Phrasing and prominence: A case study of the post-focal context in English

1. Introduction

At the sentence level prosodic structure is defined by prominence and phrasing. A standard hypothesis posits an intimate relationship between prominence and phrasing such that every phrase has a head (the most prominent sub-constituent). I call this *the headedness hypothesis*. Selkirk (1995) and Ito and Mester (2003), for example, propose that headedness holds universally, and is an inviolable constraint. This paper tests this hypothesis by studying the intermediate phrase (iP) in English. If the headedness hypothesis holds, then every iP must have a head. If every iP has a head, then we may further ask how the head of an iP is marked. If it must be marked by pitch accent (I call this *the prominence by pitch accent hypothesis*), then every iP must contain a pitch accent. Beckman (1996) and theories underlying ToBI transcription conventions (Beckman and Ayers Elam 1997; Silverman et al. 1992) have adopted this assumption, and claim that accent-less iPs do not exist.

If every iP must contain a pitch accent, one interesting implication is that constraints that primarily affect pitch accent can have effects on phrasing: if pitch accents are prevented from appearing in some context, then iP boundaries will also be restricted. This paper presents an experiment that tests precisely this prediction, searching for evidence of phrasing in an accent-less context.

The prominence by pitch accent hypothesis has another interesting implication. If the head of an iP must be marked by pitch accent, then in a context with no pitch accent, there should be no head and thus no prominence distinction. This is the position of Bolinger (1958), who proposed that prominence is only realized by pitch accent. A different view hypothesizes that the head of an iP may be marked by something other than pitch accent, for example phrasal stress. This view would make the opposite prediction from the prominence by pitch accent hypothesis, predicting that there can still be prominence distinctions in the absence of pitch accent, as long as the head still has phrasal stress. The experiment to be presented in this paper tests these predictions by searching for evidence of prominence distinctions in an accent-less context.

1.1. Preview of the result and theoretical implications

Our experiment tests the headedness hypothesis and the prominence by pitch accent hypothesis through the study of an accent-less context. To preview the result, we have found evidence suggesting that accent-less iPs exist. This result is still compatible with the headedness hypothesis if we give up the prominence by pitch accent hypothesis, and say that these accent-less iPs still have a head, but that this head is marked not by pitch accent, but by something else such as phrasal stress.

We will now describe how we create an accent-less context to test these questions. In English, words following narrow focus are known to be deaccented, a phenomenon known as ‘post-focal deaccenting’ (Ladd 1980, 1996; de Jong 2004; Xu and Xu 2005). To show what this means, let us first examine a sentence uttered in a neutral context, and then compare its prosody with that of a sentence containing narrow focus. We can control the position of focus by making the sentence an answer to a question and changing what the question asks about. For example, when the sentence ‘Jane will be moving over the chairs’ (1B) answers the

question in (1A), the entire sentence in (1B) is focused (focus is marked by the subscript F; I will refer to (1B) as having *broad focus*). In this case, *Jane*, *over*, and *chairs* tend to carry pitch accents, as indicated by the high pitch accent H* below.

(1)*Broad focus*

Context: The moving company is having a meeting about plans for tomorrow.

A: What will happen?

B: [Jane will be moving over the chairs]$_F$.

H* H* H*

Compare the prosody of (1B) with that of a sentence containing narrow focus. The question in (2A) triggers narrow focus on the subject, *Jane*, in (2B). The narrow focus is realized by the nuclear accent on *Jane*, the strongest pitch accent in the sentence. Furthermore, there is no pitch accent on any word following *Jane*. Typically, this means that F0 falls sharply after *Jane*, and remains low and level through the rest of the sentence.

(2) *Narrow focus*

Context: The moving company is having a meeting about plans for tomorrow.

A: Who will be moving over the chairs?

B: Jane$_F$ will be moving over the chairs.

H*

A simple analysis of this phenomenon posits that focus must be marked by a nuclear accent; and given that nuclear accent is defined as the last pitch accent in a phrase, there can be no pitch accent following focus in the phrase. When the entire sentence is focused, as in (1B), we do not observe post-focal deaccenting because nothing follows focus. Regular pitch accent assignment occurs in the focused phrase (the entire sentence in this case), and nuclear accent falls on the last word *chairs*.

Post-focal deaccenting is particularly interesting for the question of prominence and phrasing. First, let us consider one possible way to phrase a broad-focus sentence. Following Beckman's (1996) view that every intermediate phrase (iP) must contain at least one pitch accent, the broad-focus sentence (1B) may be phrased as follows, with the parentheses indicating iP boundaries:

(3)*Broad focus*

A: What will happen?

B: (Jane) (will be moving over) (the chairs).

H* H* H*

Assuming that every iP must contain at least one pitch accent, the fact that there is no pitch accent after narrow focus implies that there is no iP phrasing post-focally. In other words, post-focal material should be in the same iP as the focus. This is illustrated below, which shows possible ways of phrasing (2B). Due to post-focal deaccenting (i.e. nothing but *Jane* carries pitch accent), (4B) is the only phrasing that is allowed. (4B') has an accent-less iP

(*will be moving over the chairs*); and so do (4B'') (*the chairs*) and (4B''') (*will be moving over* and *the chairs*).

(4)*Narrow focus*

A: Who will be moving over the chairs?

B: (Jane will be moving over the chairs).

B': (Jane) (will be moving over the chairs).

B'': (Jane will be moving over) (the chairs).

B''': (Jane) (will be moving over) (the chairs).

H*

In the experiment to be presented below, we have found evidence of the iP boundary immediately after *over*, suggesting that (4B'') or (4B''') is possible, and hence that there can be an iP with no pitch accent. As we discussed at the beginning of this section, accent-less iPs should not exist if we accept both the headedness hypothesis and the prominence by pitch accent hypothesis. The fact that accent-less iPs exist then suggests that at least one of these hypotheses should be revised.

One possibility is that the headedness hypothesis should be rejected, entailing that not every phrase requires a head. If the head of an iP is marked by pitch accent, then a head-less iP has no pitch accent. This would allow head-less iPs (i.e. accent-less iPs), and allow the phrasing in (4B'-B'''). One theory that exemplifies this possibility is Wagner (2005), which we will discuss in depth in the discussion in section 4.

Another possibility is to reject the prominence by pitch accent hypothesis, and say that the accent-less iPs in (4B''-B''') still have a head. The head does not have to be marked by pitch accent, but can be marked by something else such as phrasal stress. Assuming that pitch accent is generally assigned to phrasally stressed syllables, then by the pitch accenting pattern in (1B), let us assume that the stressed syllables of *Jane*, *over*, and *chairs* carry phrasal stress. Furthermore, let us assume that focus only affects pitch accenting, but not phrasal stress, then these syllables still carry phrasal stress in a post-focal context (2B) and (4B-B'''). The accent-less iPs in (4B''-B''') can still have a head. For example, in the accent-less iP *the chairs*, phrasal stress falls on the first syllable of *chairs*, making it the head of the iP.

1.2. Overview of previous research in the literature

Having presented the hypotheses, how they are tested and a preview of our results, we will now discuss previous research that is related to the current study, and how the current study differs from them. They include studies on post-focal phrasing (in English and other languages), as well as studies on post-focal prominence distinctions. Studies on post-focal phrasing include Wagner and McAuliffe (2019) and Norcliffe and Jaeger (2005), who have found similar results to ours, i.e. that is there is evidence of post-focal phrasing. However, neither of these studies examine prominence distinctions in a post-focal context, or discuss whether the accent-less iPs are still headed. Furthermore, the experiment to be presented in this paper differs from these previous works in experimental design. Previous works in the literature that studied prosodic phrasing often focused on coordinated structures. Wagner and McAuliffe continued this tradition by studying phrasing in coordination such as *A, and B or C*. Our experiment examines a less-studied syntactic structure, the verb phrase domain, and shows that phrasing in this domain is also preserved post-focally. Norcliffe and Jaeger, on the

other hand, do study the verb phrase domain, and are in fact an inspiration for us. But our work extends theirs in significant ways, as discussed in more detail at the end of section 2.

There has been more work on post-focal phrasing in other languages, reaching the same conclusion as in English that phrasing is not completely removed post-focally (e.g. Féry 2010, Hayes and Lahiri 1991, Ishihara 2003, 2007, 2016, Jun 2011, Jun and Fougeron 2000, Kügler and Féry 2017, Sugahara 2003). However, to our knowledge, the question we ask for English in this paper was not asked for those languages, that is whether the post-focal phrases in those languages have heads, and if they do, how these heads are marked.

