UCLA Working Papers in Phonetics, No. 108, pp. 203-225

Information Structural Expectations in the Perception of Prosodic Prominence*

Jason Bishop
j.bishop@ucla.edu

Abstract

A number of previous investigations using context matching (e.g., Gussenhoven 1983) and appropriateness rating tasks (Birch and Clifton 1995, Welby 2003) suggest that English-speaking listeners lack expectations regarding how the size of a focused constituent (broad versus narrow) can be expressed prosodically in certain constructions. In the present study English-speaking listeners were presented with the same SVO sentence (e.g., *I bought a motorcycle*) presented in either broad or narrow question contexts, and were asked to rate the prominence of the words in those sentences. In general, listeners reported sentence-final objects to be relatively more prominent than preceding verbs in the test sentences when those sentences were presented in narrow-object (*What did you buy?*) rather than broad-VP (*What did you do?*) or sentence (*What happened?*) focus contexts. This effect was found to be stronger in Experiment 2, where the answer was a correction. The findings suggest listeners do have expectations about the relationship between the size of a sentence's focus constituent and its prosodic realization. It is argued that these expectations are well founded given the listeners' likely experience with productions of this particular information structural contrast.

0. Introduction

A recurrent finding in speech perception research is that what a listener 'hears' in the acoustic signal is partially influenced by her knowledge of the utterance's linguistic structure. A clear and well-known demonstration of this is the 'phoneme restoration effect' (Warren 1970, Samuel 1981), where listeners are shown to use their knowledge of the phonological structure of a word to 'hear' phonetic information that has been masked or entirely removed from the signal. A similar restorative-like finding is 'perceptual epenthesis' (Dupoux et al. 1999), which suggests listeners hear–or do not hear–the presence of segmental information depending on their knowledge of what constitutes a possible phonotactic sequence in their language. What such phenomena indicate is that the way in which listeners interpret the acoustic signal is informative as to their representations of its linguistic structure, because they exhibit expectations for that structure.

In the present study we probed listeners for expectations about the way certain prosodic information relates to certain semantic-pragmatic aspects of a sentence's meaning. In particular, we examined whether listeners exhibit any expectations regarding how prosodic prominence can express an information structural distinction: the size of the focus constituent. The size of the focus constituent (ranging from broad sentence to narrow object focus) was chosen as the linguistic contrast to examine because there is considerable uncertainty in the literature on English focus

*A revised version of this paper is to appear in a volume edited by Gorka Elordieta and Pilar Prieto. Parts of the research reported here were presented at the 2009 Workshop on Prosody and Meaning in Barcelona, and at the 28th West Coast Conference on Formal Linguistics at the University of Southern California. I thank audiences at both conferences for helpful comments and suggestions.

prosody as to whether foci of different sizes in fact correspond to any unique prosodic pattern at all. The primary experimental finding presented here suggests that listeners do indeed have expectations about how this information structural contrast is realized. We found that listeners treat acoustically identical utterances in systematically different ways depending on their interpretation of the size of the sentence's focus constituent. In particular, it was found that listeners reported hearing sentence-final objects as more prominent when they were under narrow focus than when part of larger focus constituents. As is discussed below, this particular result is in line with what a number of recent production studies would predict.

The paper is organized as follows. In Section 1 we describe in detail the information structural contrasts of interest, some aspects of the acoustic signal that may correlate with them, and, finally, some previous studies that have probed listeners' knowledge of such correlations. In Section 2 we review some recent studies suggesting that listeners' perception of prosodic prominence, like their perception of segmental information, to some extent reflects their expectations. In Section 3 two novel experiments explore listeners' expectations about patterns of prosodic prominence in broad and narrow focus constructions in English. Section 4 includes a discussion of the experimental results and some concluding remarks follow in Section 5.

1. The size and type of a focus constituent

Focus is an aspect of a sentence's information structure, and is construed here as the information the speaker presents as *semantically or pragmatically* prominent. There are two dimensions along which focus may vary: *size* and *type*. The size of a focus constituent, sometimes called the focus "domain"[1], refers to the syntactic constituent under focus. Which constituent that will be depends on what information the speaker regards as informative, which itself is dependent on the discourse context in which the sentence is uttered. For example, what the size of the focus constituent in (2d) will be will depend on which of the contexts in (2a-c) it is uttered in:

(2) a. What happened?
b. What did you do?
c. What did you buy?

d. I bought a motorcycle.

In response to (2a), the focus constituent of (2d) is the entire sentence, referred to here as *sentence focus*; in the context of (2b), only the verb phrase in (2d) forms the focus constituent, referred to here as *VP focus*. Finally, if (2d) is uttered as a response to (2c), the focus constituent is the noun-phrase object, that is, *object focus*. Where the size of the focus constituent is larger, such as in sentence focus or VP focus, we will refer to them as cases of *broad* focus (Ladd 1996), compared to when the focus is comparatively *narrow* on a single word, such as in object focus.

[1] "Size of the focus" will be used in this paper rather than "domain of the focus" for primarily two reasons. First, "size of the focus" is a somewhat more transparent and theoretically neutral term, and, second, in order not to confuse it with another common use of "domain", which refers to the syntactic or semantic domain of a focus sensitive operator such as "only".

By focus *type*, we can distinguish two potentially different categories: one which is non-contrastive and one which is contrastive.[2] For the purposes here, *non-contrastive focus* will be used to refer to cases like that in (2), namely, the information which is required by a WH-question. In such cases, there is no explicitly mentioned alternative to the answer in the discourse. *Contrastive focus*, will be used for cases where the focus of a sentence does have an explicitly mentioned alternative (often referred to as "corrective focus"), as in *motorcycle* in (3b):

(3) a. Did you buy a car?

b. (No) I bought a motorcycle.

Note that when cases of contrastive focus, as defined here, are discussed, they are often also cases of narrow focus (e.g., Baumann et al. 2008, Hanssen et al. 2008). However, focus size and focus type are (in principle) orthogonal properties of information structure, and thus it is possible for a focus of any size to either have or not have an explicitly mentioned alternative in the discourse. In the present case, the sentence in (3b) could have a contrastive VP focus if uttered in response to a question such as "*Did you work today?*".

A matter of much recent interest is the relation these two dimensions of focus have to prosodic realization. Regarding the size of the focus in the simple subject-verb-object constructions considered in this paper, one account, the Focus Projection Hypothesis (Selkirk 1995, see also Gussenhoven 1984) predicts an ambiguity. According to this hypothesis, a single pitch accent on the object is said to be appropriate for the answer to each of the questions in (2), and so a distinctly narrow or distinctly broad focus meaning of the sentence is not expected to be expressible prosodically. However, it has been known for some time that speakers do not necessarily pronounce foci of different sizes equivalently. Gussenhoven (1983/1984), for example, found that listeners auditorily detected differences in sentences produced in a broad or narrow focus context, such that verbs were more prominent when part of a broad VP focus than when outside of the focus constituent. Further, Ladd (1996) discussed the possibility for English speakers to disambiguate narrow focus on an object from broad VP focus by placing more "emphasis" on the focused object -pronouncing it, for example, with a higher f0 peak.

Subsequent to these earlier observations, recent controlled production studies have suggested a rather systematic use of acoustic features to distinguish focus constituents of different sizes. Hanssen et al. (2008) reported that narrow focus was distinguishable from broad focus by Dutch speakers with respect to segmental durations (longer in an object under narrow focus) and the shape of the nuclear accent on the object (a steeper fall under narrow focus). For

[2] This distinction is made with the understanding that there is much debate as to whether contrast versus non-contrast is properly speaking a grammatical distinction. While some authors have assumed the grammar encodes the difference (Chafe 1976, Vallduví & Vilkuna 1998, among others), others have suggested a single category, all focus being essentially a kind of contrast (e.g., Jackendoff 1972, Rooth 1992). Yet another view comes from Büring (2007), who suggests that the distinction is not one of grammar, but of usage. Büring suggests recognizing the possibility that, rather than reflecting a property of the meaning of focus, interpreting contrast is matter of interpreting speaker intensions in a particular pragmatic context. Although the results of the experiments presented in Section 3 may be taken as relevant to the debate, a contribution to it is not a primary goal of the present paper. Thus no position is taken here as to the best characterization of the linguistic representation of contrast. (For a recent discussion of the matter, however, see Katz and Selkirk, submitted).

