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

Abstract

This paper studies the relationship between prosodic phrasing and prominence by addressing the questions of whether every prosodic phrase must have a head (a most prominent sub-constituent), and if so, what phonological features may mark the head of a phrase, and how those phonological features are realized phonetically. We study these questions by examining the intermediate phrase (iP) in English. If every iP must have a head, and this head must be marked by a pitch accent, then in an environment without any pitch accent, there should be no head/non-head distinction. And if there is no head, there should be no iP phrasing in this context either. We conducted a production study of strings that are ambiguous between verb + preposition and verb + particle structures such as *winning over*, and found durational evidence suggesting the presence of iP boundaries and prominence distinctions in an accent-less context. We also found evidence suggesting that accent-less iPs have phrase accents, and that focused material is lengthened, while post-focal material is shortened. Our results are compatible with at least two different hypotheses: either every iP has a head, but this head can be marked by phrasal stress; or prosodic phrases can be head-less.

Keywords: prominence, phrasing, pitch accent, duration, post-focal de-accenting, phrase accent, English

1. Introduction

At the sentence level prosodic structure involves two dimensions: prominence and phrasing. A standard hypothesis posits an intimate relationship between the two such that every phrase has a head (the most prominent sub-constituent). We call this *the headedness hypothesis*. Selkirk (1995) and Ito & Mester (2003), for example, proposed that headedness holds universally, and is an inviolable constraint. We test this hypothesis by studying the intermediate phrase (iP) in English in the sense of Beckman & Pierrehumbert (1986). If the headedness hypothesis holds, then every iP must have a head. If every iP has a head, then we may further ask what phonological features may mark the head of an iP, and how those phonological features are realized phonetically. We will test a strong hypothesis that posits that every iP has a head, and this head is made prominent by pitch accent, which is realized primarily or solely by pitch excursion (e.g. Beckman, 1996; Beckman & Edwards, 1994; Bolinger, 1958; Selkirk, 1984; section 1.1.1). Following Selkirk's (1984) terminology, we call this type of hypothesis *the accent-first hypothesis*. The accent-first hypothesis will be contrasted with two others–*the stress-first hypothesis*, which also follows the headedness hypothesis, but posits that the head of an iP is marked by phrasal stress rather than pitch accent (e.g. Beaver et al., 2007; Büring, 2013; Hayes, 1995; Ladd, 1996; section 1.1.2), and *the equal stresses hypothesis*, which abandons the headedness hypothesis and labels like iP, and allows a prosodic domain to have a series of equally prominent sub-constituents (Wagner, 2005, 2010; section 1.1.3). We will test these hypotheses in a context without any pitch accent–the region after narrow focus in English. These hypotheses make different predictions about phrasing and prominence in the post-focal region (section 1.2). In particular, the accent-first hypothesis predicts that post-focally, because there are no pitch accents, there should be no iP heads, and therefore also no

iP phrasing. In contrast, the other two hypotheses predict there to be iP phrasing and prominence distinctions post-focally. If we do find post-focal iPs as those two hypotheses predict, we further ask if those accent-less iPs may have a phrase accent, since Beckman & Pierrehumbert (1986) claimed that every iP ends with a phrase accent (section 1.3). We present a production study that searches for phonetic correlates of post-focal phrasing, prominence and additional phrase accents. While there have been prior studies on the two topics of post-focal phrasing and prominence separately (section 1.4), the current experiment studies both at the same time, using the subtle prosodic differences between two syntactic structures in the verb phrase region.

1.1.Phrasal prominence and phrasing

1.1.1. The accent-first hypothesis

This subsection presents three types of accent-first hypothesis, presenting the first one with the most detail. While they differed in how pitch accent is assigned, they all considered pitch accent to be the primary marker of phrasal prominence. The first type of accent-free hypothesis includes Beckman & Edwards (1994), Beckman (1996) and theories underlying ToBI transcription conventions (Beckman & Elam, 1997; Silverman & Pierrehumbert, 1987), which claimed that every iP has a head, and this head's prominence is *given* by a pitch accent. These theories implied that every iP must have at least one pitch accent.

As examples of these theories, we discuss Beckman & Edwards' (1994) and Beckman's (1996) proposal in detail. They assumed following Pierrehumbert (1980), Pierrehumbert & Hirschberg (1990), Selkirk (1980) and Vanderslice & Ladefoged (1972) a three-level prosodic structure consisting of what they called a "stress foot", an intermediate phrase (iP) and an intonation phrase (IP). Following Selkirk (1980), they further assumed that each level must have a head syllable, whose prominence is *given* by the head-marking phonological features at that level. Different levels have different head-marking, prominence-lending features. On the stress foot level, vowel quality is such a feature: a full vowel is more prominent than a reduced one. Full specification of vowel quality has inherent phonetic consequences in formant patterns, duration and loudness. On the iP level, pitch accent is a head-marking, prominence-lending feature: it has the structural property of making accented syllables more prominent than unaccented syllables. Beckman & Edwards claimed that the main phonetic realization of a pitch accent is pitch excursion; a pitch accent can be augmented by additional effects in duration and intensity, but they are not always required. This is similar to the position of Bolinger (1958), who found pitch to be the primary cue of accent, duration merely an ancillary one, and intensity having a negligible role.

Because a stress foot must have a head, whose prominence is marked by vowel quality, Beckman & Edwards predict that a stress foot should always have a full vowel. Likewise, because an iP must have a head, whose prominence is marked by pitch accentuation, they predict that an iP must have a pitch accent.

The two prosodic levels and their heads are interrelated. Beckman & Edwards followed the assumption that stress (i.e. the head of the stress foot) is the potential site for accentuation (i.e. a pitch accent can only associate with a stressed syllable) (e.g. Beckman, 1996; Beckman & Edwards, 1994; Bolinger, 1958; Pierrehumbert, 1980). If there is only one pitch accent in an iP, it must associate to the "most stressed" syllable; if there is more than one pitch accent in an iP, then the last pitch accent associates to the "most stressed" syllable, leading to the consequence that the last accented syllable is the prominent one (i.e. the nuclear accent).

We demonstrate the interrelation of stress and pitch accent with two examples, a one-word utterance and a multi-word one. Following Beckman (1996), we represent the stress and accentuation information in a metrical grid, a hierarchy of phonological elements. Syllables project to different levels on the metrical grid, and the higher their projection, the more prominent they are. The tier names are replicated from Beckman's grid: she assumed that stressed syllables project to the heavy tier, accented syllables project to the accented tier, and nuclear accented syllables further project to the uppermost tier. We add to the grid a final column indicating the phonetic realizations of prominence following Beckman & Edwards (1994).

First, Beckman & Edwards (1994) mentioned the verb *digést*, whose second syllable has primary stress, and the first one may or may not have secondary stress, which is marked by whether the first vowel is a full diphthong [aɪ] or a reduced vowel [ə]. Following would be Beckman's grid representation of the three possible ways to pronounce the verb *digést* as a full sentence: when pronounced as a full sentence, the nuclear pitch accent falls on the second syllable. Whether the first syllable also has a pitch accent depends on whether it has secondary stress, which is in turn marked by its vowel quality: if the first syllable has a full diphthong, this indicates that it has secondary stress and therefore potential to have pitch accent–it may or may not have a pitch accent, as are indicated by the left two representations in (1); but if the first syllable has a schwa, this indicates that it does not have stress, and therefore cannot have a pitch accent, either, as is indicated by the rightmost representation in (1).

(1)*Grid representations that Beckman & Edwards (1994) would assign to the citation pronunciations of the verb* digést*.*

						Tier	Phonetic realization of prominence
	x		x		x	nuclear accented	• pitch • optional: duration, intensity
x	x		x		x	accented	• pitch • optional: duration, intensity
x	x	x	x		x	heavy	• full specification of vowel quality
x	x	x	x	x	x	syllable	
dì	gést	dì	gést	di	gést		

Contrast the verb *digést* with the noun *dígèst*, whose first syllable has primary stress and the second syllable has secondary stress. If pronounced as a full sentence, we would expect the noun's first syllable to always carry pitch accent, and the second to do so optionally:

(2) *Grid representations of the citation pronunciations of the noun* dígèst.

				Tier	Phonetic realization of prominence
x		x		nuclear accented	• pitch • optional: duration, intensity
x	x	x		accented	• pitch • optional: duration, intensity
x	x	x	x	heavy	• full specification of vowel quality
x	x	x	x	syllable	
dí	gèst	dí	gèst		

Beckman (1996) also discussed sentences like *Anna came with Manny*. The empirical observation was that *Manny* has nuclear accent, and *Anna* may or may not have a pitch accent. The fact that the first syllable of *Anna* has a full vowel suggests that it is stressed, and thus has the potential for carrying the pitch accent. It may or may not actually be pitch accented, leading to the following two representations respectively:

(3) *Grid representations that Beckman (1996) assigned to* Anna came with Manny.

a.

						Tier	Phonetic realization of prominence
				x		nuclear accented	• pitch • optional: duration, intensity
x				x		accented	• pitch • optional: duration, intensity
x		x		x		heavy	• full specification of vowel quality
x	x	x	x	x	x	syllable	
A	nna	came	with	Ma	nny		

b.