In addition, there have been previous studies on post-focal prominence distinctions (none of them examined post-focal phrasing at the same time). However, they all focused on word-level stress, and showed that there are word-level stress distinctions in an accent-less environment (e.g. Halliday (1967) and Okobi (2006) for English, and Sluijter and van Heuven (1996) for Dutch). There has been far less work on phrasal stress in an accent-less context, with Horne (1993) being one of the few on this topic; our experiment contributes to the study on phrasal stress.

The next section introduces the syntactic structures that produce the prosodic patterns that in turn allow us to probe the research questions. Section 3 presents the experimental design of our production study. Section 4 discusses the theoretical implications of the results. Section 5 concludes the paper.

2. Syntactic structures with different prosodic patterns

Our experiment looks for evidence of phrasing and phrasal stress post-focally by putting syntactic structures that are usually distinguished by accentuation and phrasing patterns in a post-focal context. Then we can see if they continue to be distinguished prosodically.

First, we identified a suitable syntactic distinction. Then we confirmed that this syntactic distinction does lead to prosodic differences under broad focus using suitable durational measures. Finally, we examined the same syntactic structures in the post-focal context using the same durational measures.

The syntactic distinction that we identified involves the verb phrase domain: it has been observed that under broad focus, when a verb is followed by a word that is ambiguous between a particle and a preposition, different syntactic structures result in different prosodies (Price et al. 1991; Norcliffe and Jaeger 2005). Take the string *moving over* as an example (these earlier studies did not examine this particular string, but it shows the same point):

(5) *Verb + Particle (V + Part)*
Jane will be moving over the chairs.

(6) *Verb + Preposition (V + Prep)*
Jane will be moving over Christmas.

The following are the waveforms of one of our speaker's productions of these two sentences. The blue lines indicate pitch.

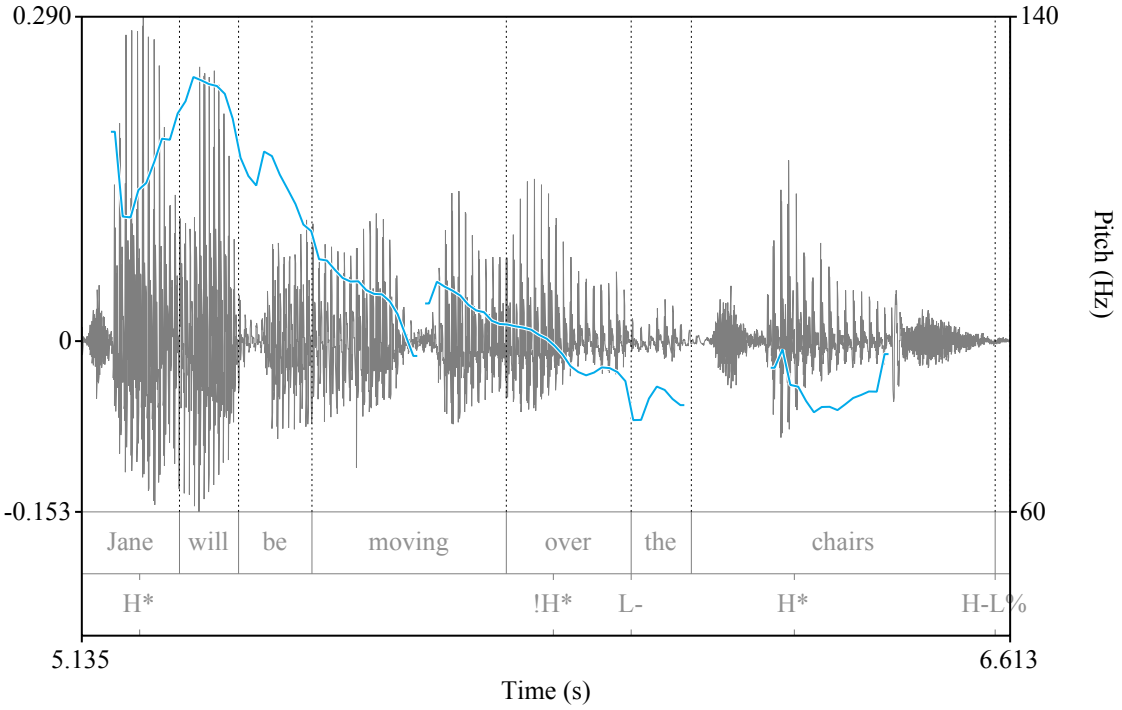

Figure 1: *Waveform of a Verb +Particle sentence.*

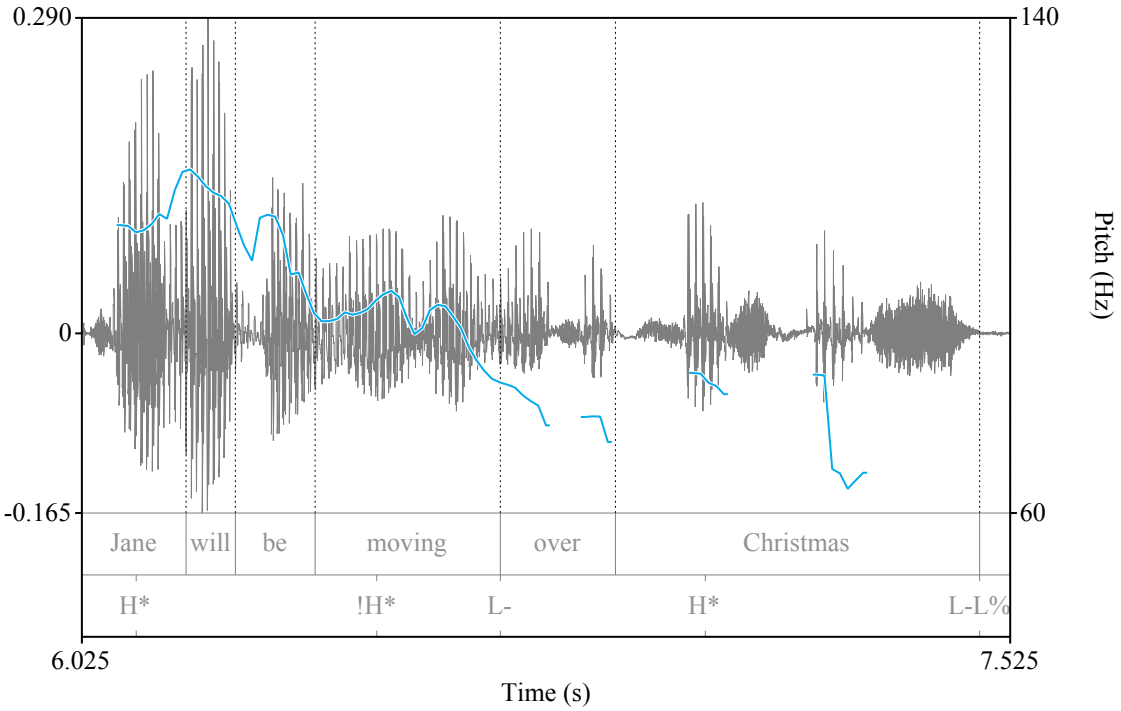

Figure 2: *Waveform of a Verb +Preposition sentence.*

We focused on the region extending from the verb to the particle / preposition, and found the following generalizations for all of our recordings. They are also consistent with Price et al.'s (1991) observations:

- The particle has a pitch accent, whereas the preposition does not.
- The verb tends to carry a pitch accent in V+Prep, but not in V+Part.
- The prosodic boundary between the verb and the particle is smaller than the boundary immediately following the particle, whereas the boundary between the verb and the preposition is greater than the boundary immediately following the preposition.

As we will see shortly, the evidence for the third bullet above is primarily durational. We also observed an L tone between the verb and the object. It is most plausible that this is a low pitch phrase accent L-, and it is associated with the stronger phrase boundary diagnosed by durational evidence (i.e. the boundary immediately following the particle and the boundary immediately following the verb in V+Prep).

