

There is no Post-focal De-phrasing in English

Danfeng Wu

Massachusetts Institute of Technology

1 Introduction

Prosodic structure at the sentence level 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, proposed headedness as an inviolable constraint that holds universally. 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, 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) adopted this assumption, and claimed that accent-less iP's 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 My experiment tests the headedness hypothesis and the prominence by pitch accent hypothesis through the study of an accent-less context. To preview the result, I have found evidence suggesting that accent-less iP's exist. This result is still compatible with the headedness hypothesis if we give up the prominence by pitch accent hypothesis, and say that the accent-less iP's still have a head, but this head is marked not by pitch accent, but by something else such as phrasal stress.

I first describe how I create an accent-less context to test these questions. In English, words following narrow focus are known to be deaccented ('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 (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.

* Thanks are due to undergraduate research assistants Aquila Simmons and Kitty Zeng for their help with collection and annotation of the data presented here. Thanks are also due to Edward Flemming for constructive comments and helpful feedback throughout this project. I also thank Athulya Aravind, Charles Chang, Martin Hackl, Michael Kenstowicz, David Pesetsky, Roger Schwarzschild, and the audiences at AMP 2020 and the MIT Phonology Circle for discussion. This project was supported by the Ken Hale Fund of MIT Linguistics.

(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 (1B) with 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 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. (4) 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).

H* H*

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

H* H*

In the experiment to be presented, I have found evidence of the iP boundary immediately after *over*, suggesting that (4B'') or (4B''') is possible, and hence there can be an iP with no pitch accent. As I discussed at the beginning of this section, accent-less iP's should not exist if we accept both the headedness hypothesis and the prominence by pitch accent hypothesis. Thus, 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 iP's (i.e. accent-less iP's), and allow the phrasing in (4B'-B'''). One theory that exemplifies this possibility is Wagner (2005), which I will discuss in depth in the discussion in section 5.

Another possibility is to reject the prominence by pitch accent hypothesis, and say that the accent-less iP's 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), *Jane*, *over*, and *chairs* carry phrasal stress. Furthermore, assuming that focus only affects pitch accenting, but not phrasal stress, these words still carry phrasal stress in a post-focal context (2B) and (4B-B''). Thus, the accent-less iPs in (4B''-B'') can still have a head. For example, in the accent-less iP *the chairs*, *chairs* has phrasal stress, making it the iP head.

Section 2 introduces the syntactic structures that produce the prosodic patterns that in turn allow me to probe the research questions. Section 3 presents the experimental design of my production study. Section 4 presents the results, and section 5 discusses their theoretical implications. Section 6 concludes the paper.

2 Syntactic structures with different prosodic patterns

The experience 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. To do so, I first identified a suitable syntactic distinction. Then I confirmed that this syntactic distinction does lead to prosodic differences under broad focus using suitable durational measures. Finally, I examined the same syntactic structures in the post-focal context using the same durational measures.

The syntactic distinction that I 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:

- (5) Jane will be moving over the chairs. *Verb + Particle (V + Part)*
 (6) Jane will be moving over Christmas. *Verb + Preposition (V + Prep)*

The following are the waveforms of one speaker's productions of (5) and (6). The blue lines indicate pitch.

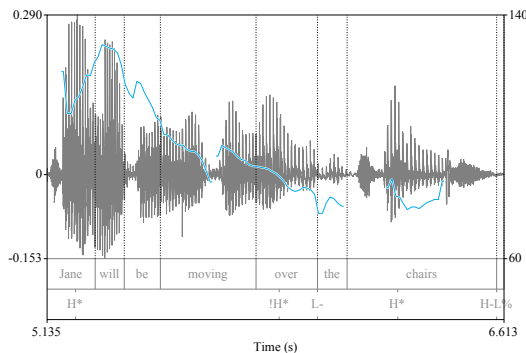


Figure 1: *Waveform of a Verb + Particle sentence (5).*

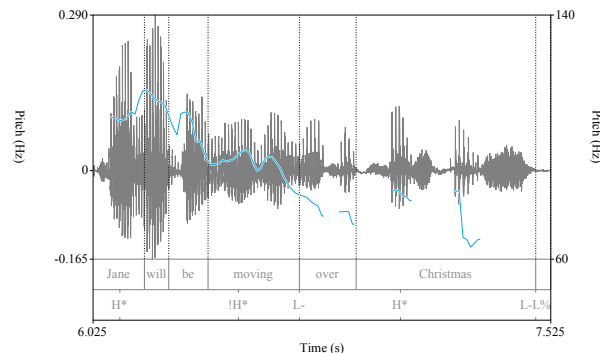


Figure 2: *Waveform of a Verb + Preposition sentence (6).*

I focused on the region extending from the verb to the particle / preposition, and found the following three generalizations for all of my recordings. They are also consistent with Price et al.'s (1991) observations. First, the particle has a pitch accent, whereas the preposition does not. Second, the verb tends to carry a pitch accent in V+Prep, but not in V+Part. Third, 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 observation above is primarily durational. I 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).