						Tier	Phonetic realization of prominence
				x		nuclear accented	• pitch • optional: duration, intensity
				x		accented	• pitch • optional: duration, intensity
x		x		x		heavy	• full specification of vowel quality
x	x	x	x	x	x	syllable	
A	nna	came	with	Ma	nny		

Beckman & Edwards' and Beckman's proposal about the prosodic structure has two important implications about environments without pitch accents. First, the requirement that every iP must contain a pitch accent implies that in a context that has no pitch accent, we may expect there to be no iP boundaries.[1] Second, because pitch accent lends prominence to the accented syllable, they predict that pitch accented syllables are more prominent than unaccented but

[1] A reviewer asked about their analysis of tag questions like *Mary will win, won't she?*. While Pierrehumbert (1980) put the entire sentence in a single IP and did not posit an iP level, Beckman & Pierrehumbert (1986) revised this analysis to one with an iP boundary separating the main clause and the tag. They argued for the presence of a pitch accent in the tag, thus satisfying the requirement that every iP has a pitch accent.

stressed syllables, and unaccented but stressed syllables are equally prominent. This implies that if we can remove the pitch accent from a syllable by putting it in a context that disallows pitch accents, we expect it to be just as prominent as a normally unaccented but stressed syllable. For example, Beckman & Edwards made the explicit prediction that in an accent-less environment, the verb *dìgést* with the full diphthong [ˌdaɪˈdʒɛst] should not be categorically different from the noun *dígèst* [ˈdaɪˌdʒɛst], presumably because both syllables in these two words are unaccented but stressed:

(4) *Grid representations of the citation pronunciations of the verb* digést *[ˌdaɪˈdʒɛst] and the noun* dígest *[ˈdaɪˌdʒɛst] in an accent-less environment.*

				Tier	Phonetic realization of prominence
				nuclear accented	• pitch • optional: duration, intensity
				accented	• pitch • optional: duration, intensity
x	x	x	x	heavy	• full specification of vowel quality
x	x	x	x	syllable	
dì	gést	dí	gèst		

It is worth mentioning that the first type of the accent-first theories does not make any prediction about where an iP phrase may occur, other than the prediction that there is no iP phrasing in an accent-less environment. Their definition of iP phrases is based on the surface phonetic form of an utterance, but they do not predict iP phrasing based on the syntactic structure of a sentence.

Other than this first type of the accent-first theories, there were others that did not assume the relationship between pitch accent and stress in the same way, but nevertheless took pitch accent to be the primary marker of phrasal prominence. We mention two of them here without going into much detail: Selkirk (1984) and Bolinger (1958). Like Beckman & Edwards and Beckman, Selkirk considered pitch accent to constitute and define phrasal prominence, and to be the "prior" and more essential marker of prominence than metrical stress. But unlike these works, Selkirk considered pitch accent assignment to be strictly separate from stress assignment, and to be based on purely syntactic-semantic constituents called "sense groups" instead. Pitch accent may in turn partially affect metrical stress assignment, causing the metrical grid to adjust so that pitch accents fall on maximally stressed syllables. Like Selkirk, Bolinger (1958) considered pitch accent to mark phrasal prominence, and also separated pitch accent assignment from the metrical grid: he considered pitch accent assignment to be guided by information-structural properties such as newness, contrast and informativity.

1.1.2. The stress-first hypothesis

The stress discussed in Beckman & Edwards (1994) and Beckman (1996) in the previous subsection was word stress, which occurs on a lower level than phrase. Selkirk (1984), Hayes (1995), Ladd (1996), Truckenbrodt (1995, 2006), Beaver et al. (2007), Büring (2013) and others posited *phrase-level stress*, a distinct marker of phrasal prominence that is different from pitch accent. All these theories except Selkirk put stress before accent, and argued that phrasal prominence is marked by stress; because there is often correlation of stress and pitch accent, it leads to the illusion that prominence is marked by pitch accent. But in environments where we

can tease the two apart like accent-less environments, these theories claimed that we should still see prominence distinctions caused by stress.

These theories all allow for phrasal stress, but differ in detail, such as whether and how pitch accent associates with stress, how to derive the prosodic domains from syntax, and whether it is every iP or every lexical XP that requires a phrasal stress. But these differences do not matter for our purposes. What matters is that they all considered phrasal prominence to be marked by stress rather than pitch accent.

As an example theory, we briefly introduce Hayes' formulation of metrical theory of phrasal stress, and then discuss how it interacts with pitch accent. After that, we will discuss Beaver et al.'s (which was based on Selkirk and Hayes) predictions about phrasing and prominence in accent-less environments.

According to Hayes, phrasal stress establishes relative prominence relations between members of a phrase, just like how word-level stresses does so for members of a word. He proposed rules to assign phrasal stress, such as the Nuclear Stress Rule that makes the rightmost member of a phrase the strongest in English. When applied recursively to nested syntactic structures, Nuclear Stress Rule derives multiple levels of stress. This leads to a direct correlation of stress and syntactic depth of embedding. But the actual correlation is not always so straightforward, which Hayes claimed was due to precedence of eurhythmic factors, focus-related factors and general principles such as one that prefers not to stress verbs. Therefore, we will only present Hayes, who discussed as examples the English hyphenated name *Farrah Fawcett-Majors* and sentences like *John saw Mary*. He observed that the hyphenated name puts the most prominence on *Majors*; while *Farrah* and *Fawcett* both contain stress, the stress on *Farrah* is stronger than that on *Fawcett*. Likewise, there are prominence distinctions in sentences with a nested syntactic structure. Assuming a right-branching syntactic structure for Anna [VP came [PP with Manny]] and additional constraints such as verbs' tendency to resist stress, we expect Hayes to represent this sentence in the following way with an unlabelled grid:

(5) *Grid representations that Hayes (1995) would assign to* Anna came with Manny.

				x	
x				x	
x		x	x	x	
A	nna	came	with	Ma	nny

Below we add to this grid the labels and the 'syllable' tier following Beckman. Hayes (1995) and Beaver et al. (2007), which built on Hayes, did not make any explicit proposal of how to assign pitch accents to the metrical grid, except for Hayes' remark that pitch accents be associated with the strongest stress. Here we suggest a way to incorporate pitch accent and nuclear accent to Hayes' metrical grid that is consistent with Hayes and Beaver et al.: we add the 'accented' and 'nuclear accented' tiers on top of Hayes' stress grid, and assume that syllables that have projected to the topmost phrasal stress tier must carry pitch accents, which project them further to the 'accented' tier. The final accent in a phrase is then projected to the 'nuclear accented' tier. The syllables that have projected to the lower phrasal stress tiers may carry pitch accents. Following would be a possible grid representation of *Anna came with Manny*. Beaver et al. (2007) suggested that phonetically, pitch accent is only realized by pitch, while phrasal stress is realized by greater duration, intensity and more extreme formant frequencies, leading to better differentiated vowel quality. We incorporate these assumptions in the rightmost column of the grid:

(6) *Grid representations that Hayes (1995) and Beaver et al. (2007) would assign to the citation form of* Anna came with Manny *with labels.*

						Tier	Phonetic realization of prominence
				x		nuclear accented	• pitch
				x		accented	• pitch
				x		phrasal stress	• duration, intensity, vowel quality
x				x		phrasal stress	• duration, intensity, vowel quality
x		x	x	x		word stress	• full specification of vowel quality
x	x	x	x	x	x	syllable	
A	nna	came	with	Ma	nny		

Beaver et al. predicted that we should still be able to find prominence distinctions due to phrasal stress in an accent-less environment, and showed that this was borne out by second-occurrence focus. Beaver et al. predict that putting a sentence like *Anna came with Manny* in an accent-less environment would remove the pitch accents on these words, but keep the phrasal stresses and word stresses, as the following grid indicates. Specifically, the first syllable of *Manny* should still be more prominent than that of *Anna*, which is realized not by pitch, but by duration, intensity and vowel quality.

(7) *Grid representations that Hayes (1995) and Beaver et al. (2007) would assign to the citation form of* Farrah Fawcett-Majors *in an accent-less environment.*

						Tier	Phonetic realization of prominence
						nuclear accented	• pitch
						accented	• pitch
				x		phrasal stress	• duration, intensity, vowel quality
x				x		phrasal stress	• duration, intensity, vowel quality
x		x	x	x		word stress	• full specification of vowel quality
x	x	x	x	x	x	syllable	
A	nna	came	with	Ma	nny		

Beaver et al. (2007) suggested that pitch accent is realized only by pitch excursion, while phrasal stress is realized by duration, intensity and vowel quality. But we can imagine a slightly different hypothesis, where pitch accent may be realized not only by pitch, but also by duration, intensity and vowel quality. Then this version of the hypothesis may predict that in an accent-less environment, the first syllable of *Manny* is still more prominent than that of *Anna*, but this unaccented first syllable of *Manny* is less prominent than the accented syllable of *Manny* in a normal environment in terms of duration, intensity and vowel quality.

Beaver et al. (2007) did not discuss the headedness hypothesis, therefore technically they were mute regarding the correlation of iP phrasing and pitch accent or phrasal stress. But other stress-first hypotheses followed the headedness hypothesis explicitly. For example, Büring (2013) assumed that prosodic constituents must have heads, and they are marked by *stress*. Ladd (1996) also said explicitly that every iP must have a phrasal stress. Let us thus assume that the stress-first approaches take a different position from the accent-first approaches: every iP must have a head, but this head is marked by phrasal stress rather than pitch accent. This implies that in a context without pitch accent, we may still find iP phrasing, as long as there is still phrasal stress. Therefore, these stress-first hypotheses predict that in an accent-less context, we may still find phonetic correlates of phrasal stress and iP phrasing.

1.1.3. The equal stresses hypothesis

Wagner (2005, 2010) proposed a totally different approach from the accent-first and the stress-first hypotheses, with three key differences. First, Wagner got rid of the labels of the grid such as iP or IP. Since there was no concept of iP or IP, there was no requirement that the head of an iP must be marked by pitch accent, either. Second, Wagner put pitch accent in 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); weaker prosodic phrases don't carry any pitch accent. Finally, Wagner abandoned the headedness hypothesis, and allowed a phrase to have a series of equal stresses. While he allowed for phrasal stress–syllables that project to a phrasal level would carry phrasal stress–he did not consider there to be a nuclear stress acoustically: there can be multiple syllables that project to the same level in the grid, and acoustically they are equally prominent. We thus call this *the equal stresses hypothesis*.