We annotate our observations in ToBI transcription terms (see (7) and (8) below), where H* indicates a high pitch accent, L- indicates a low pitch phrase accent, and ')' indicates an iP boundary. Accents mark phrasal stress. We call the stressed/first rime of the verb "V1" (e.g. the *o* of *moving*); the unstressed/final rime of the verb is termed "V2" (e.g. the *ing* of *moving*); the stressed/first rime of the word that is ambiguous between a particle and a preposition (short for P) is termed "P1" (e.g. the *o* of *over*); and lastly the unstressed/final rime of P is "P2" (e.g. *er* of *over*). All the V words and P words that we use in the experiment are disyllabic and trochaic, so the stressed rime is always the first rime, and the unstressed rime is always the final/second rime.

(7) *Verb + Particle (V + Part)*
Jane will be moving over the chairs.

		V1		V2	P1		P2	
…	m	o	v	ing	ó	v	er	…
					H*		L-)	

(8) *Verb + Preposition (V + Prep)*
Jane will be moving over Christmas.

		V1		V2	P1		P2	
…	m	ó	v	ing	o	v	er	…
		H*		L-)				

Our transcriptions of pitch accenting and phrasing can be substantiated using durational measures. First, phrasing triggers pre-boundary lengthening effects: elements that precede a prosodic boundary are lengthened in English and other languages (e.g., Price et al. 1991; Shattuck-Hufnagel and Turk, 1996; Wightman et al., 1992). In particular, Wightman et al. found that the segmental lengthening that is most significantly correlated with the perceived size of a boundary falls on the rime of the final syllable before the boundary. The bigger this break is, the longer the rime. Thus, we take the duration of the final rime as an indication of the strength of the prosodic boundary.[1]

[1] Turk and Shattuck-Hufnagel (2007) found that while the majority of the phrase-final lengthening effects fall on the final rime, a non-final stressed rime is lengthened as well. Thus, we take the duration of the final rime to be the *main* indicator of boundary strength. The first rime also reflects the boundary strength, but to a much lesser

If our hypothesis is correct that the prosodic boundaries immediately following V and P differ in strength, then this difference should lead to different pre-boundary lengthening effects. Specifically, because in V+Part, the boundary after P is stronger than the boundary after V, P2 should be lengthened more than V2. Conversely, in V+Prep, the boundary after V is stronger than the boundary after P, so V2 should be lengthened more than P2. And if we compare these two syntactic structures, the ratio of V2 to P2 should be greater in V+Prep than in V+Part.

One might think that the pre-boundary lengthening effects should also lead to an absolute difference in durations, that is, V2 should be longer in V+Prep than in V+Part, and P2 should be longer in V+Part than in V+Prep. The reason we compared the ratio of V2 to P2 rather than their absolute durations is due to Carlson, Clifton and Frazier's (2001) and Wagner's (2004) observation that the strength of a phrase boundary is not marked absolutely, but relative to surrounding phrase boundaries in the utterance. For example, in V+Prep the boundary immediately following V is stronger than the boundary immediately following P; but if they are correct, then we can't say the boundary immediately following V must be stronger in absolute terms in V+Prep than in V+Part.

In addition, our observation about pitch accent placement can be substantiated by durational measures as well. Dimitrova and Turk (2012) have observed that the rime of an accented syllable is longer than the rime of an unaccented syllable. Since we observed that the verb tends to be pitch accented in V+Prep but not in V+Part, V1 should be on average longer in V+Prep than in V+Part. Since we observed that the particle has a pitch accent and the preposition does not, P1 should be longer in V+Part than in V+Prep.

The research question addressed in this paper is whether these prosodic distinctions found under broad focus are maintained in a post-focal context. If phrasing is preserved post-focally, we expect the relative strength of phrase boundaries to be preserved as well, and thus the ratio of V2 to P2 should still be greater in V+Prep than in V+Part. If prominence distinctions are preserved post-focally, then elements that are lengthened due to accentuation in broad focus should still be lengthened post-focally (i.e. post-focal V1 should be longer in V+Prep than in V+Part, and post-focal P1 should be longer in V+Part than in V+Prep). Such a result would suggest that phrasal prominence can still be realized in the absence of pitch accent.

If these prosodic distinctions are preserved post-focally, we can further ask whether they are similar in size compared to the prosodic distinctions under broad focus. If so, the differences between the two syntactic structures should not differ significantly between broad focus and narrow focus. On the other hand, if the difference in a post-focal context is significantly smaller than the difference under broad focus, then the prosodic differences in terms of phrasing and prominence between the two syntactic structures will be minimized post-focally.

Our study was inspired by Price et al. (1991) and Norcliffe and Jaeger (2005). In particular, one of our research questions is the same as Norcliffe and Jaeger's, that is, whether an iP can lack pitch accents. The environment in which we investigated this question is the same as theirs as well: it is the verbal domain. However, the current study differs from theirs in important ways. Norcliffe and Jaeger did not control the syllable and stress structure of the words in their stimuli, but instead used a mix of monosyllabic and disyllabic words such as *won* and *over*. Using monosyllabic words like *won* might confound two separate effects,

---

degree than the final rime. As we discuss shortly, we take the duration of the first rime to mostly reflect prominence.

lengthening due to prominence, and lengthening due to occupying a phrase-final position. In its V+Prep use, *won* both has pitch accent and immediately precedes an iP boundary. In its V+Part use, *won* neither has pitch accent nor immediately precedes an iP boundary. Thus, when Norcliffe and Jaeger measured the length of the monosyllabic verb and compared its duration between syntactic conditions, the lengthening effects they found were actually cumulative and a result of both prominence and phrase-finality. Hence, they confounded these two effects. For this reason, the current study only uses disyllabic and trochaic words, so that the first/stressed rime reflects lengthening effects due to prominence (mostly, see footnote 1), and the second/unstressed rime reflects pre-boundary lengthening effects.

3. The experiment

3.1.Materials

The speech materials for our experiment consisted of dialogs in four conditions (2 syntactic structures x 2 focus structures). The two syntactic conditions were V+Part and V+Prep, and the two focus conditions were broad focus and narrow focus (on the sentential subject). Each item was a short story with three roles: the narrator, who introduces the context of the dialog, the question-asker, and the question-answerer. (9) and (10) are examples. The answer to the question was the target sentence of interest. The syntactic structure of the target sentence was controlled by the context sentence. Its focus structure was elicited with the *wh*-question. (9) is an item that elicits V+Part and broad focus in the answer. (10) elicits V+Part and narrow focus.

(9) Context: The moving company is having a meeting about plans for tomorrow.
Question: What will happen?
Answer: Jane will be moving over the chairs.

(10) Context: The moving company is having a meeting about plans for tomorrow.
Question: Who will be moving over the chairs?
Answer: Jane will be moving over the chairs.

The speaker/participants in our experiment were instructed to play the three different roles (narrator, question-asker, question-answerer) by uttering the context, the question, and the answer sentences in the given order. We had 20 sets of experimental items that varied across the four conditions (V+particle vs. V+preposition and broad vs. narrow focus). The dialogs differed in lexical content, for a total of 80 target sentences. There were 73 filler items, which were items for two other experiments.

3.2.Participants

We conducted a production study with 13 native speakers of North American English (8 female, 5 male); our subjects were all students at XXX University. They were remunerated with a small sum for their time. At the end of the experiment, we asked each participant to tell us what they thought the purpose of the experiment was, and no one was able to make a correct guess. We suspect the large number of filler items and the diverse form of items helped conceal the purpose of the experiment.

3.3.Data collection

Recording took place in a sound-attenuated booth at XXX University. Participants were seated in front of a computer, which displayed one context-question-answer trio at a time. The stimuli plus fillers were presented in pseudo-randomized order, and minimal pairs were not placed next to each other. Participants were given instructions about the task at the beginning of the experiment, which asked them to first read each trio quietly to themselves, and only proceed to read it out loud when they were ready. They were asked to imagine they were playing three different roles in each trio, and to act out the dialogues naturally rather than reading the sentences mechanically. If the participants were not satisfied with their rendition of an item (a common reason was that they stumbled over some words), they were allowed to say it again. If they asked to repeat an item, we only considered the rendition they were happy with, and discarded the previous renditions.

3.4.Data annotation

Two research assistants labeled V1, V2, P1, and P2 manually in Praat (Boersma and Weenink 2018), each assistant annotating half of the recordings. Before annotation, they listened to every recording, and marked the ones that needed to be excluded (the exclusion criterion to be explained in the next subsection). Then they labeled the rimes following an annotation manual that was developed for this experiment that detailed the cues to segmental boundaries for each rime (the manual is included in the supplementary materials along with the list of items). For example, the criterion for the boundary between a stop and a following vowel (e.g. the beginning of V2, [ing] in *moving over*, is the end of frication noise created by [v]).

