of focal stress

Lexical item	# of times produced by child	# of times previously used by mother ¹⁴	# Uf/Ow	# pp	# stressed
C.J.15	2	0	0	0	0
Thank you	1	0	0	0	0
Rock	1	4	3	3	3
Flower	1	0	0	0	0
Stick	1	1	1	1	1

arbitrary limits on the investigation of such possible influences.

¹⁴ For the purposes of this analysis, only maternal tokens of a word occurring prior to the infant's first use of it were counted, even in those cases where the infant went on to use the word more than once. For virtually all infant words analyzed in these sample portions, mothers repeated each word at least once directly following the infant's initial use of it, and I felt that including these later maternal tokens without being able to specify whether they occurred before or after the infant's first use of the word would have substantially distorted the picture presented by the data.

¹⁵ This is the name of someone the child knows, and the reason she produced it in this sample is that she saw a truck that she thought belonged to him and pointed it out to her mother.

Table 5: Previous Maternal Tokens of Words Produced by C2 at 15 Months

Lexical item	# of times produced by child	# of times previously used by mother	# Uf/Ow	# pp	# stressed
Ball	1	1	1	1	1
Cat	4	1	1	0	0
A16	1	7	7	1	1
Read	1	4	2	1	1
Book	3	10	9	3	0
G	1	1	1	0	0
I	1	1	1	0	0
Nana	1	0	0	0	0
Kiki ¹⁷	1	3	3	1	1
Bite	2	1	0	0	0

While the patterns of results for the two dyads are somewhat different, some generalizations about the prior maternal tokens of infant words can be made. For one thing, those infant words which had previously been spoken by the mother were overwhelmingly likely to have been produced utterance-finally or as a one-word utterance. Secondly, words spoken by the mothers on utterance f_0 peaks tended not to appear in their infants' later speech unless they were also recipients of focal stress. This adds credence to the hypothesis advanced in the previous section that not all pitch peaks are equally salient for infants and that they possess some method, either prosodic, (e.g., magnitude of pitch peaks), or pragmatic, (e.g., direction of maternal gaze or other contextual factors), for determining which f_0 peaks have semantic significance and which do not.

¹⁶ All letters listed in this column refer to letters of the alphabet whose names were produced by both mother and child as they read an ABC book together.

¹⁷ I assume that this is the name of someone the child knows, although this is not made clear anywhere in the sample.

Several cautionary notes must be struck, however, before the findings reported here are taken to imply that words placed in utterance-final position or marked with emphatic stress by mothers will be acquired more quickly or easily by their infants. It must be kept in mind, for example, that while the infants in this study did not have very large lexicons at the time the last speech samples were made, they were old enough to be producing words consistently, meaning that the tokens collected in this analysis probably did not represent the infants' first attempts to produce the words in question. Thus what was actually measured in these samples was whether various prosodic properties in the input increased the probability that infants would produce words that they already knew, and this study has nothing to say about any role that such properties might play in helping infants to segment and produce words for the first time. Such a facilitative effect of input prosody would have to be investigated by collecting data on maternal speech beginning at a very young infant age, tracking the frequencies and prosodic characteristics of numerous words in that speech, and then determining whether words spoken more often by the mother, or more highly represented in certain privileged prosodic positions such as pitch peaks and the ends of utterances, appear any earlier in the infants' own speech than less common words.

Another potential problem with this study is that in attempting to determine why infants produce certain words and not others, there is no way of teasing apart the influence of the linguistic input and particularly its prosodic character from the contributions made by pragmatic factors such as the direction of maternal and infant gaze, the objects being handled by one or both interlocutors, and other actions taking place in the environment. The infants whose speech patterns were analyzed in this study may therefore not have been responding directly to the prosodic cues used by their mothers to mark emphatic stress but instead may have chosen to produce the words they did based on the same pragmatic and discourse variables which prompted the mothers to focally stress those words in the first place. It would appear that no naturalistic data can isolate and control for these sorts of extralinguistic factors in order to pinpoint the role played by maternal prosody alone, but an experiment could be designed to address this question by asking mothers or other adults to figure out what prosodic cues they would typically use to signal focal stress and then apply those cues to words which are not