We demonstrate how Wagner would analyze a simple sentence like *Anna came with Manny*. The boundaries in Wagner's unlabelled metrical grid are mapped from syntax, specifically: more deeply embedded syntactic constituents are separated from each other by weaker boundaries in the grid than less deeply embedded constituents. Assuming the right-branching syntactic structure for this sentence, below would be its prosodic representation in Wagner's unlabelled grid. Because *Manny* is syntactically more embedded than *Anna*, the boundaries around *Manny* (i.e. the boundary between *with* and *Manny*) are weaker than the boundaries around *Anna* (i.e. the boundary between *Anna* and *came*). Wagner (2005) also suggested that monosyllabic function words such as *with* generally subordinate to their arguments in the grid and resist accentuation, while transitive verbs may or may not subordinate to their arguments. In the grid below we assume that both *with* and *came* subordinate to their arguments.

(8) *Grid representations that Wagner (2005, 2010) would assign to* Anna came with Manny.

						Phonetic realization of prominence
(x				x	)	• pitch accent, duration, intensity, pitch
(x	)	(		x	)	• duration, intensity, pitch
(x	)	(x)	(	x	)	• duration, intensity, pitch
(x	x)	(x)	(x)	(x	x)	• duration, intensity, pitch
A	nna	came	with	Ma	nny	

Note that there are three syllables that are equally strong on the highest level. Wagner claimed that acoustically they are equally prominent (e.g. they should have the same degree of lengthening, and the last syllable, *ma*, may even have lower pitch and intensity than the

preceding syllables due to declination, downstep and intensity down-drift), and the reason why the last stress (the stress on *Manny*) *sounds* the strongest is due to perception. Following Newman (1946) and citing Pierrehumbert's (1979) and Liberman & Pierrehumbert's (1984) findings, Wagner claimed that the last stress in a series of equal stresses is perceived to be the strongest. This perceptual principle creates the impression that *Manny* has the nuclear stress of the sentence, despite the fact that acoustically it is as prominent as the previous stresses.

What Wagner would predict to happen to phrasing and prominence in an accent-less context depends on his analysis of such contexts. The next subsection introduces one such context and what his analysis would be.

1.2.Focus prominence and phrasing

The accent-first and the stress-first hypotheses make concrete predictions about accent-less contexts: while the former hypothesis predicts that there should be no iP phrasing and prominence distinctions in these contexts, the latter predicts that there should still be iP phrasing and prominence distinctions in duration and other phonetic measures. These predictions need to be tested in a context that has no pitch accent. In an English sentence the area following the nuclear accent (we call this *the post-focal region*) is precisely such a context. To demonstrate this, we will first examine a broad-focus sentence, and then compare its prosody with that of a sentence with narrow focus. Then we will discuss the predictions by the accent-first and the stress-first hypotheses. After that, we will also review Wagner's (2005, 2010) analysis of focus, which differs in crucial ways from the two hypotheses. This leads to Wagner's unique predictions about prosody of the post-focal region.

In English, broad focus is generally realized by multiple accents in a sentence, with nuclear accent falling on the last content word. 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 answering the question in (9A), the entire sentence in (9B) is focused (semantic focus is marked by the subscript F), and thus has broad focus (following the views of e.g. Gussenhoven, 1983; Ladd, 1980; Schmerling, 1976; Selkirk, 1984). The context in (9) leads to the verb + preposition interpretation of *winning over*, which has been previously shown to have a particular intonational melody (Norcliffe & Jaeger, 2005; Price et al., 1991): the subject, the verb and the object tend to carry pitch accents, which fall on their primarily stressed syllables; the prosodic boundary between the verb and the preposition tends to be larger than the boundary following the preposition. Price et al. (1991) found in a production study that 60% of the time, the boundary between the verb and the preposition was an IP in the sense of Pierrehumbert (1980) and Beckman & Pierrehumbert (1986) because a boundary tone occurred there. Figure 1 is the waveform of our recording of a speaker's production of (9B). The pitch perturbations are caused by obstruents and glottalization.

(9)*Broad focus*
Context: Heartless violence led to a bloody victory.
A: What was happening in the story?
B: [The Vikings were winning over their enemies]$_F$.

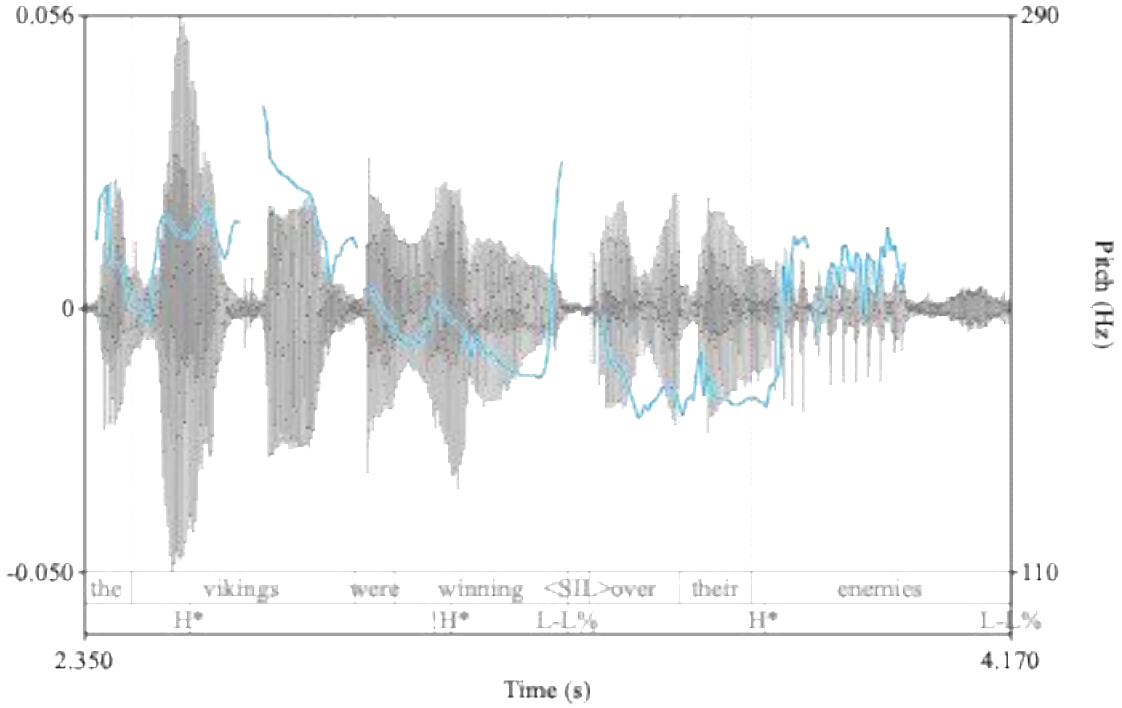

Figure 1: *Waveform of a Verb + Preposition sentence in broad focus.*

In contrast, in a sentence with narrow focus, the focused word has nuclear accent, and all the following words are *de-accented*, including those that normally may have a pitch accent in broad focus, a phenomenon known as *post-focal de-accenting* (e.g., Beckman, 1996; de Jong, 2004; Ladd, 1980, 1996; Xu & Xu, 2005). Compare the prosody of (9B) with that of a sentence containing narrow focus. The question in (10A) triggers narrow focus on the subject, *the Vikings*, in (10B). The narrow focus is realized by the nuclear accent on *the Vikings*, the strongest pitch accent in the sentence. Specifically, it falls on the primarily stressed syllable, the first syllable of *Vikings*. Furthermore, there is no pitch accent on any word following *Vikings*. Typically, this means that F0 falls sharply after *Vikings*, and remains low and level through the rest of the sentence. Figure 2 is the waveform of our recording of the same speaker's production of (10B).

(10) *Narrow focus*
Context: Heartless violence led to a bloody victory.
A: Who was winning over their enemies?
B: [The Vikings]$_F$ were winning over their enemies.