To check for consistency between annotators, each assistant cross-annotated a random sample of the other person's work (96 target sentences) without looking at the other person's annotations; the result was agreement on most of the exclusion judgments (93% of the cross-annotated recordings) as well as the locations of the rime boundaries. For the syllable rimes that were labeled by both annotators, on average the duration annotated by one assistant differed from the other assistant's annotation by 11.7 ms.

3.5.Data exclusion

Disfluent utterances were excluded, which were 1% of total (N=1040) target sentences. One speaker constantly produced disfluent speech, putting a pause between every pair of words, so this speaker was excluded. One of our item sets used *blowing over* as the V+P combination, and one speaker consistently pronounced *blowing* as *blown*, so this set of items for this speaker was excluded, which was 0.4% of the total number of target sentences.

Because we were interested in what happens in an accent-less region, we made sure there were no pitch accents in the target region. We excluded any utterance that was supposed to be narrow-focus, but nevertheless had a pitch accent in the post-focal region extending from the word immediately following the subject of the sentence up to P; this exclusion comprised 6% of the total number of target sentences. The research assistants based their exclusion judgments by listening to the recording to make sure they heard no prominence up to P, and examining the pitch tracks in Praat to make sure the pitch fell after the focused subject and remained low and level without any spike. Figure 3 is an example of a good post-focal target sentence. In contrast, Figure 4 is an example of an item that was supposed to have narrow focus on the subject and was excluded because it has a post-focal pitch accent.

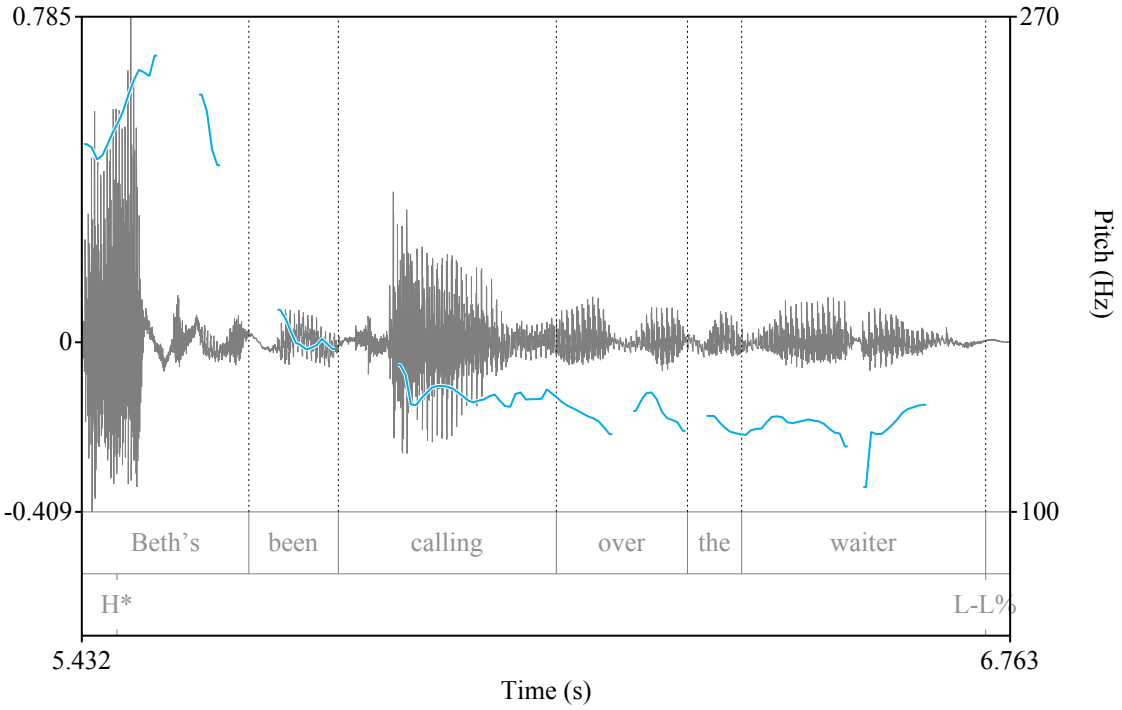

Figure 3: *Waveform of an utterance with no post-focal pitch accent.*

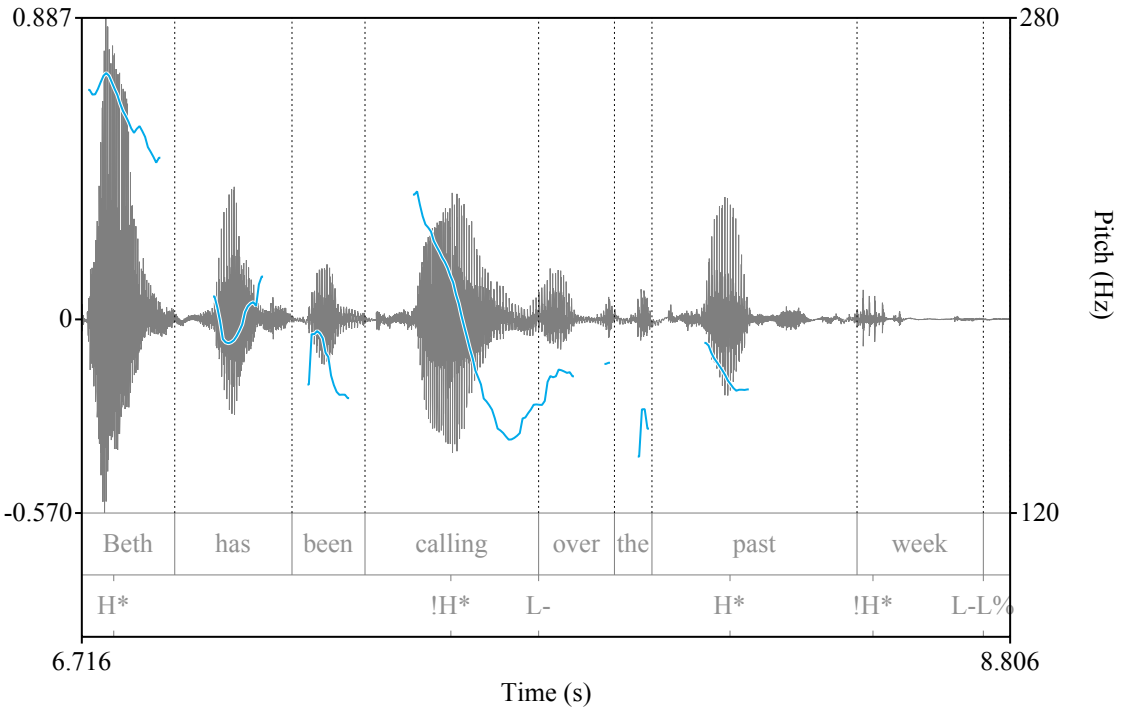

Figure 4: *Waveform of an utterance with post-focal pitch accent.*

Note that the research question for our study is not whether under broad focus V+Part and V+Prep *always* create the prosodies we expected, as in (7) and (8). The research question is: when a speaker produces the expected prosody for a broad-focus sentence, whether the they

preserve the prosodic differences post-focally. So broad-focus utterances whose intonational patterns are the opposite of what we'd expect in (7) and (8) are not good test grounds of our research question. These are V+Part sentences with a pitch accent on the verb but not the particle, and V+Prep sentences with a pitch accent on the preposition but not the verb. We excluded these broad-focus sentences as well as their narrow-focus counterparts, which totaled 4% of the total number of target sentences.

3.6.Data analysis

We fitted three linear mixed-effects models. We calculated p values using Satterthwaite's degrees of freedom method (using the R package 'lmerTest'), with the durations of V1 and P1, and the log of the ratio of the duration of V2 to the duration of P2 (*log(V2/P2)*) as the dependent variable. The fixed effects were syntax (V+Part vs. V+Prep), focus (broad vs. narrow), and their interaction. We chose log(V2/P2) instead of V2/P2 because V2/P2 was not normally distributed. In order to directly see the difference between the syntactic conditions under each focus condition, we contrast coded syntax, and dummy coded focus. Following Matuschek et al. (2017), the models included random intercepts and slopes by speaker and item where those effects made a significant contribution to model fit ($p < 0.2$). In addition to the main models, we also conducted three post hoc tests for the broad-focus material only. They are linear mixed effects models that have the identical structure as the main models, with the only difference being that there is only one fixed effect variable, syntax.