central to the discourse, thus providing their infants with a prosodic-pragmatic mismatch. The results of such an experiment, while interesting, would be only marginally relevant to the study of normal language acquisition, however, for even the most forceful proponents of the prosodic bootstrapping hypothesis acknowledge that prosodic information is integrated with contextual cues as well as previous linguistic knowledge in the decisions infants ultimately make about how to parse the speech stream and about which acoustic units to map onto which real-world referents. All the prosodic bootstrapping hypothesis asserts is that input prosody is one source of information that infants use to organize and interpret the speech they hear and that this information enables the language acquisition process to proceed more smoothly than it otherwise would, and the findings described above, while not providing conclusive proof by any means, are certainly consistent with this theory.

5. Conclusion

The intonational and temporal properties that characterize the speech register we use to address infants are markedly different from those used in ordinary speech, occur in speakers of both genders using a variety of languages, and have been shown to generate both positive affect and increased attention in infants as young as 52 hours of age. There is also evidence that infants respond in emotionally appropriate ways to the different canonical melodic contours of ID attentionals, approvals, prohibitions and expressions of comfort. The relationship between these prosodic variables and indices of lexical acquisition is less straightforward, with individual differences in infants' cognitive abilities and in the prosodic aspects emphasized by different maternal IDS registers making it difficult if not impossible for any universal conclusion about such a link to be reached.

The foregoing analysis does suggest, however, that 15-month-old infants are far more likely to produce words previously spoken by their mothers in isolation or utterance-final position than words occurring elsewhere in their mothers' speech. Infants at this age also demonstrated a tendency to reproduce words occurring on f_0 peaks in their mothers' speech only when those f_0 peaks were also identified as loci of focal stress, and such loci of stress were in turn found to be more reliably signaled by pitch prominence

and utterance-final position in combination than by pitch prominence alone. These findings lend support to the claim that certain prosodic aspects of the input, perhaps acting in tandem, have an effect on young children's linguistic behavior, thus reaffirming the accuracy of an observation made by Dwight Bolinger in 1965, who defined prosody as "those components of speech that come first for the child but last for the analyst." He continued:

"If the child could paint the picture, these [components] would be the wave on which the other components [of speech] ride up and down. But the linguist is older and stronger and has his way. He calls them suprasegmentals and makes the wave ride on top of the ship."(7, cited in Fernald 1991).

In the years since Bolinger made this statement, prosody has gained considerably more attention in the linguistic and psychological literature, particularly as it is realized in IDS. A great deal more research needs to be done and a number of complex interrelationships must be untangled, however, before we will fully understand the precise role that this wave of prosody plays in charting the course of language acquisition.

Appendix A
Productive Lexicons of C1 and C2 as Evidenced in Their Full 15-Month Samples

In these tables, the words are listed in the order in which the child first produced them.

Words Produced by C1 in Hour-Long Sample Taken at 15 Months 4 Days of Age

Lexical item	Frequency in sample
Flower	3
Doggie	1
C.J.	5
Go	5
Push	3
Beep beep	2
Thank you	2
Rock	1
Stick	1
Car	1
Mom/mama	3
Hi Zach	2
Nana	3
No	3
Dirty	1
Fish	1
Baby	1

Words Produced by C2 in Hour-Long Sample Taken at 15 Months 13 Days of Age

Lexical item	Frequency in sample
Two	3
Ball	1
Cat	5
A	1
Read	2
Book	3
G	1
I	1
Nana	4
Kiki	1
Bite	2
K	1
Dog	3
Q	4
Meow	5
Baby	4
B	1
Eat	1
Baba	10
Cup	2
Go	4
Mama/mommy	12
Bye/byebye	13

Appendix B
Portion of M-C1 15-Month Transcript Used in Focal Stress Analysis

[Syllables or words appearing in all capital letters occurred on their carrier utterance's f_0 peak. Words appearing in all capitals and in bold type occurred on the f_0 peak of an utterance and were also judged to have received focal stress. One-word utterances by the mother were not included in the analysis of f_0 peaks and focal stress.]