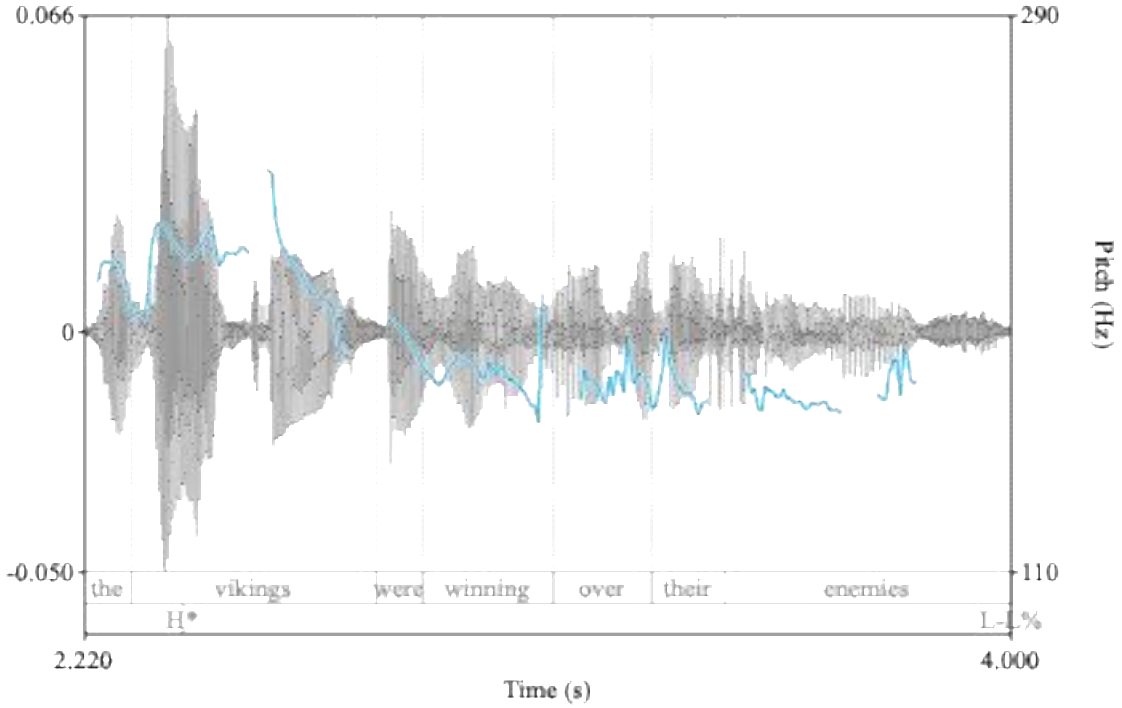

Figure 2: *Waveform of a Verb + Preposition sentence in narrow focus.*

The common analysis of post-focal de-accenting was that focus must be marked by a nuclear accent (e.g. Beaver et al., 2007; Beckman, 1996; Büring, 2013), and that nuclear accent is the last pitch accent in a phrase (e.g. Cohan, 2000; Ladd, 1996). When the entire sentence is focused, as in (9B), we do not observe post-focal de-accenting 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 *enemies*. But when the subject is narrowly focused, as in (10B), no word following it may be accented to ensure that the subject has the final pitch accent in the phrase.

1.2.1. The accent-first hypothesis

One way to analyze post-focal de-accenting in the metrical grid is removal of post-focal pitch accents (Beaver et al., 2007; Beckman, 1996). Here we discuss the first type of accent-first hypothesis' predictions about post-focal prominence, but the other types make the same predictions. Based on the accentuation facts, following would be Beckman's grid representations of the broad-focus sentence (9B) and the narrow-focus sentence (10B) respectively:

(11) *Grid representations that Beckman (1996) would assign to the broad-focus sentence (9B).*

									x			nuclear accented
	x			x					x			accented
	x	x		x		x		x	x			heavy
x	x	x	x	x	x	x	x	x	x	x	x	syllable
the	Vi	kings	were	win	ning	o	ver	their	e	ne	mies	

(12) *Grid representations that Beckman (1996) would assign to the narrow-focus sentence (10B).*

	x											nuclear accented
	x											accented
	x	x		x		x		x	x			heavy
x	x	x	x	x	x	x	x	x	x	x	x	syllable
the	Vi	kings	were	win	ning	o	ver	their	e	ne	mies	

As explained in section 1.1.1, the first type of the accent-first hypothesis only requires every iP to have a pitch accent, but does not have a predictive theory of where iP phrasing should occur in a broad-focus sentence like (9B). Whether it is able to predict iP phrasing there does not actually matter to us. All that matters is that the iP phrasing after the verb in the broad-focus sentence (9B) (if there is any) is predicted to disappear in the narrow-focus sentence (10B). This is because the absence of pitch accent after narrow focus implies that there is no iP phrasing post-focally (i.e. *post-focal de-phrasing*). Thus, among (13B-B'''), the logically possible ways of phrasing (10B), (13B) is the only one that is permitted by this hypothesis. (13B') has an accent-less iP (*were winning over their enemies*); and so do (13B'') (*their enemies*) and (13B''') (*were winning over* and *their enemies*).

(13) *Narrow focus*
A: Who was winning over their enemies?
B: (The Vikings were winning over their enemies).
B': (The Vikings) (were winning over their enemies).
B'': (The Vikings were winning over) (their enemies).
B''': (The Vikings) (were winning over) (their enemies).
H*

Besides post-focal de-phrasing, this hypothesis makes another prediction about post-focal prominence distinctions. Under broad focus, the verb *winning* is more prominent than the preposition *over* because the former is accented but the latter is not, as we can see from (11). In contrast, with narrow focus on the subject, the verb and the particle should be equally prominent because both are stressed but unaccented, as we can see from (12).

1.2.2. The stress-first hypothesis

The stress-first hypothesis also analyzed post-focal de-accenting as removal of pitch accents in the post-focal region. Following would be this hypothesis' grid representations of the broad-focus sentence (9B) and the narrow-focus sentence (10B) respectively. Based on the empirical

fact that *Vikings*, *winning* and *enemies* have pitch accents, it would assign phrasal stress to these words because they have to carry phrasal stress in order to have the potential for accentuation. Then with narrow focus on the subject, only the pitch accents on *winning* and *enemies* are removed, but they retain their phrasal stresses.

(14) *Grid representations that Beaver et al. (2007) would assign to the broad-focus sentence (9B).*

									x			nuclear accented
	x			x					x			accented
	x			x					x			phrasal stress
	x	x		x		x		x	x			word stress
x	x	x	x	x	x	x	x	x	x	x	x	syllable
the	Vi	kings	were	win	ning	o	ver	their	e	ne	mies	

(15) *Grid representations that Beaver et al. (2007) would assign to the narrow-focus sentence (10B).*

	x											nuclear accented
	x											accented
	x			x					x			phrasal stress
	x	x		x		x		x	x			word stress
x	x	x	x	x	x	x	x	x	x	x	x	syllable
the	Vi	kings	were	win	ning	o	ver	their	e	ne	mies	

If Beaver et al. (2007) require every iP to have a head, and this head to be marked by phrasal stress instead of pitch accent, then we do not expect post-focal de-phrasing just because there is no pitch accent, but we do expect to find post-focal prominence distinctions due to phrasal stress (i.e., post-focally, we expect the first syllable of *winning* to be more prominent than that of *over* in terms of duration, intensity and vowel quality). If pitch accent is only realized by pitch accent and phrasal stress is realized durationally as Beaver et al. suggested, then we expect the post-focal prominence distinctions realized in duration to be the same as those in broad focus; but if pitch accent is also realized durationally, then we expect the post-focal prominence distinctions to be smaller than those in broad focus because of the absence of pitch accent on *winning*.

Just like with the accent-first hypotheses, it does not actually matter where the stress-first hypotheses predict iP boundaries should occur in the broad-focus sentence.[2] All that matters is

[2] Some stress-first theories do make predictions about iP phrasing in broad focus (e.g. Büring, 2013), but Büring (2013) actually fails to predict there to be an iP boundary following the verb in (9B), contrary to the empirical observations. Specifically, Büring followed Selkirk's (1986) edge based approach and aligned the right edge of every NP and VP with a right iP boundary. Let us assume a right-branching syntactic structure for (9B): [The Vikings] [$_{TP}$ were [$_{VP}$ winning [$_{PP}$ over [$_{DP}$ their enemies]]]]. Büring would not put an iP boundary between *winning*

that it predicts that if there is any iP boundary after the subject in broad focus, then those boundaries should still be present post-focally.

1.2.3. The equal stresses hypothesis

In contrast to these proposals that remove the accents in the post-focal region, Wagner (2005, 2010) promoted the narrowly-focused constituent to another level in the grid, and subordinated the unfocused material to the focus, keeping the prosodic structure of the unfocused material intact. We assume that (9B) has a right-branching syntactic structure: [The Vikings] [$_{TP}$ were [$_{VP}$ winning [$_{PP}$ over [$_{DP}$ their enemies]]]]. Following Wagner's (2005, 2010) claim that the metrical grid replicates the syntactic embeddings and his assumption that function words can subordinate to their following arguments, following would be Wagner's grid representation of the broad-focus sentence (9B) that is consistent with its actual intonational melody. According to Wagner, the primarily stressed syllables of *Vikings*, *winning* and *enemies* have pitch accent because they project to the highest level in the grid, and *enemies* sounds like it has nuclear accent because of a perceptual principle that highlights the last stress. The prosodic boundary between *winning* and *over* is larger than the boundary following *over*, which matches the syntactic structure.

(16) *Grid representations that Wagner (2005, 2010) would assign to the broad-focus sentence (9B).*

(	x			x					x		)
(	x	)	(	x					x		)
(	x	)	(x)	(x					x		)
(	x	)	(x)	(x	)	(			x		)
(	x	)	(x)	(x	)	(x	)	(	x		)
(x	x	x)	(x)	(x	x)	(x	x)	(x)	(x	x	x)
the	Vi	kings	were	win	ning	o	ver	their	e	ne	mies

When the subject has narrow focus as in (10B), Wagner (2005, 2010) proposed adding a level on the top to the grid of the broad-focus sentence to ensure that the subject projects to a higher level than the other constituents in the sentence. This allows the first syllable of *Vikings* to be the final stress on the topmost level, making it perceptually the strongest stress in the sentence. Thus, unlike the other proposals that remove the accents of post-focal material, Wagner adds a level to which only the narrowly-focused word projects, and the unfocused elements are subordinated, so that phrasing and headedness on the lower levels remain intact. Wagner would thus predict that post-focally, there are still iP phrasing and prominence distinctions just like there are in broad focus.

---

and *over* because there is no right NP or VP edge between them. We leave to future research what sorts of mapping theories can or cannot account for the empirical observations of an iP boundary in (9B).

(17) *Grid representations that Wagner (2005, 2010) would assign to the narrow-focus sentence (10B).*

(	x										)
(	x			x					x		)
(	x	)	(	x					x		)
(	x	)	(x	(x					x		)
(	x	)	(x	(x	)	(			x		)
(	x	)	(x	(x	)	(x	)	(	x		)
(x	x	x)	(x	(x	x)	(x	x)	(x)	(x	x	x)
the	Vi	kings	were	win	ning	o	ver	their	e	ne	mies

To summarize, so far we have presented three different hypotheses and their different predictions about post-focal phrasing and prominence. Because there is no pitch accent post-focally, the accent-first hypothesis predicts there to be no iP phrasing and no prominence distinctions post-focally, either. The stress-first hypothesis and the equal stresses hypothesis both predict there to still be iP phrasing and prominence distinctions in phrasal stress, which are realized acoustically in durations.