3.7.Predictions and results

Our experiment is testing several hypotheses. First, if a larger prosodic phrasing break induces more pre-boundary lengthening, and if as we observed, under broad focus, the boundary immediately following the particle is stronger than the boundary between the verb and the particle, then we expect P2 (the final rime of P) to be lengthened more than V2 (the final rime of the verb) in V+Part structures. The reverse should be true for V+Prep structures, where V2 should be lengthened more than P2. This means that the ratio of the lengths of V2 to P2, as well as the log of this ratio, should be greater in V+Prep structures than in V+Part structures.

Our results confirmed this prediction. Under broad focus, log(V2/P2) is significantly greater in V+Prep than in V+Part ($p<0.01$) (see the box plots on the left side of Figure 5). Specifically, the ratio of the durations of V2 to P2 in V+Part structures is 70% of the ratio in V+Prep structures.

The second hypothesis we tested can be expressed as follows. If pitch-accented rimes are lengthened, and if as we observed with our ToBI transcriptions under broad focus, the verb is more likely to have a pitch accent in V+Prep structures than in V+Part structures, then we expect V1 to be longer in V+Prep structures than in V+Part structures. Likewise, since we observed that the particle has a pitch accent while the preposition does not, then P1 should be longer in V+Part contexts than in V+Prep contexts.

Our results confirmed these predictions as well. Under broad focus, V1 is on average 5.3 ms longer in V+Prep structures compared to V+Part structures ($p<0.001$) (see the box plots on the left side of Figure 6). Moreover, P1 is on average 10.6 ms longer in V+Part contexts than in V+Prep contexts ($p<0.01$).

Having seen the predictions for broad focus are borne out, we can proceed to the answer to our principal the research question: what happens when the target material follows narrow focus? If boundaries are removed post-focally, we expect no pre-boundary lengthening effects. Neither V2 nor P2 should be lengthened, whether in V+Part or in V+Prep

environments. Due to the lack of lengthening of V2 or P2, the ratio of the durations of V2 to P2, as well as the log of this ratio, should not differ between V+Part and V+Prep constructions. On the other hand, if iP boundaries persist post-focally, we expect pre-boundary lengthening effects on V2 and P2 with more lengthening on V2 than P2 in V+Prep contexts, and more lengthening on P2 than V2 in V+Part contexts. This implies a significant difference between the two syntactic structures, where log(V2/P2) should be greater in the V+Prep context than in V+Part context. Furthermore, if we find such a difference post-focally, we may ask if the size of this difference is larger or smaller than the difference under broad focus.

Our results suggest that iP boundaries are retained post-focally. We find that in post-focal position, log(V2/P2) is significantly greater in V+Prep contexts compared to V+Part contexts ($p<0.01$).[2] The ratio of V2's duration to P2's duration in the V+Part context is 79% of that in the V+Prep context. The post-focal difference in the log(V2/P2) ratio between the V+Prep and V+Part contexts does not differ significantly from the ratio difference under broad focus. Figure 5 shows the log(V2/P2) ratios under broad focus and narrow focus.

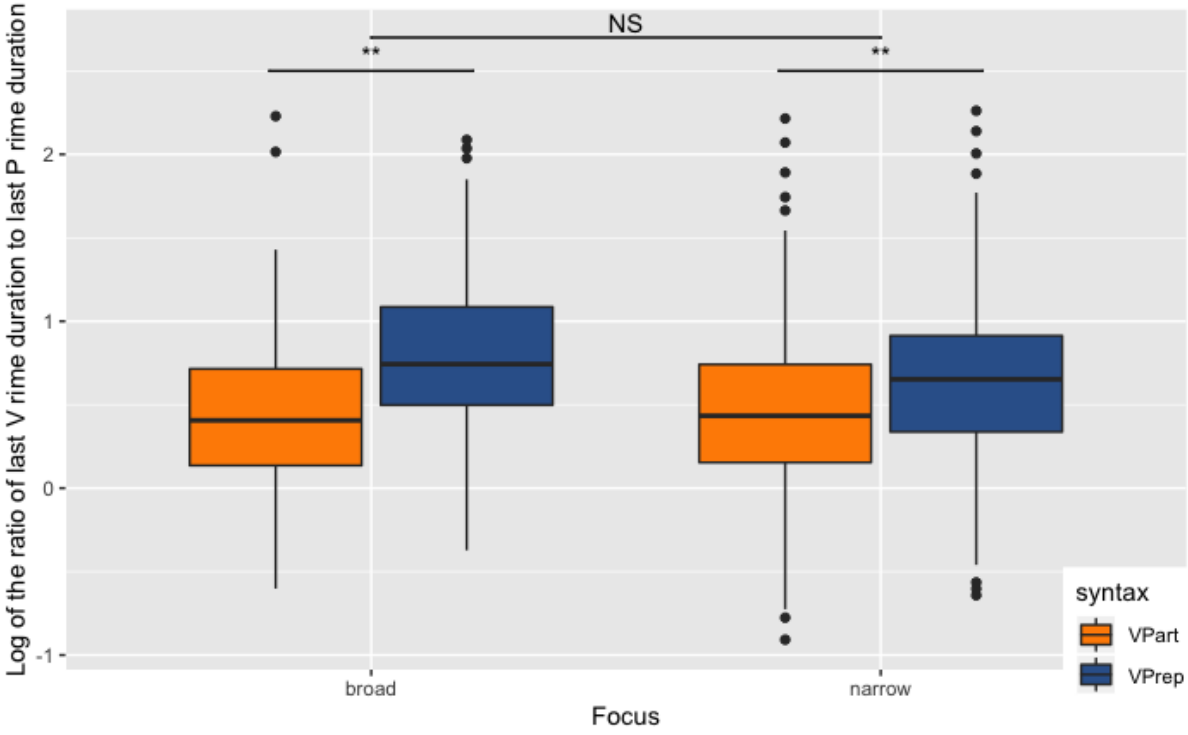

Figure 5: *Log of ratio of verb's second rime duration to P's second rime duration in broad focus and narrow focus.*

Having seen the results for the log(V2/P2) ratio, let us now turn to V1 and P1. There is no pitch accent after narrow focus. If the phrasal prominence distinctions also disappear, then the verb should not be more prominent than the preposition, and neither should the particle be

[2] The full model including an interaction between the random variables gave a convergence warning and had a singular fit, suggesting that its random effects' structure was too complex. A simplified model gave no convergence warning and had no singular fit, so we present the results of the simplified model here. The full model yielded similar results to the simplified model, producing similar coefficients and identical significance levels.

more prominent than the verb. We should thus expect V1 to be no longer in V+Prep than in V+Part, and P1 to be no longer in V+Part than in V+Prep. On the other hand, if the phrasal prominence distinctions still exist despite deaccentuation in the post-focal region, then the verb should still be more prominent than the preposition, and the particle should be more prominent than the verb.

Concerning this prediction, our results are mixed, so it was difficult to determine whether phrasal prominence is preserved post-focally. We found no significant difference in the duration of post-focal V1 rimes between the V+Part and the V+Prep contexts (Figure 6). Nor did we find a significant difference in the duration of post-focal P1 rimes (Figure 7). These results might suggest that there is no evidence for post-focal prominence on the stressed V rime or the stressed P rime. However, there was also no significant interaction between syntax and focus. In other words, the difference between the syntactic structures does not align with the broad focus and narrow focus distinction. Although the difference in post-focal V1 duration does not reach statistical significance, it is also not significantly smaller than the difference in broad focus V1 duration. The same is true for P1 duration.

Furthermore, we found that focus significantly reduced the duration of most rimes, with V1 being on average 6.8 ms shorter post-focally compared to the broad focus context ($p<0.01$), and with P1 on average 22.0 ms shorter post-focally compared to the broad focus context ($p<0.001$). Post hoc tests suggest that V2 is on average 11.73 ms shorter post-focally than in the broad focus context ($p<0.05$), and P2 is not significantly shorter post-focally than in broad focus. This overall finding is consistent with our impression that post-focal material is shorter in duration and more reduced compared to the same material in the broad focus context.

One possible explanation for the failure to detect a post-focal prominence distinction between the syntactic conditions may be post-focal shortening. Under broad focus, the effect size of the difference between the syntactic conditions was already small (a 5.3 ms difference for V1 and 10.6 ms for P1). Since post-focal rimes are shorter than pre-focal rimes, the already small difference between the syntactic conditions might be reduced further post-focally.