German, Baumann et al. (2006/2008) found that as focus narrowed from broad sentence focus to narrow object focus, the nuclear accented object was pronounced with longer segmental durations, with a higher f0 peak relative to prenuclear accents, and the probability of prenuclear accents on verbs decreased. Results similar to Baumann and colleagues', which considered the realization of objects in relation to surrounding phonetic material, have been reported for English by Sityaev & House (2003), Löfstedt (2006), Jun (in progress), and by Breen et al. (submitted).

Fewer studies have examined the prosodic realization of focus type, and among those that have, the results are somewhat less uniform compared to what was been reported for focus size. While two of the studies mentioned above, Hanssen et al. (2008) and Baumann et al. (2008), found no differences in speakers' productions of narrow contrastive and narrow non-contrastive focus, some studies of English speakers have. Bartels and Kingston (1994), for example, claim that contrastive focus is most reliably distinguished from non-contrastive focus by the height of the accent peak, f0 being higher for contrastive focus. Consistent with this, Ito et al. (2004) found speakers more likely to use a prominent pitch excursion (the ToBI L+H* rather than the H*) to mark foci as contrastive. Finally, Breen et al. (submitted) reported finding acoustic differences between productions of contrastive and non-contrastive focus when speakers were deliberately trying to communicate the distinction and were given feedback on their success. Although their findings do not support those of the previous two studies of English in terms of f0's role (indeed they find the opposite), Breen and colleagues did find that when speakers intentionally tried to express contrastive rather than non-contrastive focus, they did so using greater intensity relative to surrounding words.

Having discussed recent investigation into the phonetic realization of focus size and focus type, it is possible to make two generalizations regarding how speakers might encode these aspects of information structure prosodically. The first is that speakers, perhaps only when they are made conscious of an ambiguity, will tend to use increased acoustic prominence on a focused constituent when it is being used contrastively. The second, similarly, is that speakers will produce a sentence-final object with more acoustic prominence if it is narrowly focused than if it is part of a larger focus constituent. Note that what these generalizations would lead us to predict is that listeners should exhibit some corresponding expectations when faced with the task of interpreting the information structure intended by the speaker.

However, in the case of the size of the focus, some studies have been more in line with the ambiguous status predicted by the Focus Projection Hypothesis, mentioned above. For example, although Gussenhoven (1983/1984) found that productions of broad versus narrow focus were auditorily distinguishable to listeners in a prominence rating experiment, he found no evidence that any detectable differences were used in a task requiring listeners to match answer sentences with the correct question contexts. This basic result has been replicated in more recent studies using similar methodology. Birch and Clifton (1995) and Welby (2003), for example, found that listeners rated an answer sentence with a nuclear accented object and optionally prenuclear accented verb as equally "appropriate" in broad VP or narrow object focus contexts. Worth noting however, is that Birch and Clifton found that listeners' response time for rating appropriateness was somewhat slowed for VP focus sentences lacking a prenuclear accent on the verb. This interesting result suggests that, at least in processing, listeners seem to exhibit a preference for a particular broad focus pronunciation (that is, one in which the verb was more prominent) that was not detectable in context matching or appropriateness rating tasks. Thus it may be that when asked to decide whether or not a production is appropriate for a context, listeners are more generous with respect to what constitutes an acceptable pronunciation.

However, a recent study by Breen et al. (submitted) presents an exception to this generalization. As in Gussenhoven's (1983) experiment, they asked listeners to match a speaker's answer productions with the correct question contexts. Unlike Gussenhoven, however, the authors found that listeners were able to do this with considerable success, rarely (13% of the time at most) confusing narrow object focus with broad focus. While the experimental design was similar to that used by Gussenhoven, Welby, and Birch and Clifton, it differed in one important way. As mentioned above, the answer stimuli in Breen and colleagues' study came from speakers whose goal it was to disambiguate their productions. The fact that listeners were able to distinguish the meaning in this kind of speech is not a trivial result. Rather, it provides us with evidence that listeners have expectations about how a narrow focus production should differ from a broad focus production: the cues speakers gave with the intention of disambiguating were effective in leading listeners to recover the intended meaning.

Evidence that listeners have similar expectations about the prosodic realizations of focus *type* is also available, and can be found in Ito and Speer (2008). In their eye-tracking study, the authors showed listeners to respond differently to prominent accent peaks (labeled L+H* by ToBI-trained transcribers) than to less prominent accent peaks (labeled H*), such that the *more* prominent accent peaks evoked a contrastive interpretation. Using the visual world paradigm, listeners in their study were presented with an array of ornaments of different colors and were given instructions for hanging them on a Christmas tree. Ito and Speer found that when asked to hang, for example, a *blue drum* ornament, a more prominent accent on the adjective *blue* in a context like "*Hang the green drum…now hang the BLUE drum."* elicited a contrastive interpretation, indicated by listeners' speeded visual search for (and fixations on) suitable alternatives. This was far less likely to happen if the adjective carried the *less* prominent accent ("*Hang the green drum…now hang the blue drum."*). Indeed an infelicitous use of a prominent accent on an adjective ("*Hang the red angel… now hang the BLUE drum."*) evoked contrastive responses from listeners, causing them to fixate on alternatives to the previously mentioned ornament type, even while hearing conflicting segmental information. Thus, listeners show expectations regarding the prosodic realization of focus type, much like they do for focus size. In the case of focus type, the expectation seems to be that contrastive foci are more prominent than non-contrastive foci.

We have so far discussed a considerable amount of evidence bearing on how the meaning distinctions of focus size and focus type may correspond to prosodic realization. In particular, it was noted that both narrow focus and contrastive focus seem to correlate with greater acoustic prominence (relative to adjacent material in the utterance) compared with broad focus and non-contrastive focus. Further, we have reviewed some of the available evidence for listeners' *knowledge* of these relationships. Before going on to present two experiments that probe the consequences of this knowledge for the perception of prosodic prominence, it is first necessary to distinguish perceived prominence of prosodic units from acoustic prominence, and briefly discuss how the two might relate.

2. Perceived Prominence

Perceived prominence is the listeners' subjective impression of "prosodic strength". Although it is not defined in terms of the acoustic signal, it is well established that a number of acoustic cues serve as predictors of perceived prominence, some of which have been alluded to above in relation to productions of focus size and type. In particular, increased segmental durations,

greater intensity (both contributing to "loudness"), and salient aspects of the fundamental frequency (f0) contour -namely, peaks, valleys and movements- are all associated with perceived prominence. Although there is some disagreement across experimental studies with respect to their relative importance to English listeners (see, for example, Beckman 1986, Gussenhoven et al. 1997, Kochanski et al. 2005), recent studies show that much of the variance in listeners' judgments of prominence can be accounted for on the basis of a model that includes some combination of these features. To take a recent example, Cole et al. (to appear) found that intensity and duration were strong predictors of the probability that linguistically untrained listeners would judge words from unscripted Buckeye (Pitt et al. 2007) corpus excerpts as prominent. Findings consistent with this are reported in Kochanski et al. (2005) and Mo (2008). What this would seem to suggest is that a listener's impression of prosodic prominence can be modeled as a *signal-based* process, the result of acoustic cues.