While the stress-first hypothesis and the equal stresses hypothesis make the same predictions about post-focal phrasing and prominence, there is a subtle difference in their predictions about the duration of the primarily stressed syllable of the subject compared to that of the verb. Beaver et al. (2007), proponents of the stress-first hypothesis, suggested that pitch accent may be realized phonetically only by pitch, while phrasal stress is realized by other effects such as duration. In contrast, the equal stresses hypothesis considers there to be prominence effects in duration on every level of the grid. Beaver et al. would thus predict that under broad focus, the durational effect due to phrasal stress on the subject should equal that durational effect on the other phrasally stressed words (e.g. the verb). Since under narrow focus, only pitch accents are removed, but the phrasal stresses on these words remain the same, the durational effects due to phrasal stress on these words should be the same as under broad focus. In contrast, the equal stresses hypothesis predicts that under broad focus, the subject and the verb are equally prominent acoustically. But under narrow focus, the subject is promoted to a higher level than the other words, and therefore should be more prominent, and this is realized durationally. Therefore, Beaver et al. predict that the relative prominence of the subject compared to the verb should be the same for broad focus and narrow focus, but the equal stresses hypothesis predicts that this relative prominence should be greater for narrow focus than for broad focus.

1.3.Whether post-focal iPs may end with a phrase accent

If we do durational evidence for iP phrasing post-focally, we may further ask if those post-focal iPs may contain a phrase accent. The common assumption that every iP must end with a phrase accent (e.g. Beckman & Pierrehumbert, 1986) leads us to expect that they do.

To test this question, we need to search for additional phrase accents in the post-focal area. We can do so in a systematic way because as it turns out, we cannot find those phrase accents in the utterances produced by a speaker that always produces L-. Instead, we would need a speaker who sometimes produces L- and other times H-. The rest of this subsection explains this.

If a speaker always produces L- phrase accents in broad-focus sentences like the pattern in Figure 1, then we may infer that if they were to put a phrase accent under narrow focus, that phrase accent is an L- as well. Just based on the intonational melody of their narrow-focus

sentences, we cannot tell if their post-focal iPs have a phrase accent because whether or not it has a phrase accent, we will see the same realization in pitch movement. Consider the narrow-focus sentence (10B) again as an example. If an accent-less iP can end with a phrase accent, and there is an iP boundary immediately following the verb, then these speakers may be able to produce an L- at this iP boundary (18b). But if an accent-less iP cannot contain any phrase accent, then (18b) is not possible, but only (18a) is. Let us also assume that narrow-focus sentences generally put an H* on the subject and end with an L%.

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

	The Vikings	were	winning	over	their	enemies.
a.	H*					L-L%
b.	H*		L-			L-L%

Assume that an L- phrase accent spreads its tone, filling the space between the last pitch accent and the phrasal boundary (Beckman & Pierrehumbert, 1986; Pierrehumbert, 1980). If the F0 target for the two L- phrase accents is the same in (18b), the tunes in (18a&b) should have the same pattern in pitch movement.[3] For both tunes, we should observe that the pitch falls shortly after the H*, and remains low and steady until the end of the sentence.

Thus, to test the question of whether a post-focal iP may end with a phrase accent, we would need to examine speakers whose phrase accent is not always L- because they will create more varied pitch movements than speakers who only produces L-, thus creating more opportunities to identify a sentence-medial phrase accent. Suppose there is such a speaker who in broad focus sometimes produces L- and sometimes H-. We infer that if this speaker were to produce a phrase accent post-focally, it would sometimes be H- and sometimes be L-, making it a great place to search for the phrase accent associated with a post-focal iP. If this speaker could produce the following tune, then it would be evidence that accent-less iPs can have a phrase accent:

(19) *Key evidence for phrase accent in an accent-less iP in a narrow-focus V+Part sentence.*

The Vikings	were	winning	over	their	enemies.
H*		L-			H-L%

There is no theory that we know of that predicts the tune of two consecutive phrase accents that are not separated by any pitch accent, which in this case are L- H-. Following the assumption that phrase accents spread leftward, and fill the interval between the last pitch accent and the final boundary tone (Beckman & Pierrehumbert, 1986; Pierrehumbert, 1980), we could imagine that the pitch may stay low due to the first L- for some time, followed by a rise to H-. The tune in (19) may thus be realized in the following way: the pitch falls sharply after the H* and stays low for some time before it rises back up to an H-, and stays nearly flat in a middle range or falls slightly (L%) at the end of the sentence. This tune crucially differs from the tunes that lack a phrase accent in the middle, such as H* H-L%, H* L-L% and H* L-

[3] If the two L- phrase accents in (18b) have different F0 targets, (18b) may actually be realized differently from (18a). Presumably the two L- phrase accents in (18b) would both spread. There is no existing theory for what happens when a space is filled with two L-s. Suppose the F0 target of the second L- is lower than the first, then perhaps the pitch falls shortly after H* to the F0 level of the first L-, remains low for some time before it falls sharply again 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 (18b), and the first pitch drop may differ in slope from the second drop.

H%. H* H-L% is realized as a steady pitch at the high level after the H*; H* L-L% is realized as a pitch fall shortly after the H*, which then stays low and steady; H* L-H% is realized as a pitch fall shortly after the H*, after which the pitch stays steadily low at the L- before it rises at the end of the sentence to an H%. In contrast to H* H-L% and H* L-L% where after the H* the pitch stays steady at either the high or low level, the tune in (19) requires a fall from the H* and then a rise back up. H* L-H% also has a fall and then a rise, but this rise occurs much later at the end of the sentence, compared to the tune in (19), where this rise occurs much earlier.

In sum, if we do find post-focal iPs, we may further test whether they may end with a phrase accent. This would require speakers who sometimes produce H- and sometimes L-. If such speakers can produce the unique tune in (19) in narrow focus, then post-focal iPs have phrase accents.

1.4. Previous studies of post-focal phrasing, prominence and additional phrase accents

Having presented the three hypotheses and their predictions about post-focal phrasing and prominence, we will now discuss previous research that examined post-focal phrasing (in English and other languages) and prominence experimentally, and how the current study differs from them.

Studies on post-focal phrasing in English included Norcliffe & Jaeger (2005) and Wagner & McAuliffe (2019), who did find evidence of post-focal phrasing. However, neither of these studies tested post-focal prominence distinctions and discussed the question of whether those accent-less iPs are still headed.

Furthermore, the experiment to be presented in this paper also differs from these previous works in experimental design. Previous works in the literature that studied prosodic phrasing often focused on coordinated structures. Wagner & McAuliffe continued this tradition by studying phrasing in coordinations in sentences like *But we were told that Megan, and Lauren or Morgan would help,* and varied focus by putting broad focus on the entire coordination, or narrow focus on one of the three names. These coordinations tend to involve significant prosodic boundaries, and there is also evidence that those boundaries may be boosted even more by the comma in text (Cooper & Paccia-Cooper, 1980; Price et al., 1991). In contrast, our experiment examines the lesser-studied syntactic structures of verb + preposition (V+Prep) and verb + particle (V+Part) by taking advantage of the lexical and structural ambiguities of strings like *winning over*. Compared to coordinations, Price et al. (1991) found that V+Prep and V+Part constructions do contrast in boundary strengths, but they tend to involve more subtle prosodic boundaries that ranged from "a minor grouping of words", a boundary slightly smaller than an iP in Beckman & Pierrehumbert (1986), to an IP boundary.

Norcliffe & Jaeger (2005) did study precisely post-focal phrasing effects in the V+Prep and V+Part constructions, which partly inspired this research. But the current study differs from theirs in four important ways. First, Norcliffe & Jaeger did not control for the syllable and stress structure of the words in their stimuli, but instead used a mix of monosyllabic and disyllabic verbs and particles. Using monosyllabic words like *won* might confound two separate effects–lengthening due to prominence and lengthening due to occupying a phrase-final position. For this reason, the current study only uses disyllabic trochaic words. Second, Norcliffe & Jaeger only tested narrow-focus sentences, but not broad-focus sentences, but the current study examines both. We begin with a systematic study of the V+Prep and V+Part sentences under broad focus, and ask if their different syntactic structures lead to different prosodic realizations. Then we ask if the prosodic differences we see in the broad-focus sentences remain in the narrow-focus sentences. Third, Norcliffe & Jaeger's experiment had a sample size of 40, while the current one has a larger size of 2040. Finally, Norcliffe & Jaeger's

experiment did not have fillers (E. Norcliffe, p.c.), but the current study does, which hopefully helps mask the syntactic differences between the V+Prep and V+Part sentences for the participants.

Besides these studies in English, there were studies on post-focal phrasing in other languages, which also found that phrasing is not completely removed post-focally (e.g. Féry, 2011; Hayes & Lahiri, 1991; Ishihara, 2003, 2007, 2016; Jun, 2011; Jun & Fougeron, 2000; Kügler & 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 phonologically and phonetically.

Besides studies on post-focal phrasing, there were also studies on post-focal prominence, though none of them examined post-focal phrasing at the same time. Most of the studies on post-focal prominence focused on word-level stress, and showed that there are word-level stress distinctions in an accent-less environment (e.g. Halliday (1967), Horne (1993) and Okobi (2006) for English, and Sluijter & Van Heuven (1996) for Dutch).