@Participants: CHI Morgan Child, MOT Brenda Mother, OTH unknown other
@Birth of CHI: 28-Mar-1996
@Date: 02-Jul-1997

1*MOT: Morgie , did you have blueBERRY-S ?
*CHI: C+j .
*CHI: C+j .
2*MOT: THAT'S not C+j , sweet-ie .
3*MOT: THAT'S NOT C+j .
4*MOT: we're not to C+j-'s HOUSE yet .
5*MOT: did you have blueBERRY-S for BREAKFAST ?
6*MOT: Morgie , did you have PANCAKE-S ?
*MOT: pancake-s .
7*MOT: well you're just interested in this TRUCK .
8*MOT: PRETTY truck isn't it ?
9*MOT: it's a BLUE truck but it's not C+j-'s truck .
10*MOT: C+j has a RED truck .
11*MOT: (o)kay it's a BLACK truck .
12*MOT: C+j-'s truck is RED .
13*MOT: where you GO-ing ?
14*MOT: are you JOG-ING ?
15*MOT: that's ORANGE .
*MOT: rock .
16*MOT: THAT'S right .
17*MOT: that's a ROCK .
18*MOT: come ON !
19*MOT: let's go see BRISCO !
*MOT: you're xxx .
20*MOT: go-ing the wrong WAY, sweet-ie .
21*MOT: Brisco-'s THIS way .
22*MOT: let's go see BRISCO .
*MOT: Sammy .
23*MOT: come ON !
*CHI: thank you [?] .
24*MOT: what are you THANK-ing me for ?
25*MOT: come ON !
26*MOT: is that a **ROCK** ?
27*MOT: (i)s that a **ROCK** that you HAVE ?

*CHI: rock .
 *MOT: rock .
 28*MOT: that's **RIGHT** .
 29*MOT: that Morgan-'s **ROCK** ?
 *CHI: flower .
 *MOT: flower .
 30*MOT: that's **RIGHT** .
 31*MOT: those are **NOT OUR** flower-s .
 32*MOT: we don't **PLAY** with other people-'s flower-s .
 33*MOT: they're **PRETTY** though aren't they ?
 34*MOT: they're pink and purple and **YELLOW** .
 35*MOT: **VERY** pretty .
 36*MOT: **LOOK** , morgan !
 37*MOT: Mommy has a **STICK** .
 *CHI: (s)tick .
 38*MOT: **THAT'S** right .
 *MOT: stick .
 39*MOT: **COME** on !
 40*MOT: let's go see **BRISCO** !

Key to Symbols Used in This Transcript

- - separates morphemes within a single word; e.g., walk-ed
- Consonants or vowels surrounded by parentheses were absent or reduced in the output; e.g., goin(g) where /n/ and not /N/ (read “ng”) is the final consonant
- xxx = portion of an utterance which was unintelligible
- + appearing between two words means that they were spoken as a compound; e.g., kitty+cat

Appendix C
Portion of M-C2 15-Month Transcript Used in Focal Stress Analysis

[Syllables or words appearing in all capital letters occurred on their carrier utterance's f_0 peak. Words appearing in all capitals and in bold type occurred on the f_0 peak of an utterance and were also judged to have received focal stress. This segment is representative of the sample as a whole in its large number of one-word maternal utterances, utterances which were partially or entirely unintelligible, (denoted by "xxx" in the transcript), and utterances which occurred in environments of significant background noise, (denoted by "noisy" in brackets following the affected utterance). None of these types of utterances were included among the 40 for which pitch peaks and focal stress were analyzed.]