I annotate my observations in ToBI transcription terms in (7) and (8), where H* indicates a high pitch accent, L- indicates a low pitch phrase accent, and ')' indicates an iP boundary. Accents mark phrasal stress. I 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 I 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) Jane will be moving over the chairs. *Verb + Particle (V + Part)*
- | | | | | | | | | | |
|-----|---|----|---|-----|----|----|----|-----|--|
| | | V1 | | V2 | | P1 | | P2 | |
| ... | m | o | v | ing | ó | v | er | ... | |
| | | | | | H* | | | L-) | |
- (8) Jane will be moving over Christmas. *Verb + Preposition (V + Prep)*
- | | | | | | | | | | |
|-----|---|----|---|-----|---|----|----|-----|--|
| | | V1 | | V2 | | P1 | | P2 | |
| ... | m | ó | v | ing | o | v | er | ... | |
| | | H* | | L-) | | | | | |

My 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, I take the duration of the final rime as an indication of the strength of the prosodic boundary.¹

If my 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 I 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 I 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. 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, my 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 I 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 I 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, I 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, I 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

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

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.

My study was inspired by Price et al. (1991) and Norcliffe and Jaeger (2005). In particular, one of my research questions is the same as Norcliffe and Jaeger's, that is, whether an iP can lack pitch accents. The environment in which I 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 – 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 materials 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). I had 20 sets of experimental items that differed in lexical content and varied across the four conditions, for a total of 80 target sentences. There were 73 filler items.

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. The answer to the question was the target sentence, whose syntactic structure was controlled by the context sentence, and whose 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 speakers/participants 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.

3.2 Participants I conducted a production study with 13 native speakers of North American English (8 female, 5 male), who were all students at MIT. They were remunerated with a small sum for their time. At the end of the experiment, I asked each participant to tell me what they thought the purpose of the experiment was, and no one was able to make a correct guess. I 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 MIT. 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 words), they were allowed to say it again, in which case I only considered the rendition they were happy with, and discarded the previous renditions.

3.4 Data exclusion and annotation Before annotation, I excluded disfluent utterances (1% of total target sentences). Because I was interested in what happens in an accent-less region, I excluded utterances with a pitch accent in the post-focal region (6% of total). In addition, I did not intend to study whether V+Part and V+Prep *always* create the prosodies described in (7) and (8), but to ask when they do, whether these prosodic differences are preserved post-focally. Therefore, I excluded broad-focus utterances whose prosody was not as expected by (7) and (8), as well as their narrow-focus counterparts (4% of total).

Two research assistants labeled V1, V2, P1, and P2 manually in Praat (Boersma and Weenink 2018), with each assistant annotating half of the recordings. 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 analysis I fitted three linear mixed-effects models. I 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. I chose $\log(V2/P2)$ instead of $V2/P2$ because $V2/P2$ was not normally distributed. 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, I 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.

4 Results

My experiment tested several hypotheses. First, if a larger prosodic phrasing break induces more pre-boundary lengthening, and if as I observed, under broad focus, the boundary immediately following the particle is stronger than the boundary between the verb and the particle, then I 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.

My 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 3). 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 I tested can be expressed as follows. If pitch-accented rimes are lengthened, and if as I observed with my 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 I expect V1 to be longer in V+Prep structures than in V+Part structures. Likewise, since I 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.

My 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 4). 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, I can proceed to answer my principal research question: what happens when the target material follows narrow focus? If boundaries are removed post-focally, I 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, I 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 I find such a difference post-focally, I may ask if the size of this difference is larger or smaller than the difference under broad focus.

My results suggest that iP boundaries are retained post-focally. I find that in post-focal position, $\log(V2/P2)$ is significantly greater in V+Prep contexts than 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 3 shows the $\log(V2/P2)$ ratios under broad focus and narrow focus.