There has been less work on phrasal stress in an accent-less context, with Beaver et al. (2007) being one of the few on this topic. But Beaver et al. may have mixed up post-focal prominence effects with post-focal phrasing effects. They found post-focal prominence distinctions in sentences with two foci such as *Even [the state prosecutor]$_F$ only named Sid$_F$ in court today.* This sentence was produced after the sentence *But the defendant only named Sid$_F$ in court today*. This makes *the state prosecutor* in the target *even*-sentence the first-occurrence focus that associates with *even*, and *Sid* in the post-focal region the second-occurrence focus that associates with *only*. Beaver et al. found that *Sid*, despite being post-focal, was still slightly more prominent than the other post-focal but non-focused material such as *court*. But Beaver et al. did not control for the syllable and stress structure of the words in their stimuli, and all their targets words were monosyllabic such as *Sid* and *court*. If there is post-focal phrasing immediately following *Sid* (and there may well be), then what they found may not be phonetic effects due to *Sid*'s prominence, but phonetic effects due to pre-boundary lengthening effects created by this boundary following *Sid*.[4] Our experiment keeps these two effects strictly apart with carefully-controlled stimuli, and thus tests post-focal phrasing and post-focal prominence separately.

To our knowledge whether accent-less iPs may have a phrase accent is a question that has not been asked before in the literature. We suspect this is because even the question whether accent-less iPs exist has not been quite settled in the first place. But we hope that the current study as well as previous works on English and other languages suggest that they do exist. Also, as was explained in section 1.3, even if this question may have been on some researchers' minds, to test it would require quite unusual intonational melodies involving a mix of L- and H- in narrow focus.

The next section presents the production study, which takes advantage of the prosodic differences created by different syntactic structures. Section 3 presents the results. Section 4 discusses the theoretical implications of the results. Section 5 concludes the paper.

[4] We are grateful to a reviewer for pointing out this potential issue of Beaver et al. (2007) to us.

2. The experiment

2.1.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 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.

Prosody can disambiguate many types of syntactic distinctions (e.g. Carlson, 2009; Cooper & Paccia-Cooper, 1980; Kraljic & Brennan, 2005; Lehiste, 1973; Price et al., 1991; Schafer et al., 2000). The syntactic distinction that we identified involves the verb phrase domain: it was observed that, when a verb is followed by a word that is ambiguous between a particle and a preposition, it has different syntactic structures, which correspond to different prosodies (Norcliffe & Jaeger, 2005; Price et al., 1991). Take the string *won over* as an example: when interpreted as persuaded, it has the verb + particle (V+Part) structure; when interpreted as won by violence, it has the verb + preposition (V+Prep) structure:

(20) *V+Part*
Gentle persuasion led to a friendly settlement. The men won over their enemies.
(21) *V+Prep*
Heartless violence led to a bloody victory. The men won over their enemies.

In a production study, Price et al. (1991) found that these syntactic differences led to different prosodic realizations. Following are their prosodic representations of the typical realizations of these two structures, using their ways of marking. They assigned break indices from 0 to 6: 0 is boundary within a clitic group; 1 is normal word boundary; 2 is boundary marking a minor grouping of words; 3 is equivalent to iP boundary in Beckman & Pierrehumbert (1986); 4 is IP boundary; 5 is boundary marking a grouping of IPs; and 6 is sentence boundary. They distinguished among four levels of prominence: no prominence at all, minor phrasal prominence (P0), major phrasal prominence (P1) and nuclear stress (FF). Major phrasal prominence may be accompanied by what they called *contrastive stress*.

(22) *V+Part*

the	0	men	3	won	0	over	2	their	1	enemies	6
		P0				P1				FF	

(23) *V+Prep*

the	0	men	2	won	4	over	1	their	1	enemies	6
		P0		P1-Contrastive						FF	

More generally, they found that prosodic boundary played a major role in distinguishing between these two structures, while prominence played a supportive role. Among all the productions of V+Part and V+Prep structures like (20)–(21), the break index between the verb and the particle was always smaller than the one following the particle, and conversely, most of the time the break index between the verb and the preposition was larger than the one

following the preposition. This break index was quite large: for example, 60% of the time, an IP boundary immediately followed the verb in V+Prep. Furthermore, 80% of the particles had a prominence, whereas only 5% of the prepositions did.

The phrasing and prominence distinctions observed by Price et al. (1991) should be reflected phonetically in duration. 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 & Turk, 1996; Wightman et al., 1992). In particular, Wightman et al. (1992) 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.[5] The bigger this break is, the longer the rime. Thus, based on Price et al.'s (1991) observation that the prosodic boundary between the verb and the particle was smaller than the one following the particle, we expect the final rime of the particle to be lengthened more than the final rime of the verb; conversely, based on their observation of the reverse pattern in V+Prep, we expect the final rime of the verb to be lengthened more than the final rime of the preposition. If we can compare the same rime across syntactic structures, then we may expect the final rime of the verb to be longer in V+Prep than in V+Part, and the final rime of the particle to be longer than that of the preposition.

Dimitrova & Turk (2012) found that the rime of an accented syllable is longer than the rime of an unaccented syllable. Price et al.'s (1991) observation that the particle rather than the verb tends to be pitch accented in V+Part, and conversely the verb rather than the preposition tends to be pitch accented in V+Prep leads to the prediction that the primarily stressed rime of the verb should be longer in V+Prep than in V+Part, and conversely, the primarily stressed rime of the particle should be longer than that of the preposition.

Thus, there are two independent predictions about lengthening effects–lengthening due to prosodic phrasing and lengthening due to prominence. The former effect primarily falls on the final rime of the word before the boundary, and thus requires comparing the final rimes of the verb and the particle / preposition; the latter effect primarily falls on the primarily stressed rime of the accented word, and thus requires comparing the primarily stressed rimes of the verb and the particle / preposition.

To tease apart these two effects and predictions, we thus need verbs and particles / prepositions with disyllabic, trochaic structure, so that the first/stressed rime reflects lengthening effects due to prominence (mostly, see footnote 5), and the second/unstressed rime reflects pre-boundary lengthening effects. This makes verbs like *won* not ideal to test these predictions, and thus in our experiment we changed *won* to *winning*. Only two disyllabic, trochaic words have the lexical ambiguity of a particle and a preposition–*over* and *under*, thus our materials are quite limited for this reason. We call the first rimes of the verb and the particle / preposition V1 and P1 respectively, and their second rimes V2 and P2 respectively. We also call the verb V, and the word that is ambiguous between a particle and a preposition P.

It is worth mentioning that Norcliffe & Jaeger (2005) did test the lengthening effects, but only those effects in narrow focus, and they used a mix of monosyllabic and disyllabic words including *won*. Thus, when they measured the duration of the monosyllabic verb and compared its duration between syntactic conditions, they mixed up the lengthening effects due to prominence and phrase-finality.

---

[5] Turk & 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 will discuss shortly, we take the duration of the first rime to mostly reflect prominence.

After comparing the durations of these rimes under broad focus, which will hopefully confirm Price et al.'s (1991) observations about the prosodic differences between V+Part and V+Prep, we can then address the main research question in this paper–whether these durational differences due to phrasing and prominence that we find under broad focus still persist in the post-focal context. We can test this question by putting narrow focus on the subject of the sentence, thus putting the verb phrase in the post-focal region.

If distinctions in phrasing are preserved post-focally, we expect the relative strength of phrase boundaries to be preserved as well. We thus expect that post-focally, V2 should be longer in V+Prep than in V+Part, and P2 should be longer in V+Part than in V+Prep. Assuming that many boundaries we find in broad focus are iP boundaries, and that there is no pitch accent post-focally, this would suggest that there may be accent-less iPs, contrary to the predictions of the accent-first hypothesis, but consistent with the stress-first hypothesis and the equal stresses hypothesis. But if V2 and P2's durations do not differ between the syntactic structures post-focally, then this would support the accent-first hypothesis, but not the other two hypotheses.

If prominence distinctions are preserved post-focally, then elements that are lengthened due to phrasal prominence in broad focus should still be lengthened post-focally. We thus expect that post-focally, 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. This would suggest that there is still phrasal prominence in the absence of pitch accent, contrary to the predictions of the accent-first hypothesis, but consistent with the stress-first hypothesis and the equal stresses hypothesis. But if V1 and P1's durations do not differ between the syntactic structures post-focally, then this would suggest that there is no phrasal prominence, supporting the accent-first hypothesis, but not the other two hypotheses.

If those 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. This would be consistent with the versions of the stress-first hypothesis and the equal stresses hypothesis that consider pitch accent to be realized only by pitch, but not by duration. If the prominence distinctions are smaller post-focally than under broad focus, then this suggests that pitch accent is realized by pitch and duration, and de-accentuation leads to less prominence, which may be detected durationally.

To tease apart the stress-first hypothesis and the equal stresses hypothesis, we can examine the prominence of the subject in broad focus and narrow focus, and compare it with the prominence of V and P in those contexts. The stress-first hypothesis predicts that the subject's prominence relative to V and P should not differ by focus, whereas the equal stresses hypothesis predicts that the subject should more prominent than V and P in narrow focus than in broad focus. We compare their prominences by comparing the durations of their primarily stressed rimes.

2.2.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. (24)–(27) are examples of a set of four conditions. 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. (24) and (26) elicit V+Part and V+Prep in broad focus in the answer respectively. (25) and (27) elicit V+Part and V+Prep in narrow focus respectively.

(24) Context: Gentle persuasion led to a friendly settlement.
Question: What was happening in the story?
Answer: The Vikings were winning over their enemies.

(25) Context: Gentle persuasion led to a friendly settlement.
Question: Who was winning over their enemies?
Answer: The Vikings were winning over their enemies.

(26) Context: Heartless violence led to a bloody victory.
Question: What was happening in the story?
Answer: The Vikings were winning over their enemies.