However, there is also evidence that the perception of prominence emerges from non-signal-based factors. In addition to duration and intensity, Cole and colleagues' study examined correlations between listeners' prominence ratings and lexical and discourse variables. They found that a word's lexical frequency and number of previous occurrences in the experimental materials were negatively correlated with the probability of its being judged as prominent. Although this pattern was consistent with the acoustic prominence of those words (word frequency was also negatively correlated with duration and intensity), the relationship between word frequency and prominence judgments was partly independent of the acoustic features.[3] A result similar to Cole and colleagues' was reported in Eriksson et al. (2001) for Swedish. Their study compared a model of non-signal-based linguistic factors with a model of acoustic predictors. The model with only non-signal-based predictors of listeners' prominence judgments -which included whether or not a syllable was capable of carrying an accent, or whether or not the word was used contrastively- accounted for 57% of the variance in listeners' judgments, an improvement over the 48% accounted for by the acoustics-only model. As in Cole and colleagues' data set, the signal-based and non-signal-based variables in Ericksson et al. (2001) were surely correlated with each other. However, the higher performance of the non-signal-based model is an interesting and suggestive finding. Results of this type suggest that listeners may be making prominence judgments that are consistent with patterns found in productions, but not necessarily the acoustic properties of the stimulus at hand. The question we now wish to ask is whether listeners' experience with productions of different information structures might influence their impressions of prosodic prominence in a similar manner.

3. Probing Listeners' Expectations for Prominence

The previous section discussed evidence that listeners' perception of prominence can be influenced by factors not directly found in the acoustic information they receive, but could be seen to reflect their expectations based on experience with speakers' productions. In Section 1 we discussed evidence bearing on what their experience with productions might include with respect to foci of different sizes. For narrow focus on an object, speakers may tend to use greater acoustic prominence on the object (and less on other material) compared to the same object within a broader VP or sentence focus. In what follows, we present two experiments that were designed to test whether listeners showed any expectation for such a prosody-meaning relation–

[3] A similarly independent effect for the number of repetitions was not yet explored by the authors.

ship. In particular, we asked whether listeners would judge the same auditory stimuli differently depending on their interpretation of the size of the sentence's focus constituent. This was tested in English making use of cases such as (2), repeated here as (5):

(5) a. What happened?
b. What did you do?
c. What did you buy?

d. I bought a motorcycle.

The same pronunciation of (5d), with an (optionally) prenuclear accented verb and (obligatorily) nuclear accented final object, is said to be appropriate in each of the contexts in (5a-c) (Gussenhoven 1983, Selkirk 1995, Birch and Clifton 1995, Welby 2003). However, as discussed above, the information structure of each differs in terms of the size of the focus constituent. Based on previous studies of listeners' prominence judgments, and in line with expectations listeners could have about productions, it is predicted that a narrowly focused object ((5d) heard in the context of (5c)), will be judged as more prominent than an object that is part of a broader focus ((5d) heard in the context of (5a) or (5b)). This might be true even in the absence of any acoustic cues that could disambiguate the three contexts. This is tested in Experiment 1. In Experiment 2, this very same prediction is tested, but for contrastive focus. As noted in Section 1, these two dimensions vary independently, and it is possible for contrastive foci, just like non-contrastive foci, to differ in terms of the size of the constituent that is focused. In both experiments, untrained listeners were asked to listen to sentences in different contexts, each intended to evoke a different interpretation of their information structure, and to assign prominence ratings to the verbs and objects in those sentences. Evidence was found that the size of the focus constituent was a predictor of prominence ratings for both verbs and objects, such that an object was more prominent relative to the verb when narrowly focused. In the case of contrastive focus tested in Experiment 2, this effect was found to be somewhat more robust.

4. Experiment 1

4.1. Method

4.1.1. Materials

Recorded sets of mini-dialogues, question-answer pairs as in (5), above, were used as experimental materials for a prominence rating task. A dialogue set included one version of a subject-verb-object answer sentence such as *I bought a motorcycle*, and three different set-up questions, all WH-questions such as *What happened yesterday?, What did you do yesterday?,* or *What did you buy yesterday?* The answer sentence was to be used as the test sentence for which linguistically naïve listeners would provide prominence ratings. The set-up questions allowed for the pragmatic manipulation of the size of the focus in test sentences, and resulted in three experimental conditions: one in which the entire answer sentence was to be interpreted as a focus (sentence focus), one in which the verb-phrase was the focus (VP focus), and one in which only the object was the focus (object focus) (Table 1; for full list of stimuli used, see the Appendix). The dialogues were recorded and used to create stimuli as follows. Two native speakers of American English (both linguistics graduate students) read 17 sets of question-answer exchanges

Table 1. Example set of question-answer dialogues used in the Experiment 1.

Focus Condition	Question Context	Answer Sentence
Sentence-Foc	What happened yesterday?	I bought a motorcycle.
VP-Foc	What did you do yesterday?	
Obj-Foc	What did you buy yesterday?	

from a printed booklet. The printed dialogues contained no intonational annotations and neither the speaker of the questions (a female), nor the speaker of the answers (a male) was instructed on how to pronounce the sentences, except to read the exchanges as naturally as possible. The speakers were recorded (over two separate channels) while speaking into head-mounted microphones in a sound-attenuated booth. Recordings of the speakers' questions and answers were digitized at 22.05 kHz, saved and stored as separate WAV files.

Because the purpose of the study was to test the independent effect of information structure on the perception of prominence for words in the answer sentences, it was necessary to hold all acoustic information constant in those sentences across the conditions in which they would be presented. This was accomplished by extracting the recordings of answer sentences produced in response VP focus questions in the original recordings, and using them as the test sentences for all three focus conditions. Thus, for example, the production of the answer sentence *I bought a motorcycle* (Figure 1), originally produced as an answer to the VP focus question *What did you do yesterday?*, was made to follow each of the three different questions recorded in that set. Based on previous research, it was highly expected that each of the VP focus answer sentences to be used as stimuli would be pronounced as a single prosodic phrase with the nuclear accent on the object (Birch & Clifton 1995, Beckman 1996, Welby 2003, Breen et al. submitted). Note, however, that since the only restriction on the speakers' readings of the dialogues was that they were impressionistically natural and felicitous, no control was exerted over how the test sentences might have been produced within those simple limitations. In this sense, no *a priori* assumptions about how speakers should produce the test sentences entered directly into the form of the stimuli. However, some prosodic aspects of an "appropriate" answer to these kinds of questions are not predictable for an individual utterance. As discussed earlier, the presence or absence of a prenuclear accent, and also the type of nuclear accent (e.g., Gussenhoven 1983, Selkirk 1995, Ladd 1996, Jun in progress), can vary. It was therefore necessary to identify the intonational pattern of the sentences used as stimuli. Those data are reported here in the form of ToBI (Tones and Break Indices) annotations.[4] Two linguists trained in using the model for Mainstream American English independently transcribed tones (not break indices) for the verb phrase of the 17 answer sentences used as test stimuli. Labeling of test sentences was done without any question context included in the sound file. The tones assigned

[4] The ToBI system for transcribing intonation and prosody (Beckman & Hirschberg 1994), like a number of other such systems, including raP (Dilley and Brown 2005) and ToDI (Gussenhoven 2005), is based on a phonological model of the target language. As such, it necessarily assumes an indirect acoustics-meaning relationship, one mediated by phonological categories. This is not uncontroversial (e.g., Xu & Xu 2005). Nothing presented here crucially depends on this matter, however, and, for the purposes here, use of the transcriptions is mostly practical; MAE_ToBI is a widely used standard for prosodic transcription of American English, and it is assumed to be an effective way to communicate the form of the stimuli used in the present study.