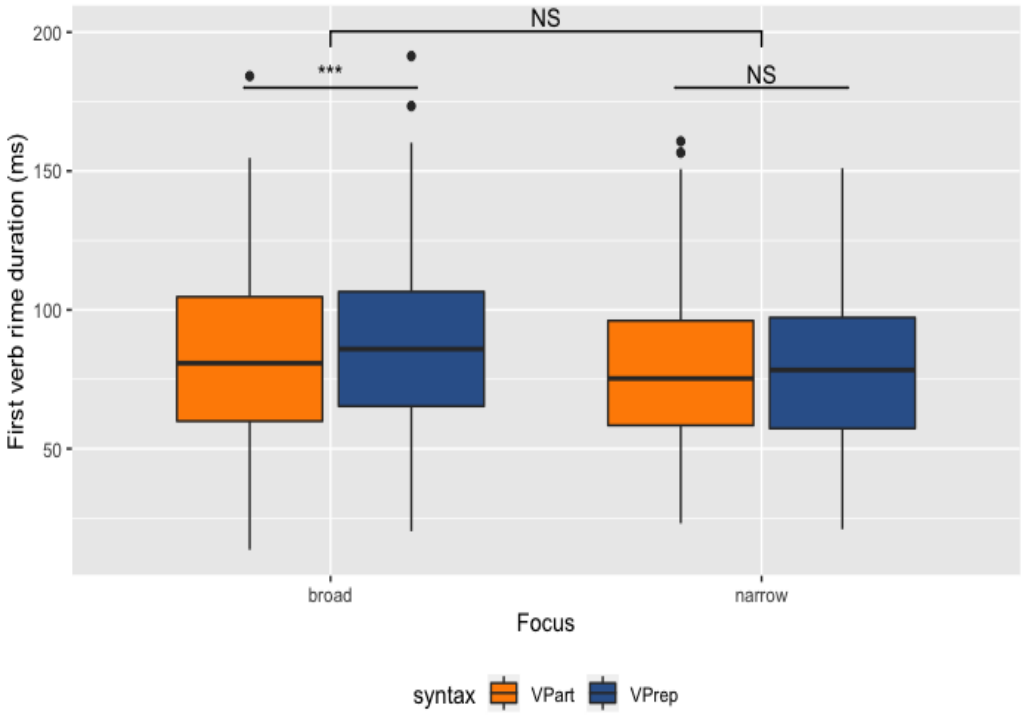

Figure 6: *Duration of the verb's stressed rime in different syntactic structures in broad focus and narrow focus.*

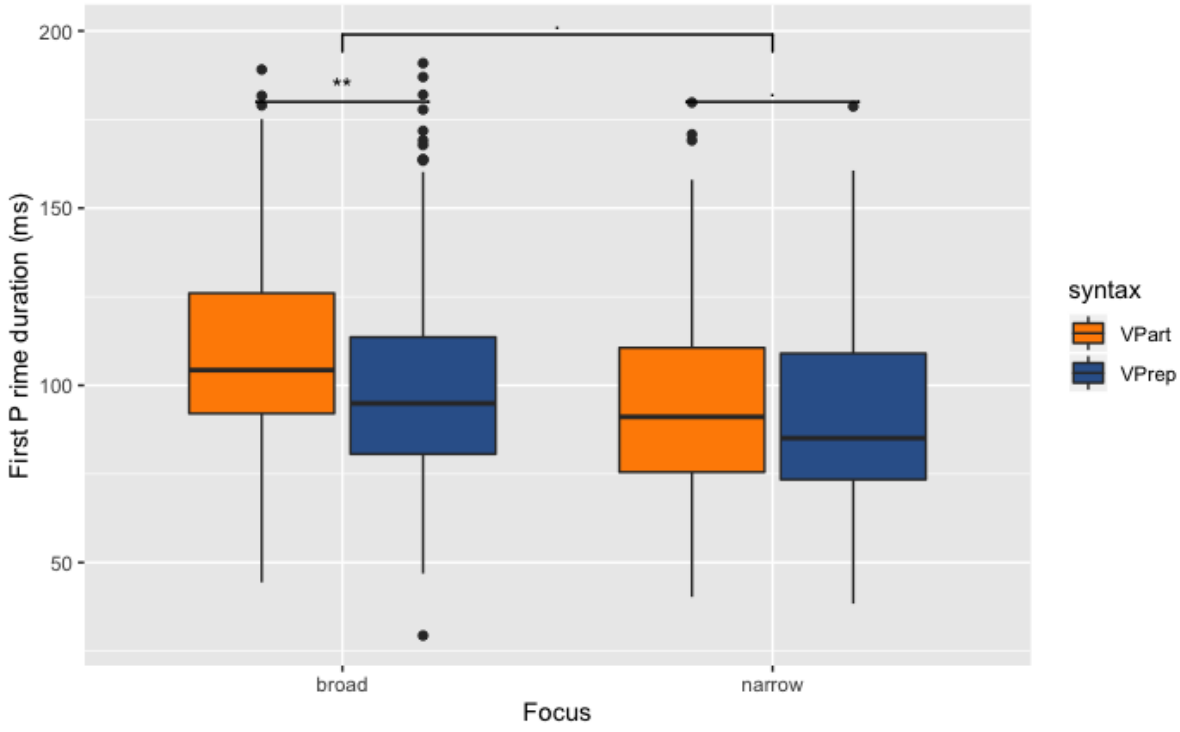

Figure 7: *Duration of P's stressed rime in different syntactic structures in broad focus and narrow focus.*

4. Discussion

We now turn to a discussion of the implications of our experimental results, and evaluate some logically possible theories that lie behind them. After that, we will briefly address the question of whether the post-focal iPs that we found might contain a phrase accent.

The relative strength of phrase boundaries is preserved post-focally, as is indicated by the preservation of the relative degree of pre-boundary lengthening (Figure 5). The next question is exactly what these phrase boundaries are, whether they are iP or larger phrase boundaries. Because the pitch in the post-focal context is generally flat, it is difficult for us to observe boundary tone phenomena, which usually indicates the phrase boundary. Thus, we approach this question indirectly, by first examining these phrase boundaries under broad focus, a context which has more pitch movements and cues to phrasing. Then we infer what the post-focal phrase boundaries should be based on the phrase boundaries under broad focus.

Under broad focus, we often found a phrase accent at the boundary following the particle on the one hand, and at the boundary between the verb and the preposition on the other. Since a phrase accent is generally assumed to occur at an iP boundary, we infer that under broad focus, the stronger phrase boundary distinguishing these two contexts is at least an iP. A transcription of a random sample of sentences indicated that the weaker boundary (the boundary that does not have a phrase accent, i.e. the boundary following the verb in V+Part, and the boundary following the preposition) is a prosodic word boundary. Thus, we take the stronger boundary between V and P to be an iP boundary, and the weaker boundary to be a prosodic word boundary. This is consistent with Price et al.'s (1991) observations.

Post-focally, we found that the relative boundary strength is preserved. Since the relative boundary strength is the same as under broad focus, if the weaker boundary between V and P is at least a prosodic word boundary (and we confirm this is true), then by inference the stronger boundary between V and P must be at least an iP.

Since post-focal contexts undergo deaccentuation (loss of pitch distinctions), this result suggests that an iP can have no pitch accent, contra Beckman (1996). This result is consistent with the view that prosodic phrasing tracks syntactic constituency, and is not affected by focus. Since in English, focus does not affect syntactic phrasing, prosodic phrasing should not be affected by focus, either. This is the position held by Wagner (2005).

As discussed in the introductory section 1, this result bears on the headedness hypothesis and the prominence by pitch accent hypothesis, and calls for a revision of at least one of these proposals. Either prosodic phrases don't have to have a head, or alternatively if every phrase still has a head, then the head of an iP can be marked by phrasal stress rather than by a pitch accent.

For concreteness, we will evaluate a few specific and logically possible theories, and will ground this discussion by putting the theories that we compare in metrical representations. We follow the common assumption that the metrical grid provides the basis for stress assignment, and that pitch accent is assigned to the syllables with the greatest stress (syllables on the highest levels of the metrical grid) (Pierrehumbert 1980).

The metrical grid consists of a hierarchy of phonological elements. The levels relevant to us are syllable (σ), prosodic word (PW), intermediate phrase (iP), and intonational phrase (IP). I assume the following link between prosodic structure and stress: syllables project to different levels on the metrical grid, and the higher their projection, the stronger their stress.