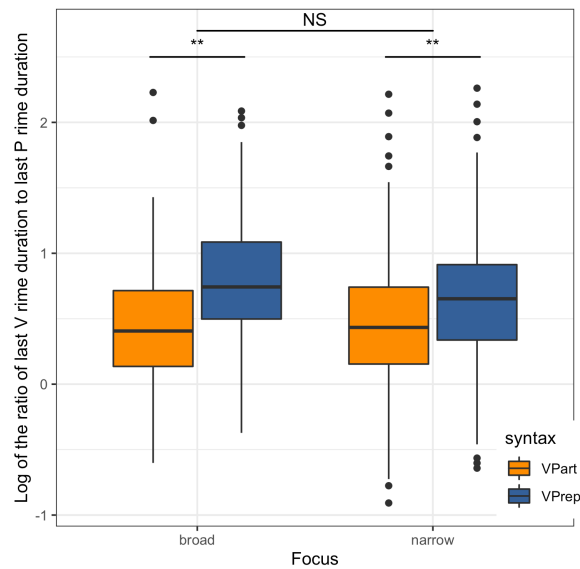


Figure 3: *Log of ratio of verb's second rime duration to P's second rime duration in broad focus and narrow focus.*

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

Concerning this prediction, my results are mixed, so it was difficult to determine whether phrasal prominence is preserved post-focally. I found no significant difference in the duration of post-focal V1 rimes between the V+Part and the V+Prep contexts (Figure 4). Nor did I find a significant difference in the duration of post-focal P1 rimes (Figure 5). 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.

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.

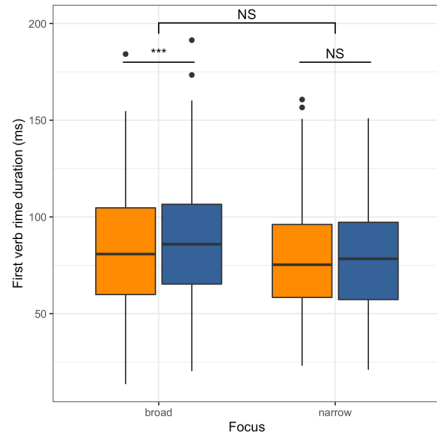


Figure 4: Duration of the verb's stressed rime in different syntactic structures in broad focus and narrow focus.

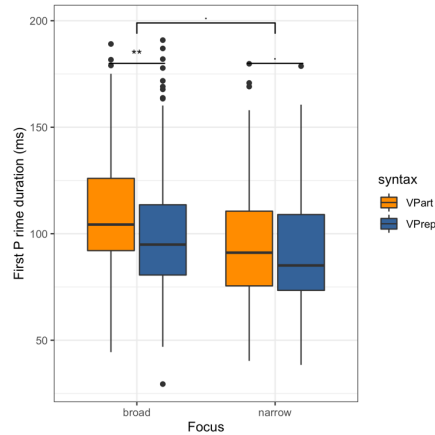


Figure 5: Duration of P's stressed rime in different syntactic structures in broad focus and narrow focus.

Furthermore, I 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 my impression that post-focal material is shorter in duration and more reduced compared to the same material in broad focus.

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.

5 Discussion

I first discuss the implications of the experimental results, and then evaluate some logically possible theories that lie behind them. 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 3). 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 to observe boundary tone phenomena, which usually indicates the phrase boundary. Thus, I 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 I infer what the post-focal phrase boundaries should be based on the phrase boundaries under broad focus.

Under broad focus, I 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, I 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 (i.e. the boundary following the verb in V+Part, and the boundary following the preposition) is a prosodic word boundary. Thus, I 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, I 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 I confirm this is true), then by inference the stronger boundary between V and P must be at least an iP.

Since post-focal contexts lose 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, I will evaluate a few logically possible theories, and will ground this discussion by putting the specific theories that I compare into metrical representations. I 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 that 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)	(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. Assuming 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 (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)	(x))	σ
Jane	will	be	moving		over		the	chairs.			
H*											

This representation is not compatible with my experimental findings, as I 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 my 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)	(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 my 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). I 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) |) | σ |
| Jane | will | be | moving | | over | | the | chairs. | | | |
| H* | | | | | | | | | | | |

Thus, there are two possible ways to accommodate my experimental findings, that is by abandoning either the headedness hypothesis, as in (13), or the prominence by pitch accent hypothesis, as in (14). These options 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, my experimental results are mixed in this respect, and do not distinguish between these alternatives.

Finally, I 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. A 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 I 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 no prominence by pitch accent requirement.