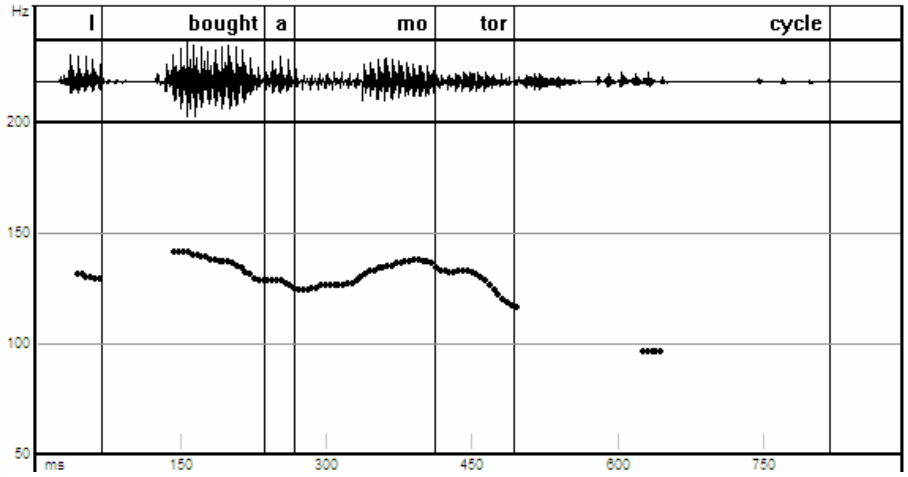

Figure 1. Example waveform and pitch track for the sentence *"I bought a motorcycle."*, recorded as answer to the VP focus question *"What did you do yesterday?"* This production was presented as an answer to each of the questions in the three focus conditions.

Table 2. Intonational structure of the 17 test sentences, described in the form of ToBI transcriptions.

Verb	Object	ip-Boundary	IP-Boundary	# of items
H*	H*	L-	L%	7
H*	!H*	L-	L%	4
H*	!H*	L-	H%	3
H*	H*	L-	H%	2
H*	L+H*	L-	L%	1

by the first labeler are shown in Table 2. In calculating interrater agreement for the two labelers, !H* and L+!H* were collapsed, but the H* versus L+H* distinction was not. Agreement between the two labelers for verbs in the test sentences was 100%, all being transcribed with a prenuclear H*. Agreement for tones on objects was 76.4%. While the two labelers agreed that all objects carried a phrase-final pitch accent, the disagreements that arose regarded whether that pitch accent was a H* or !H*, and Table 2 shows the labeler who tended to label H*.[5] Agreement for boundary tones was 100% (for both intermediate and intonational phrases). The ends of sentences were in most cases marked by L targets, although in five cases the speaker's productions showed a fall from the nuclear pitch accent (associated with the intermediate phrase), followed by a slight rise (associated with the intonational phrase).

[5] In one case, the second labeler annotated an object as ambiguous between a !H* and a delayed L*. The !H* annotation was used here to calculate agreement between raters.

MS PowerPoint presentations were created to present the recorded stimuli. The completed 51 recorded dialogues (17 test sentences, each occurring in three question contexts) were arranged in three different pseudorandomizations, intermixed in each with non-experimental filler dialogues that closely resembled the experimental dialogues in format. In each ordering of the stimuli, members of a crucial set, such as the three in Table 1, were separated by at least 6 dialogue items. The PowerPoint presentations of the stimuli contained only an item number and a play button on each slide which participants used to listen to the items; no visualization of the prosody or acoustics appeared on the screen.

4.1.2. Participants

30 native speakers of American English were recruited from the University of California, Los Angeles to participate as listeners in a prominence rating experiment. All were undergraduate students or (non-academic) employees at the university. Many of the student participants were linguistics or psychology majors, although none had any training in intonational phonology or the transcription of prosody. All participants confirmed they had no previous diagnosis of a hearing or communication disorder, and all were paid for their participation.

4.1.3. Procedure

Listeners participated in a naïve prominence "transcription" task. Listeners were able to proceed through experimental items in the PowerPoint presentation at their own pace, listening to each recorded dialogue as many times as they wished, although they were discouraged from listening more than two or three times. As they played each item, they were to listen for how "stressed" words in the male speaker's answers sounded. The experimenter emphasized to participants that their task was to listen to how the answer sentences were pronounced, and this was described to each participant in the following way:

> "*This experiment is about how speakers pronounce words in a sentence. Your task is to tell us as accurately as possible how stressed the underlined words sound relative to other words in the sentence. By "stressed" we mean "how much did the speaker use his voice to make the word stand out*?"

In many experiments involving prominence judgment tasks, the participant is asked to judge which of two words or which of two accent peaks in a sentence is more prominent (e.g., Pierrehumbert 1979, Rietveld and Gussenhoven 1985, Gussenhoven et al. 1997), or to pick out only prominent rather than non-prominent words in a recording (Cole et al. to appear). However, as mentioned above, all the test sentences used here were transcribed with a nuclear accent on the sentence-final object, and no acoustic manipulation of the stimuli was carried out. Thus, for the stimuli used here, it is highly likely that asking participants to identify the most prominent word in the test sentence would result in 'object' responses in most or nearly all cases. Therefore, a more gradient method of response was used to collect participants' judgments across conditions. Participants were provided with printed transcripts of the dialogues they heard, ordered and numbered as they appeared on the PowerPoint lists presented to them. Participants were instructed to follow along on the transcript and to provide ratings of "stress" from 1 ("not at

all stressed") to 5 ("very stressed") for words that were underlined (verbs and objects in the answer sentences) by writing their ratings above the word. An example of how these items appeared on the transcript, and how they might be rated, is shown in (6):

(6) a. Q: What did you do yesterday?
2 4
A: I bought a motorcycle.

b. Q: What did you eat at the picnic?
2 5
A: I ate a hamburger.

Before beginning the experiment, participants completed a short practice session of three dialogue items to familiarize themselves with the style of the dialogues, the speakers and the general set-up for the task. After completing the practice session and asking questions, participants listened to the 51 test and 37 filler dialogues binaurally over Sony MDR-V500 closed, dynamic headphones at a comfortable listening volume (held constant across participants) in a sound attenuated booth in the UCLA Phonetics Laboratory. They provided prominence ratings as above for verbs and objects in each sentence (30 listeners × 17 test sentences × 2 words (verbs and objects) × 3 focus conditions (sentence focus, VP focus, object focus) = 3,060 ratings). Four measures were considered in evaluating the possible effect of the experimental manipulation on participants' judgments: (a) prominence ratings for verbs, (b) prominence rating for objects, (c) the value of the object's rating as a proportion of the verb's in the same sentence (OV_{ratio}), and (d) the difference between the object's rating and the verb's rating for each sentence (OV_{diff}). The two relative measures, OV_{ratio} and OV_{diff}, were included because it was assumed that prominence is a relative notion, and indeed participants were explicitly instructed to think of it as such. Results for each of the measures were submitted to ANOVA (all were repeated measures). An effect was regarded as significant if $p < .05$ both by subjects and by items, and significant ANOVA effects were further explored using post-hoc Tukey-Kramer pairwise comparisons.

4.2 Results

Figure 2 shows average ratings for verbs and objects in the test sentences for each of the three focus conditions. Repeated measures ANOVA indicated a significant main effect of condition on participant ratings for verbs both by subjects $F1(2,29) = 8.1$, $p < .001$), and by items $F2(2,16) = 8.13$, $p = .01$). A significant effect was found on object ratings by subjects $F1(2,29) = 3.8$, $p < .02$), but only a marginal effect by items $F2(2,16) = 2.5$, $p = .08$). A similar pattern was found for the OV_{ratio} measure, for which an effect was significant by subjects $F1(2,29) = 3.8$, $p < .02$), but only marginally by items $F2(2,16) = 2.4$, $p = .09$). A highly significant effect was found on the OV_{diff} measure both by subjects $F1(2,29) = 9.37$, $p < .001$) and by items $F2(2,16) = 7.68$, $p < .001$). There were no significant interactions between item and condition for any of the measures.