The following would be the metrical representation of a V+Part sentence under broad focus such as (1B). *Jane*, *over*, and *chairs* are pitch-accented, and their stressed syllables

project to iP. Nuclear accent falls on *chairs*, which projects to the IP level. Pitch accent H* is assigned to syllables on the iP level.

(11) *Metrical analysis of V+Part under broad focus.*

(								x)	IP
(x)	(				x	)	(	x)	iP
(x)	(x)	(x)	(x	)	(x	)	(x)	(x)	PW
(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)	σ
Jane	will be		moving		over		the	chairs.	
H*					H*			H*	

When the subject is narrowly focused, the subject bears the nuclear accent, so it must be the most prominent in the IP. Meanwhile, all post-focal elements lose their pitch accents. If we assume further that an iP must have at least one pitch accent (the position held by Beckman 1996), then all the post-focal material must belong to the same iP as the subject, as in the following representation in (12). This representation satisfies both the headedness hypothesis (every phrase has a head) and the prominence by pitch accent hypothesis (the head of iP, *Jane*, carries a pitch accent).

(12) *Analysis of V+Part under narrow focus if every iP must have at least one pitch accent.*

(x								)	IP
(x								)	iP
(x)	(x)	(x)	(x	)	(x	)	(x)	(x)	PW
(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)	σ
Jane	will be		moving		over		the	chairs.	
H*									

This representation is not compatible with our experimental findings, as we found evidence of an iP phrasing boundary between *over* and *the*. One way to resolve this conflict is to abandon the prominence by pitch accent hypothesis, and assume instead that every iP only needs to contain at least one *phrasal stress*. Then we can derive the following metrical representation (13) for the narrow-focus sentence, where there is iP phrasing post-focally that will account for the boundary lengthening differences uncovered in our experiment.

(13) *Analysis of V+Part under narrow focus if every iP must have at least one phrasal stress.*

(x) IP

(x) (x) (x) iP

(x) (x) (x) (x) (x) (x) (x) PW

(x) (x) (x) (x) (x) (x) (x) (x) (x) σ

Jane will be moving over the chairs.

H*

Suppose the pitch accent is assigned to the highest level of this metrical grid (IP level); then only *Jane* gets a pitch accent. This representation satisfies headedness: the IP has a head (*Jane*), and each iP has a head (*Jane*, *over*, *chairs*). While the head of the IP (*Jane*) is marked by pitch accent; the heads of the iPs (*over* and *chairs*) are marked only by phrasal stress.

Another way to accommodate our findings is to abandon headedness, but maintain prominence by pitch accent. Since prosodic phrases are not required to contain a head, we can have head-less iPs post-focally. Since these iPs have no head, they contain no pitch accent, either. An advocate for the prominence by pitch accent hypothesis is Bolinger (1958), who claimed that with the removal of pitch accent, prominence distinctions disappear as well. Under this view, the narrow-focus sentence may be analyzed as in (14). We keep the iP phrasing post-focally, but these iPs now have no head. Pitch accent is assigned to the iP heads, which in this case is only *Jane*.

(14) *Analysis of V+Part under narrow focus without post-focal prominence distinctions.*

(x) IP

(x) ( ) ( ) iP

(x) (x) (x) (x) (x) (x) (x) PW

(x) (x) (x) (x) (x) (x) (x) (x) (x) σ

Jane will be moving over the chairs.

H*

Thus, there are two possible ways to accommodate our experimental findings, that is by abandoning either the headedness hypothesis, as in (13), or the prominence by pitch accent hypothesis, as in (14). These alternatives make different predictions about the prominence distinctions post-focally. While (13) still allows for post-focal prominence distinctions (e.g. *over* is the head of the iP and therefore more prominent than *moving*), (14) does not (e.g. *over* should be as prominent as *moving*). Unfortunately, our experimental results are mixed in this respect, and do not distinguish between these alternatives.

Finally, I will discuss a third logical possibility that combines these two approaches, and challenges both the headedness hypothesis and the prominence by pitch accent hypothesis.

This theory does not require phrases to be headed. Nor does it require the head of an iP to be marked by pitch accent. An existing theory that exemplifies this idea is Wagner (2005). Wagner only requires the “strongest prosodic phrase” out of all the phrases in a sentence (i.e., the phrase with the strongest boundary, which we will explain shortly) to contain a pitch accent; weaker prosodic phrases don’t have to contain any pitch accent. Furthermore, the metrical grid proposed by Wagner does not contain any prosodic labels like iP and IP. Since there is no concept of iP, there is certainly no prominence by pitch accent requirement.

In order to understand Wagner’s theory, we will first show how it would analyze a broad-focus sentence. The boundaries in Wagner’s unlabelled metrical grid are mapped from syntactic boundaries. We assume the following syntactic structures for V+Part and V+Prep: the particle forms a constituent with the verb, which then merges with the object (15a); by contrast, the preposition merges first with the object, and then this prepositional phrase merges with the verb (15b).

(15) *Syntactic constituency of V+Part and V+Prep*

a. [moving over] [$_{DP}$ the chairs] *Verb + Particle*
b. [moving] [$_{PP}$ over the chairs] *Verb + Preposition*

Here is Wagner’s recipe for mapping from syntax to the metrical grid: more deeply embedded syntactic constituents are separated from each other by weaker boundaries in the grid than less deeply embedded constituents. The following are the representations of the V+Part and the V+Prep sentences under broad focus in an unlabelled grid. In V+Part (15a), the verb head *moving* is more deeply embedded than the object *the chairs*, therefore the boundary around *moving* (the boundary between *moving* and *over*) is weaker than the boundary around the object (the boundary between *over* and *the*) in the corresponding grid (16). The reverse is true for V+Prep (15b), so the boundary between *moving* and *over* is stronger than the boundary between *over* and *the* in the corresponding grid (17).

(16) *Analysis of V+Part under broad focus in an unlabelled grid.*

```
(x)  (                         x       ) (     x   )
(x)  (x)  (x)   (x         ) (x        ) (     x   )
(x)  (x)  (x)   (x)  (x )  (x)  (x )  (x)  (x    )
Jane   will be    moving      over     the    chairs.
H*                            H*              H*
```

(17) *Analysis of V+Prep under broad focus in an unlabelled grid.*

```
(x)  (             x       ) (           x       )
(x)  (x)   (x)   (x        ) (x        ) (x      )
(x)  (x)   (x)   (x)  (x ) (x)   (x ) (x)   (x)
Jane    will be     moving      over     Christmas.
```

H* H* H*

There are two important differences between Wagner's unlabelled grid and traditional metrical grid representations. First, the levels are not labelled, so there is no concept like iP and IP. Second, there are three syllables that are equally strong on the highest level. Traditional metrical grid representations distinguish nuclear stress from non-nuclear stresses. How does the unlabelled grid distinguish nuclear stress from non-nuclear stresses then? Wagner argues that they are not distinguished in the metrical grid, and the reason why the last stress (the stress on *chairs*) sounds the strongest is due to perception. Following Newman (1946) and citing Pierrehumbert's (1979) and Liberman and Pierrehumbert's (1984) findings, Wagner claims that the last stress in a series of equal stresses is perceived to be the strongest. According to Wagner, this perceptual principle creates the impression that *chairs* has the nuclear stress of the sentence.

If we adopt Wagner's approach, the representation of a V+Part sentence under narrow focus would be the following.

(18) *Analysis of V+Part under narrow focus in an unlabeled grid.*

(x)
(x) (x) (x)
(x) (x) (x) (x) (x) (x)
(x) (x) (x) (x) (x) (x) (x) (x) (x)
Jane will be moving over the chairs.
H*

In order to generate the desired effect that the focused subject has nuclear stress, it must project to a higher level than the following syllables, and the unfocused elements are subordinated. This is represented by the creation of a fourth level, to which only the syllable of the narrowly-focused element projects. Phrasing and headedness on the lower levels remain intact. Interestingly, this representation is in fact identical to the labelled grid (13), except for the absence of labels.

Is there any difference between the labelled grid (13) and the unlabelled grid (18) that would lead to observable differences? We think there are, but the evidence would come from

the last iP head in the sentence (*chairs*), an element that we have not focused on in this paper, and therefore will be left for future research.