@Participants: CHI Vas Child, MOT Dya Mother, OTH unknown other

@Birth of CHI: 25-May-1999

@Date: 09-Sep-2000

1*MOT: you wanna do your ABC ?
*MOT: huh ?
*MOT: book .
2*MOT: yeah go GET that book !
3*MOT: WITH the abcs in it .
4*MOT: look LOOK !
5*MOT: **THIS** book , man+man .
6*MOT: **THAT** book .
7*MOT: GET **THAT** book !
8*MOT: see that BOOK ?
9*MOT: with the abC on it .
10*MOT: go get that BOOK !
*MOT: wanna read this one [=! noisy] ?
*MOT: a@l .
*MOT: a@l .
*MOT: a@l .
*MOT: a@l .
11*MOT: you have to use the **POT** ?
*MOT: huh ?
12*MOT: (a)round eleven **THIRTY** you need to get on the pot .
13*MOT: you don't wanna do the abc **BOOK** ?
*MOT: yeah let's do the other one [=! noisy] !
*MOT: I (d)on't like that one [=! noisy] .
*MOT: that's kinda advanced for you [=! noisy] .
*MOT: here .
14*MOT: I can **READ** it to you .
*MOT: read .
*MOT: read [=! noisy] .
*MOT: yeah [=! noisy] .

*MOT: be a long day , baby [=! noisy] .
 *MOT: it's make-ing me xxx .
 *MOT: it's nothing xxx xxx today [=! noisy] .
 15*MOT: COME on !
 16*MOT: come OVER here !
 *MOT: look !
 17*MOT: you like the A@l ?
 *MOT: a@l .
 *MOT: baby [=! noisy] .
 *MOT: book [=! noisy] .
 *MOT: www .
 %add: not CHI
 *MOT: www .
 %add: not CHI
 *MOT: xxx xxx .
 *MOT: why you got this xxx xxx [=! noisy] ?
 *MOT: look !
 *MOT: okay [=! noisy] .
 18*MOT: you like this **BALL** ?
 *CHI: ball .
 19*MOT: see what's on **HERE** !
 *MOT: cat .
 *CHI: cat .
 *MOT: c@l .
 *MOT: cat .
 *MOT: c@l [=! noisy] .
 *CHI: cat .
 *MOT: cat .
 *MOT: hat .
 *MOT: e@l [=! noisy] .
 *MOT: e@l [=! noisy] .
 *MOT: e@l [=! noisy] .
 *MOT: elephant [=! noisy] .
 *MOT: a@l .
 *CHI: a@l .
 *MOT: xxx .
 20*MOT: look at **MOMMY** !
 *MOT: book .
 *MOT: xxx .
 *CHI: read .
 *MOT: read .
 *MOT: you like to read [=! noisy] ?
 *CHI: book [?] .
 *MOT: read !
 21*MOT: **READ** the book !
 *CHI: book .

*MOT: g@l .
 *CHI: g@l .
 22*MOT: GO fish !
 23*MOT: WHAT was that ?
 24*MOT: a **GEESE** ?
 25*MOT: YOU LIKE duck-s ?
 26*MOT: you wanna read another **BOOK** ?
 27*MOT: go **GET** it !
 28*MOT: what **BOOK** you like to read ?
 *MOT: i@l [=! noisy] .
 *CHI: i@l .
 29*MOT: **ICE** cream .
 30*MOT: say **ICE** ["] !
 *MOT: k@l .
 *CHI: boo(k) .
 *MOT: wha(t) ?
 *MOT: book ?
 *MOT: look [=! noisy] !
 *MOT: k@l .
 31*MOT: here's a **K**@l .
 *MOT: k@l .
 32*MOT: like **KIKI** ?
 *MOT: Kiki .
 *MOT: Ken .
 *CHI: Nana .
 *MOT: Nana .
 *MOT: Kiki .
 *CHI: Kiki .
 33*MOT: say **L**@l ["] !
 34*MOT: bite **YOU** .
 *MOT: lion .
 *CHI: bite .
 *MOT: bite .
 *CHI: bite .
 *MOT: bite .
 *MOT: n@l .
 *MOT: Nana .
 *MOT: Nana .
 *MOT: n@l .
 *MOT: xxx xxx xxx , man+man [=! noisy] .
 *MOT: say xxx ["] !
 35*MOT: you wanna see the **O**@l ?
 *MOT: o@l [=! noisy] .
 36*MOT: here's **O**@l .
 37*MOT: yeah **YOU** know .
 *CHI: cat [=! noisy] .

*MOT: where you see a cat [=! noisy] ?
*CHI: cat [?] .
38*MOT: you wanna see the **CAT** ?
39*MOT: let's SEE !
40*MOT: that's a **KITTEN** .