Tukey-Kramer pairwise analyses were used to explore differences among conditions for verb ratings and OV_{diff}. These comparisons showed that verbs were rated lower by listeners in

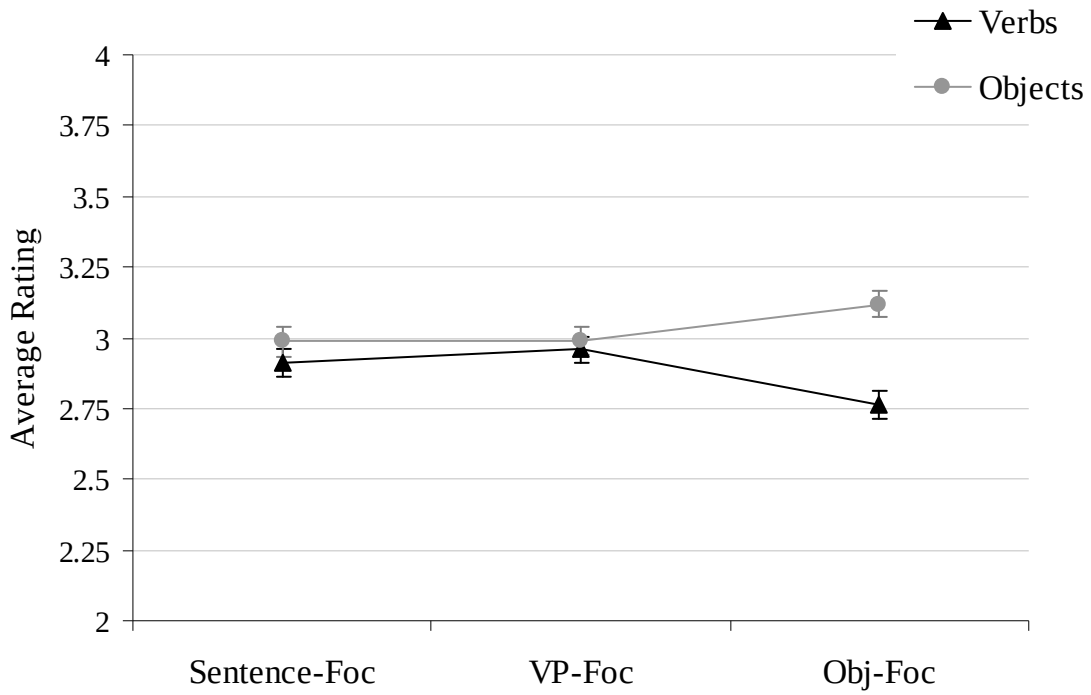

Figure 2. Average prominence ratings for verbs and objects in test sentences in the three focus conditions in Experiment 1. 1 is lowest in prominence, 5 is highest. Error bars show standard error.

the conditions. A significant difference between conditions was also found on the OV_{diff} measure: objects were rated significantly higher (relative to preceding verbs) in the object focus condition compared with either the sentence focus ($p < .001$) or VP focus ($p < .001$) conditions. For neither measure was there any significant difference between the sentence focus and VP focus conditions ($p > .1$ in both cases).

4.3 Discussion

The results of Experiment 1 showed that listeners' judgments about the prominence of verbs in the test sentences were significantly affected by the experimental manipulation of focus size. There was also some evidence that prominence judgments for objects were affected, but this evidence was statistically marginal. There was strong evidence that the *relative* prominence of verbs and objects was affected, as the strongest effect was found on the measure of the difference between object and verb ratings. More specifically, it was found that the effects reflected differences between the narrow object focus condition and the two broad focus conditions. Under narrow object focus, verbs were rated as significantly less prominent by listeners. That the strongest effect was found when objects were considered relative to verbs rather than considered independently suggests that listeners' most consistent way of judging prominence was relative to other words in the sentence, in this case preceding verbs. Note, however, that because the acoustic stimuli listeners judged were identical across the three focus conditions, we can be certain that listeners were not responding to any acoustic differences. Thus we can conclude that listeners' judgments about prominence were influenced by their interpretation of the information

structural differences of the items across conditions, namely the size of the focus. Also notable is that the patterns found are consistent with patterns found in some of the production studies discussed earlier (e.g., Breen et al. submitted).

In Experiment 1 we found that when listeners judged prosodic prominence in sentences that included a non-contrastive focus constituent, they took the size of that focus constituent into account. In Experiment 2, we attempted to replicate this result under slightly different conditions, namely when the focus constituent appeared in (explicitly) contrastive contexts.

5. Experiment 2

5.1 Method

5.1.1 Materials

A second set of 51 short dialogues were recorded by the same two speakers used in Experiment 1, and were analogous in structure to those in Experiment 1. However, in order to test for the effect of focus size for contrastive focus, the dialogues differed in the following ways. All of the questions read by the male speaker were embedded in complementizer phrases headed by *because*, which were themselves preceded by a set-up question (see Table 3; complete list shown in Appendix). For example the female speaker read questions such as *Why aren't you hungry?* and offered a possible reason which was intended as a set-up to the interpretation of the focus structure of the answer. That proposition was then corrected in the answer sentence read by the male speaker. Thus, in the context of *Why's your wife mad?... because you lost your job?"*, the answer "*No, because I bought a motorcycle."* is assumed to be a correction to the VP, *lost your job*. In the context of "*Why's your wife mad?... because you bought a car?"* that same answer is assumed to be a correction only to the object, *motorcycle*. These materials were recorded as in Experiment 1, and productions of the answer sentence spoken in a VP context (e.g., Figure 3) were removed and paired with the three different questions in the same way. ToBI annotations were again assigned to the test sentences by two labelers, and are shown in Table 4. Interrater agreement for accents on verbs was 76.5%, accents on objects 76.5%, boundary tones 100% (all sentences being transcribed with low boundary tones after the object). Disagreements in assignments for accents on verbs regarded the presence or absence of either a H* or !H*, and whether nuclear accented objects bore a H* or L+H*. Rate of agreement as to the presence of a nuclear accent on the object, however, was 100%.

Table 3. Example set of stimuli used in Experiment 2.

Focus Condition	Question Context	Answer Sentence
Sentence-Foc	Why's your wife mad… because the roof's leaking?	No, because I bought a motorcycle.
VP-Foc	Why's your wife mad… because you lost your job?	
Obj-Foc	Why's your wife mad… because you bought a car?	