In order to understand Wagner’s theory, I first show how it would analyze a broad-focus sentence. The boundaries in Wagner’s unlabelled metrical grid are mapped from syntactic boundaries. I assume that 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] | <i>Verb + Particle</i> |
| b. | [moving] [PP over the chairs] | <i>Verb + Preposition</i> |

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) | (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) | (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. The following would be the representation of a V+Part sentence under narrow focus according to Wagner.

- (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) | (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? I think there are, but the evidence would come from the last iP head in the sentence (*chairs*), an element that I have not focused on in this paper, and therefore will be left for future research. (13) and (18) make different predictions about the prominence of the last iP head *chairs*. Wagner's theory (18) would predict that post-focally, *chairs* is acoustically equal to but perceptually stronger than the particle because Wagner assumes that the last stress in a series of equal stresses is perceived to be the strongest. In contrast, the traditional metrical grid representation in (13) would predict that post-focally, *chairs* is as prominent as 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 my 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 I leave this question to future research.

6 Conclusion

This paper reports a production study of English prosody in post-focal contexts. I presented results that suggest that iP boundaries are preserved post-focally, and that an iP may have no pitch accent. I also studied whether there are prominence distinctions in an accent-free context. My 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

- Beckman, Mary E. (1996). The parsing of prosody. *Language and Cognitive Processes*, 17–67.
- Beckman, Mary E., & Gayle Ayers Elam. (1997). Guidelines for ToBI labeling, version 3: Ohio State University.
- Bolinger, Dwight. (1958). A theory of pitch accent in English. *Word*, 14, 109–149.
- Boersma, Paul, & David Weenink. (2018). Praat: doing phonetics by computer [Computer program].
- Carlson, Katy, Clifton, Charles. Jr., & Lyn Frazier. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language*, 45, 58–81.
- Dimitrova, Snezhina, & Alice Turk. (2012). Patterns of accentual lengthening in English four-syllable words. *Journal of Phonetics*, 40, 403–418.
- de Jong, Kenneth. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics*, 32(4), 493–516.
- Ito, Junko, & Armin Mester. (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 Festschrift for Professor Shosuke Haraguchi on the Occasion of His Sixtieth Birthday*, 26–65. Tokyo: Kaitakusha.
- Ladd, Robert. (1980). *The Structure of Intonational Meaning. Evidence from English*. Bloomington: Indiana UP.
- Ladd, Robert. (1996). *Intonational Phonology*. Cambridge: Cambridge UP.
- Lieberman, Mark, & Janet B. Pierrehumbert. (1984). Intonational invariance under changes in pitch range and length. In Mark Aronoff and Richard Oerhle (Eds.), *Language Sound Structure*, 157–233. MIT Press.
- Matuschek, Hannes, Reinhold Kliegl, Shravan Vasishth, Harald Baayen, & Douglas Bates. (2017). Balancing type 1 error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315.
- Norcliffe, Elisabeth, & T. Florian Jaeger. (2005). Accent-free prosodic phrases? Accents and phrasing in the post-nuclear domain. *Proceedings of Interspeech*.
- Pierrehumbert, Janet B.. (1979). The perception of fundamental frequency declination. *The Journal of the Acoustical Society of America*, 66(2), 363–369.
- Pierrehumbert, Janet B.. (1980). The phonetics and phonology of English intonation. Ph.D. dissertation, MIT.
- Price, Patti J., Mari Ostendorf, Stefanie Shattuck-Hufnagel, & Cynthia Fong. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, 9, 2956–2970.
- Selkirk, Elisabeth. (1995). Sentence prosody: Intonation, stress, and phrasing. In J. Goldsmith (ed.), *The handbook of phonological theory*, 550–569. Cambridge: Blackwell Publishers.
- Shattuck-Hufnagel, S., & Alice E. Turk. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25(2), 193–247.
- Silverman, Kim, & Janet B. Pierrehumbert. (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.
- Turk, Alice. E., & Stefanie Shattuck-Hufnagel. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, 35, 445–472.
- Wagner, Michael. (2004). Prosody as diagonalization of syntax. Evidence from complex predicates. *Proceeding of NELS*, 34, 587–602.
- Wagner, Michael. (2005). Prosody and recursion. Ph.D. dissertation, MIT.
- Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf, & Patti J. Price. (1992). Segmental durations in the vicinity of prosodic boundaries. *Journal of the American Acoustic Association*, 91, 1707–1717.
- Xu, Yi, & Ching X. Xu. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, 33(2), 159–197.