Key to Symbols Used in This Transcript

- - separates morphemes within a single word; e.g., toy-s
- Consonants or vowels surrounded by parentheses were absent or reduced in the output; e.g., wan(t) = “want” with /t/ not articulated
- xxx = portion of an utterance which was unintelligible
- @l following a letter indicates that the letter’s name was spoken rather than its corresponding sound(s) pronounced; e.g. a@L = the letter a
- + appearing between two words means that they were spoken as a compound; e.g., kitty+cat
- www = language that was not transcribed because it was a lengthy aside not addressed to the child; in this case, the mother answered a phone call at that point in the recording session

References

- Aslin, Richard N., Woodward, Uli Z., Lamendola, Nicholas P., and Bever, Thomas G. "Models of word segmentation in fluent maternal speech to infants." in Signal to Syntax: Bootstrapping from speech to Grammar in Early Acquisition. James L. Morgan and Katherine Demuth, eds. Mahwah, NJ: Lawrence Erlbaum Associates, 1996, 117-134.
- Bernstein-Ratner, Nan and Pye, Clifton. "Higher pitch in BT is not universal: acoustic evidence from Quiche Mayan." Journal of Child Language 11(3), October 1984, 515-522.
- Cooper, Robin and Aslin, Richard. "The language environment of the young infant: implications for early perceptual development." Canadian Journal of Psychology, 43(2), June 1989, 247-265.
- Cooper, Robin and Aslin, Richard. "Preference for infant-directed speech in the first month after birth." Child Development, 61(5), October 1990, 1584-1595.
- Echols, Catherine and Newport, Elissa. "The role of stress and position in determining first words." Language Acquisition, 2(3), 1992, 189-220.
- Fernald, Anne and Kuhl, Patricia K. "Acoustic determinants of infant preference for Motherese speech." Infant Behavior and Development, 20(3), July 1987, 279-293.
- Fernald, Anne. "Intonation and communicative intent in mothers' speech to infants: is the melody the message?" Child Development, 60(6), December 1989, 1497-1510.
- Fernald, Anne. "Prosody in speech to children: prelinguistic and linguistic functions." in Annals of Child Development. Ross Vasta, ed., Vol. 8. Bristol, PA: Jessica Kingsley Publishers, 1991, 43-80.
- Fernald, Anne and Mazzie, Claudia. "Prosody and focus in speech to infants and adults."

Developmental Psychology, 27(2), March 1991, 209-221.

Fernald, Anne. "Approval and disapproval: infant responsiveness to vocal affect in familiar and unfamiliar languages." Child Development, 64(3), June 1993, 657-674.

Fernald, Anne and McRoberts, Gerald. "Prosodic bootstrapping: a critical analysis of the argument and the evidence." in Signal to Syntax: Bootstrapping from speech to Grammar in Early Acquisition. James L. Morgan and Katherine Demuth, eds. Mahwah, NJ: Lawrence Erlbaum Associates, 1996, 365-384.

Fisher, Cynthia and Tokura, Hisia. "Prosody in speech to infants: direct and indirect acoustic cues to syntactic structure." in Signal to Syntax: Bootstrapping from speech to Grammar in Early Acquisition. James L. Morgan and Katherine Demuth, eds. Mahwah, NJ: Lawrence Erlbaum Associates, 1996, 343-364.

Karzon, Roanne. "Discrimination of polysyllabic sequences by one- to four-month-old infants." Journal of Experimental Child Psychology, 39(2), April 1985, 326-342.

McLeod, Peter. "What studies of communication with infants ask us about psychology: baby-talk and other speech registers." Canadian Psychology, 34(6), July 1993, 282-292.

Werker, Janet F., Lloyd, Valerie L. and Pegg, Judith E. "Putting the baby in the bootstraps: toward a more complete understanding of the role of the input in infant speech processing." in Signal to Syntax: Bootstrapping from speech to Grammar in Early Acquisition. James L. Morgan and Katherine Demuth, eds. Mahwah, NJ: Lawrence Erlbaum Associates, 1996, 427-444.