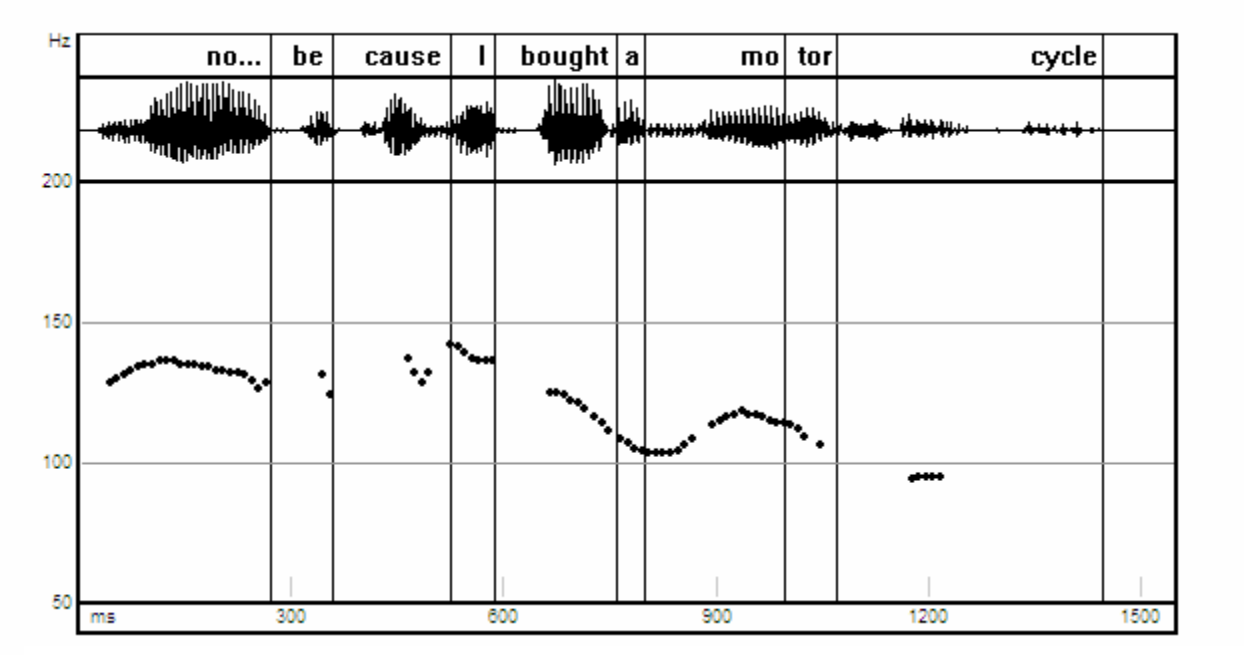

Figure 3. Example waveform and pitch track for the sentence *No, because I bought a motorcycle*, recorded as an answer to the VP focus question *Why's your wife mad…because you lost your job?* This production was presented as an answer to each of the three focus conditions.

Table 4. Intonational structure of the 17 test sentences, described as ToBI transcriptions.

Verb	Object	ip-Boundary	IP-Boundary	# of items
Ø	L+H*	L–	L%	6
H*	H*	L–	L%	3
H*	L+!H*	L–	L%	3
Ø	!H*	L–	L%	2
H*	!H	L–	L%	1
H*	L+H*	L–	L%	1
!H*	L+H*	L–	L%	1

5.1.2 Participants

30 native speakers of American English were recruited from the University of California, Los Angeles as in Experiment 1. Two had also participated in Experiment 1, approximately two months prior. No participant reported any previous diagnosis or knowledge of a hearing or communication disorder; all were paid for their participation.

5.1.3 Procedure

The procedure for Experiment 2 was carried out as for Experiment 1.

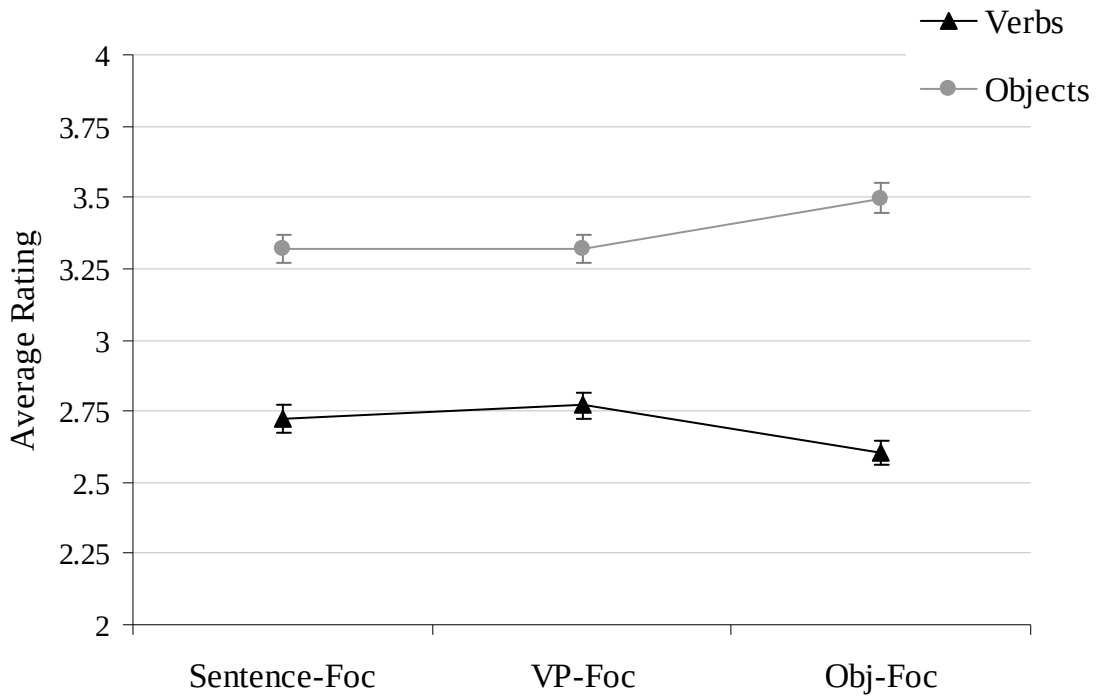

Figure 4. Average prominence ratings for verbs and objects in test sentences in the three focus conditions in Experiment 2. 1 is lowest in prominence, 5 is highest. Error bars show standard error.

5.2 Results

Figure 6 shows mean ratings for verbs and objects in each of the three focus conditions. The results of repeated measures ANOVA indicated a significant main effect of condition on all measures of prominence ratings, both by subjects and by items: verbs $F1(2,29)$, = 6.9, $p < .001$, $F2(2,16) = 3.8$, $p = .02$; objects $F1(2,29) = 9.9$, $p < .0001$; $F2(2,16) = 4.5$, $p = .01$; OV_{ratio} $F1(2,29) = 14.4$, $p < .0001$; $F2(2,16) = 10.6$); OV_{diff} $F1(2,29) = 17.1$, $p < .0001$; $F2(2,16) = 12.5$, $p < .001$. As in Experiment 1, no significant interactions between item and condition were found on any of the measures.

The results of Tukey-Kramer pairwise comparisons showed the following. Nuclear accented objects were rated significantly higher in the object focus condition than in either the sentence focus ($p < .01$) or VP focus ($p < .01$) conditions. Conversely, verbs were rated lower in the object focus condition than in the sentence focus ($p < .001$) or VP focus ($p < .05$) conditions. Both of the relative measures of prominence ratings showed highly significant differences between object focus and the other two conditions as well. According to both OV_{ratio} and OV_{diff}, objects were rated significantly higher relative to a preceding verb in the object focus condition compared to either the sentence focus ($p < .001$) and VP focus ($p < .001$) conditions. There were no significant differences found between the sentence focus and VP focus conditions on any of the measures ($p > .1$ in all cases).

5.3 Discussion

The pattern of results in Experiment 2 closely resembled that found in Experiment 1, which ex–

amined the effect of focus size under non-contrastive focus. Under contrastive focus in Experiment 2, however, the effect of focus size on listener ratings was more robust, significantly affecting all measures. As was the case for non-contrastive focus, verbs were rated lower when focus was narrow on the object, and higher in VP or sentence focus. Objects showed the opposite pattern, being judged as highest in prominence when under narrow object focus. Unsurprisingly, given this pattern, the measure of the difference between object and verb ratings showed a highly significant difference. Again, the effect for condition could not have been the result of listeners having used acoustic information, as the test sentences for which listeners made prominence judgments were held constant.

6. General Discussion

With respect to the most basic question asked, namely whether listeners would respond differently to the same acoustic stimulus depending on their interpretation of its information structure, we are able to answer in the affirmative. Relative to a preceding verb, a sentence-final object was judged as more prominent when under narrow focus than when part of a broader focus constituent. No such differences were found between the two broader focus conditions, sentence and VP focus, however. This basic pattern of results was found to hold in two separate experiments, each testing the effect of the size of the focus constituent for a different focus type, non-contrastive (Experiment 1) and contrastive (Experiment 2).