We explain briefly how it could tease apart (13) and (18). Wagner's theory would predict that post-focally, *chairs* is acoustically equal to but perceptually stronger than the particle. This is because Wagner assumes that equal stresses (including subordinated stresses) are represented equally and thus equal acoustically, but the last stress is nevertheless perceived to be the strongest. In contrast, the traditional metrical grid representation in (13) would predict that post-focally, *chairs* is equal to the particle because they both project to the iP level, and there is no perceptual principle that would lead to a greater perceived strength of *chairs* relative to the particle. A perception study comparing prominence of post-focal elements would tease apart (13) and (18).

Another result of our experiment is that all the rimes we measured are shortened post-focally, except P2. Assuming that rimes are generally shorter post-focally than pre-focally, the question is which theory so far can account for this shortening effect. Let us assume a very simple correlation between rime duration and its position on the metrical grid: the higher a rime projects in the grid, the longer its duration. Then the fact that the duration of every rime is shortened post-focally suggests that post-focally, every rime should project to a lower level than it did under broad focus. No theory that was discussed so far makes this prediction, and we leave this question to future research.

Having presented the alternative theories that might accommodate our experimental findings, the rest of this subsection will briefly discuss another question, that is whether the post-focal iPs that we found contain a phrase accent. Following the common assumption that every iP must end with a phrase accent (e.g. by Beckman and Pierrehumbert 1986), the accent-less iPs we found post-focally should contain a phrase accent.

It is difficult to tell from most speakers' productions, given that they always produced L- phrase accents in a broad-focus declarative sentence. Based on their behavior in broad-focus sentences, we infer that if they were to put a phrase accent under narrow focus, that phrase accent should be L- as well. Take a narrow-focus V+Part sentence as an example. If every iP must end with a phrase accent, and there is an iP boundary right after the particle, then these speakers should produce an L- at this iP boundary (19b), and (19a) is not allowed. If on the other hand, an iP does not have to contain any phrase accent, then (19a) is allowed. All the narrow-focus utterances from these speakers have an H* on the subject and end with L%.

(19) *Possible distribution of phrase accents in a narrow-focus V+Part sentence.*

Jane	will be	moving	over	the	chairs.
a. H*					L-L%
b. H*			L-		L-L%

The typical pitch pattern that we have observed for a narrow-focus sentence is: the pitch falls shortly after the H*, and remains low and steady until the end of the sentence. This pattern is consistent with (19a), where there is only one phrase accent after H*. This pattern is also consistent with (19b), if the F0 target for the two phrase accents is the same. Suppose, following Pierrehumbert (1980) and Beckman and Pierrehumbert (1986), that an L- phrase accent spreads its tone, filling the space between the last pitch accent and the phrasal boundary. When the two phrase accents have the same F0 target, (19b) would be realized in

the same way as (19a). Thus, our observation is consistent with both (19a) and (19b), and does not tease them apart.[3]

If the speaker's phrase accent isn't always L-, we may be able to differentiate (19a) and (19b) because that will create more varied pitch movements than a speaker that only produces L-, thus creating more opportunities to identify a sentence-medial phrase accent. One of our speakers did differ from the majority pattern, and sometimes produced H- phrase accents in broad-focus declaratives. We infer that if they were to produce a phrase accent post-focally, it would sometimes be H- and sometimes be L- as well. However, since the source of the evidence comes from only one speaker, we do not want to draw any definitive conclusion from it, but will rather leave this question to future research.

5. Conclusion

This paper reports a production study of English prosody in post-focal contexts. We presented results that suggest that iP boundaries are preserved post-focally, and that an iP may have no pitch accent. We also studied whether there are prominence distinctions in an accent-free context. Our results are mixed here. They are consistent with the hypothesis that every phrase must have a head, and the head of an iP can be marked by something other than phrasal stress. They are also consistent with a different hypothesis, that phrases don't have to have a head, but if they do, the head of an iP must be marked by pitch accent.

References

Bailly, G., & Holm, B. (2005). Sfc: a trainable prosodic model. *Speech Communication, 46*(3–4), 348–364.

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

Beckman, M. E., & Ayers Elam, G. (1997). Guidelines for ToBI labeling, version 3: Ohio State University.

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

Bolinger, D. (1958). A theory of pitch accent in English. *Word, 14*, 109–149.

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program].

Carlson, K., Clifton, C. Jr., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language*, *45*, 58–81.

Dimitrova, S., & Turk, A. (2012). Patterns of accentual lengthening in English four-syllable words. *Journal of Phonetics, 40*, 403–418.

[3] If the two phrase accents in (19b) have different F0 targets, (19b) may be realized differently from (19). (19b) has two L-s, both of which would presumably spread. The first L- would fill the space between the H* and the end of P; the second L- would fill the space between the H* and the end of the sentence. There is no existing theory for what happens when a space is filled with two L-s (in the case of (19b), it is the space between H* and the end of P). Suppose the pitch falls shortly after H* to the F0 level of the first L-, and remains low through the end of P. After that, the pitch falls sharply again right after the end of P to the F0 target of the second L-, and remains low through the end of the sentence. Thus, we may be able to observe two sudden drops in F0 for (19b): one drop might occur right after the H* and the other after the end of P; and the pitch drop in the region from the H* to the end of P may differ in slope from the drop from the end of P to the end of the sentence. We have not observed such a pattern for any of our speakers.

Fujisaki, H. (1981). Dynamic characteristics of voice fundamental frequency in speech and singing. acoustical analysis and physiological interpretations. KTH Department of Speech, Music and Hearing. Quarterly Status Report.

Gerazov, B., Bailly, G., Mohammed, O., Xu, Y., & Garner, P. N. (2018). A variational prosody model for mapping the context-sensitive variation of functional prosodic prototypes. arXiv:1806.08685.

Halliday, M. A. K. (1967). *Intonation and Grammar in British English*, The Hague: Mouton.

Horne M. (1993). The phonetics and phonology of the rhythm rule in post-focal position: Data from English. *Reports from the Department of Phonetics, University of Umea* (Sweden), 2, 69–77.

de Jong, K. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics, 32(4)*, 493–516.

Ito, J. & Mester, A. (2003). Weak layering and word binarity. In: Takeru Honma, Masao Okazaki, Toshiyuki Tabata, and Shin-ichi Tanaka (Eds.), *A New Century of Phonology and Phonological Theory. A Festschriftfor Profssor Shosuke Haraguchi on the Occasion of His Sixtieth Birthday*, 26-65. Tokyo: Kaitakusha.

Ladd, R. (1980). *The Structure of Intonational Meaning. Evidence from English*. Bloomington: Indiana UP.

Ladd, R. (1996). *Intonational Phonology*. Cambridge: Cambridge UP.

Liberman, M. Y., & Prince, A. On stress and linguistic rhythm. *Linguistic Inquiry, 8*, 249–336.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type 1 error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315.

Norcliffe, E., & Jaeger, T. F. (2005). Accent-free prosodic phrases? Accents and phrasing in the post-nuclear domain. *Proceedings of Interspeech*.

Öhmann, S. (1967). *Word and sentence intonation: A quantitative model Quarterly Status Report*. KTH Department of Speech, Music and Hearing.

Okobi, A. O. (2006). Acoustic correlates of word stress in American English. Ph.D. dissertation, MIT.

Pierrehumbert, J. B. (1980). The phonetics and phonology of English intonation. Ph.D. dissertation, MIT.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America, 9*, 2956–2970.

Silverman, K., & Pierrehumbert, J. B. (1990). The timing of prenuclear high accents in English. In John Kingston & Mary Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. 71-106. Cambridge: Cambridge University Press.

Singmann, H., & Kellen, D. (2019). An Introduction to Mixed Models for Experimental Psychology. In D. H. Spieler & E. Schumacher (Eds.), *New Methods in Cognitive Psychology* (pp. 4–31). Psychology Press.

Sluijter, A. M. C., & van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America, 100*, 2471–2485.

Wagner, M. (2004). Prosody as diagonalization of syntax. Evidence from complex predicates. *Proceeding of NELS 34*, 587–602.

Wagner, M. (2005). Prosody and recursion. Ph.D. dissertation, MIT.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic boundaries. *Journal of the American Acoustic Association*, *91*, 1707–1717.

Xu, Y. (1999). Effects of tone and focus on f0 contour formation. *Journal of Phonetics* 27, 55-105.

Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication, 46*, 220–251.
Xu, Y., & Xu., C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics, 33(2)*, 159–197.