The question we must now ask is what mechanism was driving these effects? That is, what is it that causes listeners to systematically respond to an object under narrow focus as if it were (relatively) more prominent? Here it is suggested that the answer can be found in production studies discussed in Section 1. A number of studies have now reported that speakers of Western Germanic languages, who were not thought to systematically pronounce broad and narrow focus sentences differently, do. Breen et al. (submitted) have very recently shown that English speakers, especially if they are aware of the potential ambiguity, will produce an object with more acoustic prominence if it occurs in a narrow object focus rather than in a broad VP or sentence focus context. This pattern of prominence is precisely what listeners reported hearing in the experiments above. Thus, a reasonable interpretation of the findings is that listeners' have experience that provides them with a basis for expectations about the prosodic differences between sentences with broad versus narrow focused meanings.

One might find it surprising that listeners should have expectations for a prosodic distinction that is often phonetically subtle, and perhaps only reliably found in situations where the speaker is intentionally hyperarticulating to highlight the distinction. However, that listeners can have expectations for exactly such pronunciations of a contrast is not a novel finding. Consider one relevant demonstration of the phenomenon in the perception of segmental information, namely vowels. Using the "Method of Adjustment" task, Johnson et al. (1993) asked American English-speaking participants to manipulate simple parameters on a speech synthesizer to create a series of English vowels that sounded (to the participant) like those in example words (e.g., "heed", "hid", "aid", "had", "who'd", etc). What the authors found was that participants consistently adjusted their syntheses until each vowel had relatively peripheral F1 and F2 values. Further, their syntheses exaggerated subtle distinctions such as the spectral difference between [eɪ] and [ɪ]. However, although participants' synthetic vowel spaces were more extreme and dispersed than then normal citation-form productions (from the same participants), they were in fact very much like those produced in hyperarticulated contexts. The

authors suggested this finding indicates that hyperarticulated pronunciations are used by listeners to internally represent the linguistic distinction between vowel categories. Similarly, it may be that listeners' behavior in the prominence judgment tasks in the present study reflected their representations, based on experience with the prosodic patterns that most clearly mark the linguistic distinction between broad and narrow focus. If this is in fact the case, it obviously suggests that these two information structural meanings cannot be regarded as mapping onto a single, ambiguous prosodic representation. Such a result is problematic for theories claiming that they do.

Although a direct comparison of the results from Experiment 1 and 2 is not possible because both subjects and items differed, it is worthwhile to discuss some differences before concluding. The first is that within each condition in Experiment 2, the difference between object and verb ratings was much larger than within each corresponding condition in Experiment 1. This is most likely attributable to the acoustic properties of the stimuli used in the two experiments; the ToBI annotations for the stimuli used in Experiment 1 indicated that all non-contrastive test sentences had a prenuclear accent on the verb, and nearly half of the stimuli had a non-prominent !H* as a nuclear accent on the object. This is to be compared with the contrastive test sentences used in Experiment 2, in which nearly half of the test sentences lacked a prenuclear accent on the verb, and downstepped nuclear accents on the object were comparatively rare (approximately a quarter of the stimuli). Thus it is probably safe to assume that objects were auditorily more prominent relative to verbs in the test sentences used in Experiment 2, and listeners were sensitive to this. A second difference between the two experiments, however, namely that the effect for focus size was more robust in Experiment 2, cannot be explained in terms of the acoustic signal. Although the experimental manipulation of focus size resulted in the same pattern of responses in each experiment, under contrastive focus the difference between narrow and broad focus was greater, and, unlike non-contrastive focus in Experiment 1, affected all measures of listener ratings. Because the experiments were not designed to test focus type directly, we cannot draw any strong conclusions with respect to this finding. However, we might note that when production studies find a difference between focus types, contrastive focus is usually associated with greater acoustic prominence than non-contrastive focus. If listeners have expectations for this pattern as it is suggested they do for focus size, it may be that their expectations increased correspondingly when an object is both narrowly and contrastively focused. This would suggest that speakers systematically encode contrast in the prosodic signal, and listeners systematically expect those prosodic features. While the experiments presented here both show consistent effects for the size of the focus on listeners' judgments in the prominence rating task, further research is needed to understand its possible interaction with focus type.

7. Conclusion

What prosodic structure a speaker will assign to a sentence is partially dependent on the sentence's information structure, which in turn is dependent on the context in which it is uttered. The present study explored the consequences of this prosody–meaning relationship for the listener. We probed listeners' expectations for a certain relationship between an information structural contrast -the size of the focus constituent- and patterns of prosodic prominence in simple English SVO constructions. What we found was that listeners responded to broad and narrow focus in a systematically different way. Although they were presented with the same

acoustic information, listeners judged an object as significantly more prominent (and a preceding verb as less prominent), when that object was under narrow focus. Here it was suggested that our answer to why this should be the case can be found in production studies: a very similar pattern is found when we probe speakers' pronunciations of broad and narrow focus. Thus our interpretation is that listeners' have linguistically-relevant experience with how speakers encode the meaning difference prosodically. Such a finding is difficult to understand if speakers and listeners lack any conventionalized knowledge of how this information structural distinction can be expressed in English. Some evidence was also found from the two experiments that contrastive focus, although it evoked the same basic pattern as non-contrastive focus, showed a stronger effect for focus size than did non-contrastive focus; this represents a possible interaction that merits further investigation.

Appendix: Stimuli and Contexts

1.) ***(No… because) I bought a motorcycle.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened yesterday?	Why's your wife mad? Because your roof's leaking?
VP-Foc	What did you do yesterday?	Why's your wife mad? Because you lost your job?
Obj-Foc	What did you buy yesterday?	Why's your wife mad? Because you bought a car?

2.) ***(No… because) I lost my wallet.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you upset? Because of the economy?
VP-Foc	What did you do?	Why are you upset? Because you overslept?
Obj-Foc	What did you lose?	Why are you upset? Because you lost your keys?

3.) ***(No… because) I failed my midterm.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you so worried? Because of the GREs?
VP-Foc	What did you do?	Why are you so upset? Because you're running late?
Obj-Foc	Your grade is really low... what did you fail?	Why are you so upset? Because you failed your quiz?

4.) ***(No… because) I met a girl.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you so happy? Because school's out?
VP-Foc	What did you do at the party last night?	Why's your mom so excited? Because you graduated?
Obj-Foc	Who did you meet at the party last night?	Why are you so happy? Because you met a movie star?

5.) ***(No… because) I read a book.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened at home?	Why were you up so late? Because it was noisy?
VP-Foc	What did you do at home?	Why were you up so late? Because you were doing homework?
Obj-Foc	What did you read at home?	Why were you up so late? Because you read a magazine?

6.) ***(No… because) I passed the final.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened in class?	Why are you so happy? Because it's Friday?
VP-Foc	What did you do in class?	Why are you so happy? Because you finished reading?
Obj-Foc	What did you pass?	Why are you so happy? Because you passed the quiz?

7.) ***(No… because) I bought a car.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened while I was gone?	Why are you so broke all of the sudden? Because of the economy?
VP-Foc	What did you do with all your money?	Why are you so broke? Because you started gambling?
Obj-Foc	What did you buy with all your money?	Why are you so broke? Because you bought a house?

8.) ***(No… because) I rode a Harley.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened today?	Why are you so excited? Because of the game?
VP-Foc	What did you do?	Why are you so excited? Because you went jogging?
Obj-Foc	What did you ride?	Why are you so excited? Because you rode a pony?

9.) ***(No… because) I bought a watch.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What's up?	Why are you so happy? Because of the weather?
VP-Foc	What did you do?	Why are you so happy? Because you talked to Suzie?
Obj-Foc	What did you buy?	Why are you so happy? Because you bought a hat?

10.) ***(No… because) I drove a Porsche.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you so happy? Because of the party today?
VP-Foc	What did you do?	Why are you so happy? Because you went shopping?
Obj-Foc	What did you drive?	Why are you so happy? Because you drove a Mercedes?

11.) ***(No… because) I finished my paper.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you so happy? Because class was cancelled?
VP-Foc	What did you do?	Why are you so happy? Because you went on a date?
Obj-Foc	What did you finish?	Why are you so happy? Because you finished your homework?

12.) ***(No… because) I ate a hamburger.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened at the picnic?	Why aren't you hungry? Because of the medication?
VP-Foc	What did you do at the picnic?	Why aren't you coming to lunch? Because you're dieting?
Obj-Foc	What did you eat at the picnic?	Why aren't you hungry? Because you ate a hot dog?

13.) ***(No… because) I called the doctor.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you feeling so much better? Because of the weather?
VP-Foc	What did you do?	Why are you feeling so much better? Because you slept in?
Obj-Foc	Who did you call?	Why are you feeling so much better? Because you called the nurse?

14.) ***(No… because) I pawned the stereo.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why are you so rich all of the sudden? Because of the stimulus check?
VP-Foc	What did you do?	Why are you so rich all of the sudden? Because you worked overtime?
Obj-Foc	What did you pawn?	Why are you so rich all of the sudden? Because you pawned the T.V.?

15.) ***(No… because) I fixed the roof.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why's your wife so happy? Because of the vacation?
VP-Foc	What did you do?	Why's your wife so happy? Because you took her out to dinner?
Obj-Foc	What did you fix?	Why's your wife so happy? Because you fixed the fence?

16.) ***(No… because) I painted the kitchen.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened?	Why's your wife so happy? Because it's her birthday?
VP-Foc	What did you do?	Why were you busy all day? Because you were working out?
Obj-Foc	What did you paint?	Why's your wife so happy? Because you painted the fence?

17.) ***(No… because) I kissed another cheerleader.***

	Non-Contrastive Context	Contrastive Context
S-Foc	What happened after the game?	Why are you smiling like that? Because of the game?
VP-Foc	What did you do after the game?	Why are you smiling like that? Because you played well?
Obj-Foc	Who did you kiss after the game?	Why are you smiling like that? Because you kissed another pompom girl?

References

Bartels, C. & Kingston, J. (1994). Salient pitch cues in the perception of contrastive focus. In P. Bosch & R. van der Sandt (Eds.), *Focus and natural language processing, Vol. 1: intonation and syntax*, 1-10. IBM Deutschland Informationssysteme GmbH Scientific Center, Institute for Logic and Linguistics.

Baumann, S., Becker, J., Grice, M., Mücke, D. (2008). Tonal and articulatory marking of focus in German. *Proc. of the XVIth International Congress of Phonetic Sciences*, 1029-1032.

Baumann, S., Grice., Steindamm, S. (2006). Prosodic marking of focus domains: categorical or gradient? *Proceedings of Speech Prosody 2006*, 301-304.

Beckman, M. E. (1986). Stress and Non-Stress Accent. *(Netherlands Phonetic Archives 7).* Foris.

Beckman, M. & J. Hirschberg (1994). The ToBI Annotation Conventions. Ms. The Ohio State University.

Beckman, M. (1996). The parsing of prosody. *Language and Cognitive Processes* 11, 17-67.

Bell, A., Jurafsky, D., Folser-Lussier, E., Girand, C., Gregory, M. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113, 1001-24.

Birch, S. & Clifton, C. (1995). Focus, accent and argument structure: effects on language comprehension. *Language and Speech* 38, 365-391.

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (submitted). Acoustic correlates of information structure.

Büring, D. (2007). Intonation, semantics and information structure. In G. Ramchand & C. Reiss (eds.): The Oxford Handbook of Linguistic Interfaces. Oxford University Press.

Chafe, W. (1976). Givenness, contrastiveness, subject, topic, and point of view. In C. Li, (ed.), *Subject and Topic*. New York: Academic Press.

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (to appear). Signal-based and expectation-based factors in the perception of prosodic prominence. Laboratory Phonology 11. Walter de Gruyter.

Dilley, L. & Brown, M. (2005). The RaP (Rhythm and Pitch) Labeling System, Version 1.0. Available at http://tedlab.mit.edu/rap.html.

Dupoux, E., Kaheki, K., Hirose, Y., Pallier, C. & Mehler, J. (1999). Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25, 1568-1578.

Eriksson, A., Thunberg, G., & Traunmüller, H. (2001). Syllable prominence: a matter of vocal effort, phonetic distinctness and top-down processing. *Proceedings of the European Conference on Speech Communication and Technology*, 399–402.

Gussenhoven, C. (1983). Testing the reality of focus domains. *Language and Speech* 26, 61-80.

Gussenhoven, C. (1984). On the grammar and semantics of sentence accents. Dordrecht: Foris.

Gussenhoven, C., Repp, B., Rietveld, A., Rump, W., & Terken, J. (1997). The perceptual prominence of fundamental frequency peaks. *Journal of the Acoustical Society of America* 102, 3009-3022.

Gussenhoven, C. (2005). Transcription of Dutch Intonation. In Jun, Sun- Ah (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press. 118-145.

Hanssen, J., Peters, J., & Gussenhoven, C. (2008). Prosodic effects of focus in Dutch declaratives. *Proceedings of Speech Prosody 2008*, 609-612.

Ito, K., Speer, S., & Beckman, M. (2004). Informational status and pitch accent distribution in spontaneous dialogues in English. *Proceedings of Speech Prosody 2004*, 279-282.

Ito, K. & Speer, S. (2008). Anticipatory effects of intonation: eye movements during instructed visual search. *Journal of Memory and Language* 58, 541-573.

Jackendoff, R. (1972). Semantics in generative grammar. Cambridge MA: MIT Press.

Johnson, J., Flemming, E., & Wright, R. (1993). The hyperspace effect: phonetic targets are hyperarticulated. *Language* 69, 505-527.

Jun, S.-A. (in progress). Focus: domains, types, and realizations. Ms. University of California, Los Angeles.

Katz, J., & Selkirk, E. (submitted). Contrastive focus vs. discourse-new: evidence from prosodic prominence in English.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005) Loudness predicts prominence: fundamental frequency lends little. *Journal of the Acoustical Society of America*, 118, 1038-1054.

Ladd, Robert D. (1996). Intonational phonology. Cambridge University Press.

Löfstedt, I. (2006). On focus types and focus domains. Ms. University of California, Los Angeles.

Mo, Y. (2008). Duration and intensity as perceptual cues for naïve listeners' prominence and boundary perception. *Proceedings of Speech Prosody 2008,* 639-742.

Pitt, M., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). *Buckeye Corpus of Conversational Speech* (2nd release) [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor).

Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America* 66, 363-369.

Rietveld, A. & Gussenhoven, C. (1985). On the relation between pitch excursion size and pitch prominence. *Journal of Phonetics* 13, 299-308.

Samuel, A. (1981). Phonemic restoration: insights from a new methodology. *Journal of Experimental Psychology: General* 110, 474-494.

Selkirk, E. (1995). Sentence prosody: intonation, stress, and phrasing. In J.Goldsmith (ed.). *The Handbook of Phonological Theory*. Oxford: Blackwell, 550-569.

Sityaev, D. & House, J. (2003). Phonetic and phonological correlates of broad, narrow and contrastive focus in English. *Proceedings of the XVth International Congress of Phonetic Sciences,* 1819-1822.

Vallduví, E. & Vilkuna, M. (1998). On rheme and kontrast. *Syntax and Semantics* 29, 79-107. New York: Academic Press.

Welby, P. (2003). Effects of pitch accent position, type, and status on focus projection. *Language and Speech* 46, 53-81.

Warren, R. (1970). Perceptual restoration of missing speech sounds. *Science* 167, 392-393.

Xu, Y. & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation, *Journal of Phonetics*, 33, 159–197.