(27) Context: Heartless violence led to a bloody victory.
Question: Who was winning over their enemies?
Answer: The Vikings were winning over their enemies.

Examples (24)–(27) are an example set whose target sentences have identical surface strings. In the other sets where this is not possible, we tried to make sure that the stress patterns and morphosyntactic structures of the region up to P were as similar as possible, and made sure that the total number of syllables up to P was the same.

The speakers/participants in our experiment were instructed to play the three different roles (narrator, question-asker, question-answerer) by saying the context, the question, and the answer sentences in the given order. We had 20 sets of experimental items, making it a total of 80 target sentences. Every speaker saw all 80 items. We conducted the experiment in two events, the first for 13 participants, and the second for the other 15 participants. All the participants saw the same test items, and the first event had 73 filler items, while the second had 44 filler items. They were all items for other experiments.

2.3.Participants

We conducted a production study with 28 native speakers of North American English (20 female, 8 male). Our subjects were all university students and working professionals living in X and Y Cities. 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 filler items and the diverse form of items helped conceal the purpose of the experiment.

2.4.Data collection

Recording took place in two events. The first event took place in a sound-attenuated booth at X University for 13 of the 28 participants, and the second event took place in a quiet, non-reverberant room at Y University for the other 15 participants. In each event, 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 a pseudo-randomized order, minimal pairs were not placed next to each other, and the order of items was different for every participant. 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 could take as long as they wanted. 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.

2.5.Data annotation

Two research assistants listened to the first batch of recordings of the 13 participants, and the authors listened to the second batch of recordings of the other 15 participants, annotated the pitch accents on V and P, and marked the sentences that needed to be excluded from the analysis (the exclusion criterion to be explained in the next subsection). 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).

The authors transcribed the break indices of the broad-focus sentences in the second batch of recordings following ToBI transcription conventions (Beckman & Elam, 1997; Silverman & Pierrehumbert, 1987): break index 0 is clitic boundary, 1 normal word boundary, 2 mismatch of subjective boundary strength and phrase accent, 3 iP boundary, and 4 IP boundary. All the participants produced L- phrase accents in broad-focus sentences, except one, who sometimes produced H- phrase accents. We transcribed the boundary tones of this speaker's narrow-focus sentences.

All the recordings were aligned with the Montreal Forced Aligner (McAuliffe et al., 2017) using the pretrained acoustic model English (US) ARPA acoustic model (McAuliffe & Sonderegger, 2024), and duration was measured with the forced-aligned boundaries.

2.6.Data exclusion

Out of the total number of recorded sentences (N=2240), we excluded 131 disfluent sentences (5.8%), out of which 68 sentences had broad focus and 63 sentences had narrow focus. This includes all the 80 sentences produced by one speaker, who constantly produced disfluent speech by putting a pause between every pair of words.

Because we were interested in what happens in an accent-less region, we made sure there were no pitch accents after narrow focus. We excluded 128 sentences with post-focal pitch accent (12.1% of the total number of fluent narrow-focus sentences).

2.7.Data analysis

We fitted 6 linear mixed-effects models and calculated p values using Satterthwaite's degrees of freedom method (using the R package 'lmerTest'). 4 of these models had the durations of V1, V2, P1 and P2 as the respective dependent variables, and syntax (V+Part vs. V+Prep, contrast-coded), focus (broad vs. narrow, dummy-coded), and their interaction as the fixed effects. The other 2 models had as their respective dependent variables the duration of the primarily stressed rime of the subject, and that duration divided by the average duration of V1 and P1. These 2 models had focus (broad vs. narrow, dummy-coded) as the fixed effect.

For the second batch of recordings for which we transcribed the break indices, we fitted two ordinal logistic regression models, with the break indices of the boundary between V and P and

the boundary following P as the respective dependent variables, and syntax (V+Part vs. V+Prep, dummy-coded) as the fixed effect. Following Matuschek et al. (2017), all the models included random intercepts and slopes by speaker and item group where those effects made a significant contribution to model fit ($p < 0.2$).

3. Results

3.1.ToBI transcriptions of the broad-focus sentences

Most of the broad-focus sentences had the prosody that is consistent with Price et al.'s (1991) descriptions. Out of the 1,052 fluent broad-focus sentences, 58 sentences (5.0%) had the opposite accentuation patterns of their descriptions: pitch accent on the verb but not the particle in V+Part, or pitch accent on the preposition but not the verb in V+Prep; 5 sentences (0.5%) did not put pitch accent on V or P. 313 sentences (30.0%) put pitch accent on both V and P, out of which 228 sentences (21.7%) were V+Part. But we should note that double accentuation was difficult to annotate because generally, sequences of pre-nuclear accents are notoriously difficult (Ladd, 1996). We thus ran statistical analyses both including and excluding broad-focus sentences with double pitch accent and V and P, and report them in the next subsection.

Among the 989 broad-focus sentences that had the accentuation patterns that were consistent with Price et al.'s descriptions, only 34 sentences had the opposite phrasing patterns from Price et al.'s descriptions (3.4%) by putting a stronger prosodic boundary between V and Part than after Part, or putting a stronger prosodic boundary after Prep than between V and Prep. The odds of having a larger break index between V and P is 11.84 higher in V+Prep than in V+Part ($p<0.001$) (Figure 3). The odds of having a larger break index following P is 13.83 higher in V+Part than in V+Prep ($p<0.001$) (Figure 4). Our transcriptions of the break indices followed ToBI conventions, and we often found a phrase accent at the boundary following the particle, and at the boundary between the verb and the preposition.

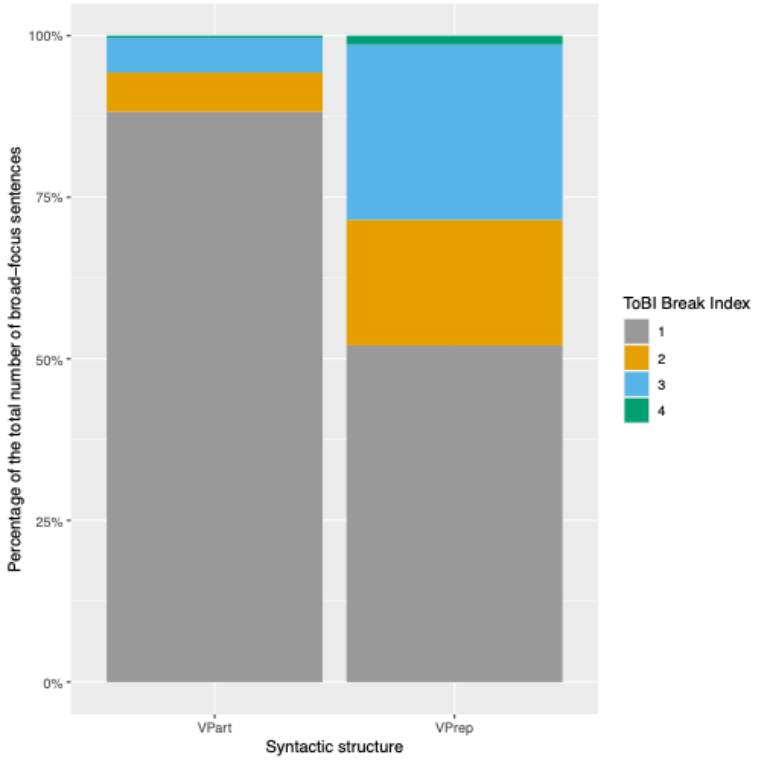

Figure 3: *ToBI break index of the boundary between V and P in different syntactic structures in broad focus.*

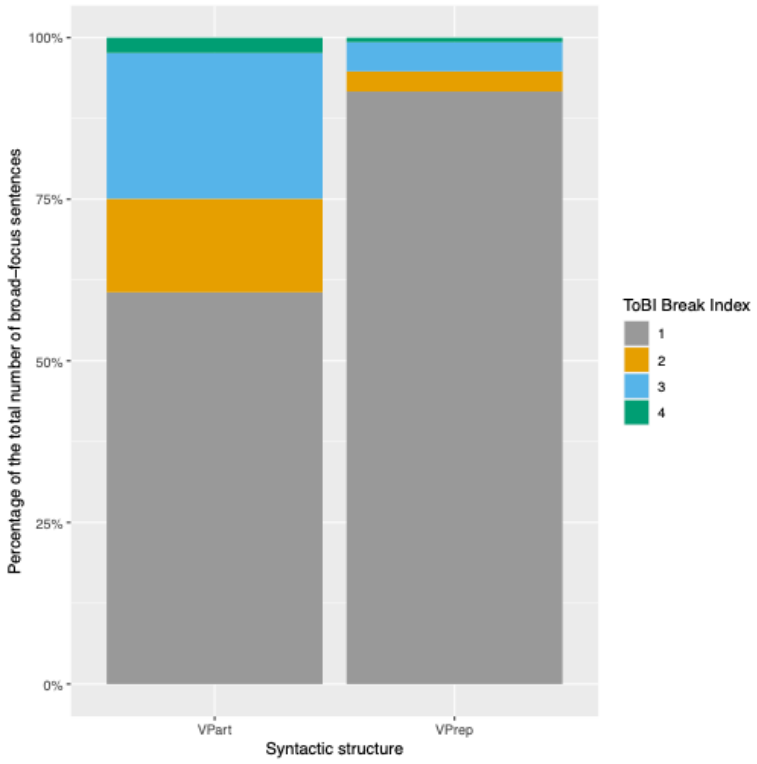

Figure 4: *ToBI break index of the boundary following P in different syntactic structures in broad focus.*

When the speakers produced phrase accents in broad-focus sentences, they were all L-, except one speaker, who sometimes produced H- phrase accents. Following is a particularly involved example of this speaker's production of a V+Prep sentence:

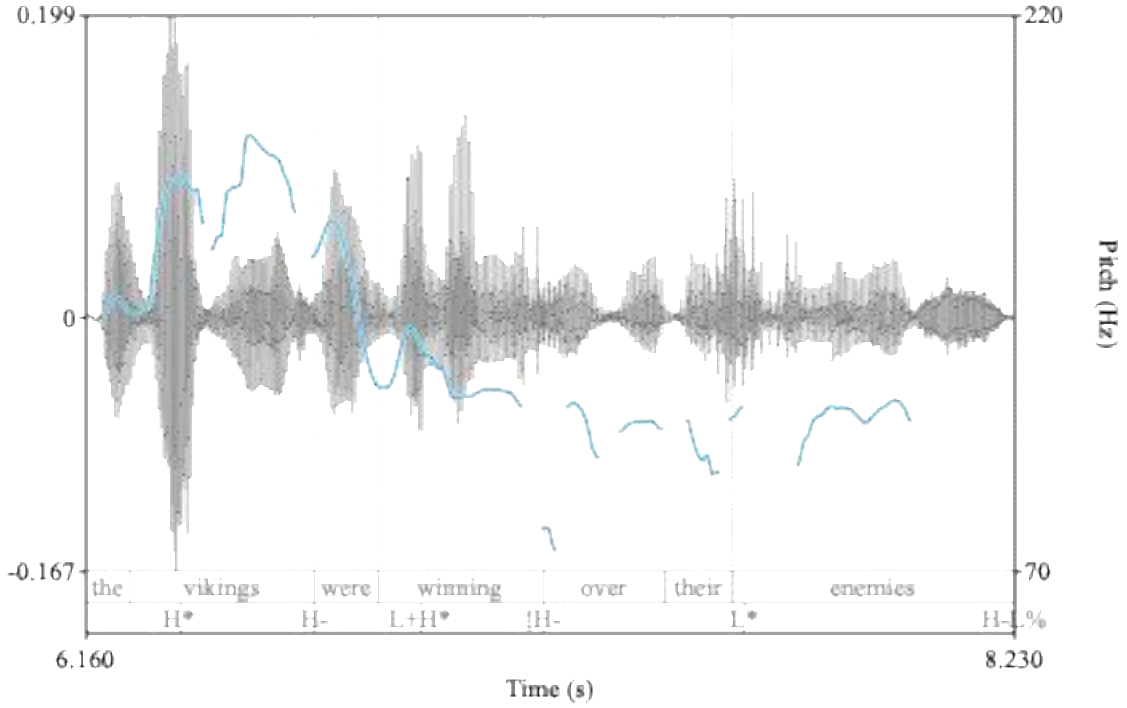

Figure 5: *Waveform of a Verb+Prep sentence with H- in broad focus.*

Among the 40 narrow-focus sentences produced by this speaker, 5 had post-focal pitch accents and were therefore excluded, and the other 35 all had the tune H* L-L% as in (19). Following is an example: we think that the sharp drop of F0 after H* suggests the presence of an L-, and the subsequent steady rise in F0 is caused by the sentence-final H-.

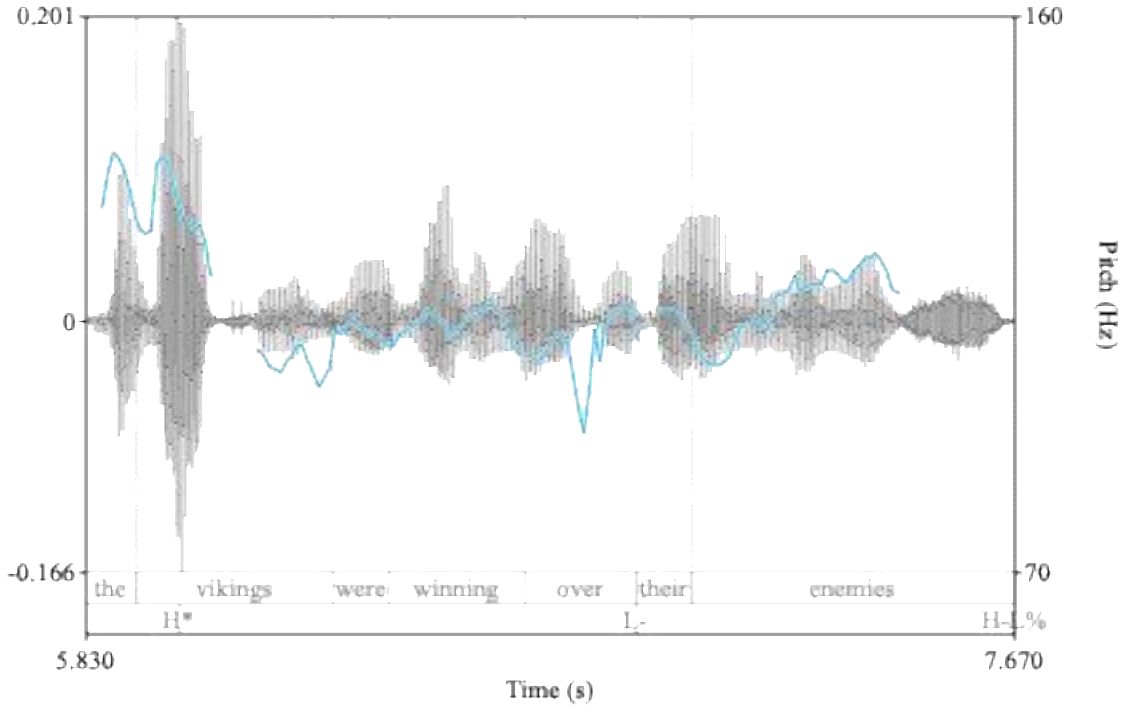

Figure 6: *Waveform of a Verb+Part sentence with the tune H* L- H-L% in narrow focus.*

3.2.Duration

Under broad focus, V1 is on average 5.21 ms longer in V+Prep than in V+Part ($p<0.001$) (Figure 7); P1 is on average 8.00 ms longer in V+Part than in V+Prep ($p<0.001$) (Figure 8); V2 is on average 7.28 ms longer in V+Prep compared to V+Part ($p<0.001$) (Figure 9); P2 is on average 13.26 ms longer in V+Part compared to V+Prep ($p<0.001$) (Figure 10).

If we do not consider the broad-focus sentences that put pitch accent on both V and P, and those with the opposite accentuation patterns from Price et al.'s (1991) descriptions (i.e. pitch accent on V but not P in V+Part and pitch accent on P but not V in V+Prep), then V1 is on average 7.13 ms longer in V+Prep compared to V+Part ($p<0.001$); P1 is on average 9.27 ms longer in V+Part than in V+Prep ($p<0.001$).

Post-focally, V1 is on average 3.36 ms longer in V+Prep than in V+Part ($p<0.01$) (Figure 8); P1 is on average 5.00 ms longer in V+Part than in V+Prep ($p<0.001$) (Figure 8); V2 is on average 4.35 ms longer in V+Prep than in V+Part ($p<0.01$) (Figure 9); P2 is on average 9.92 ms longer in V+Part than in V+Prep ($p<0.001$) (Figure 10). There is no difference between differences: the difference in V1's duration between V+Prep and V+Part post-focally does not differ significantly from that difference under broad focus. The same goes for the difference in differences for V2's, P1's and P2's durations.

Furthermore, we found that focus significantly reduced the duration of most rimes: V1 is on average 5.39 ms shorter post-focally compared to the broad focus context ($p<0.001$); P1 is on average 10.34 ms shorter post-focally than in broad focus ($p<0.001$); V2 is on average 5.61 ms shorter post-focally than in broad focus ($p<0.001$); P2 is on average 5.01 ms shorter post-focally than in broad focus ($p<0.001$). 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.

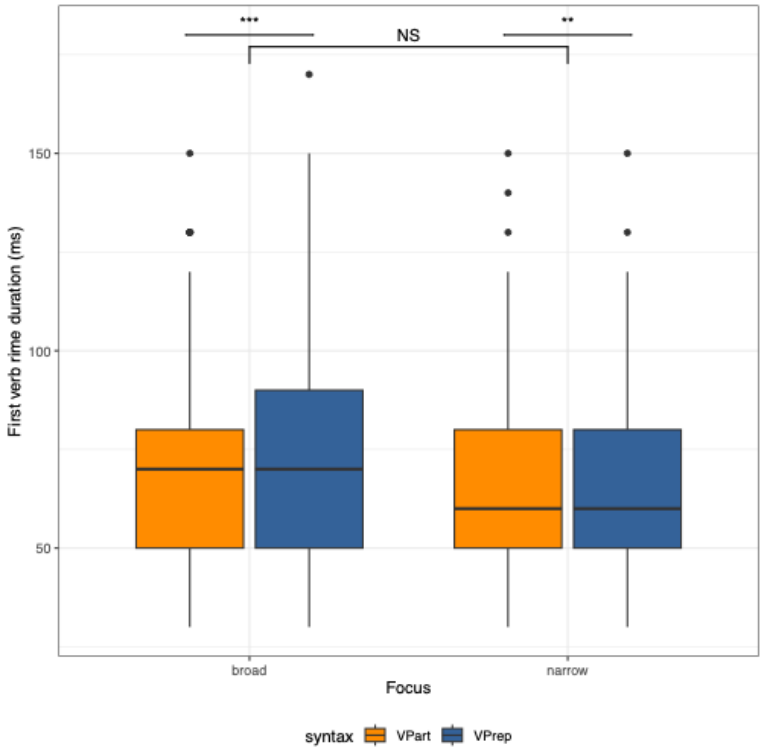

Figure 7: *Duration of the verb's stressed rime in different syntactic structures in broad focus and narrow focus.*

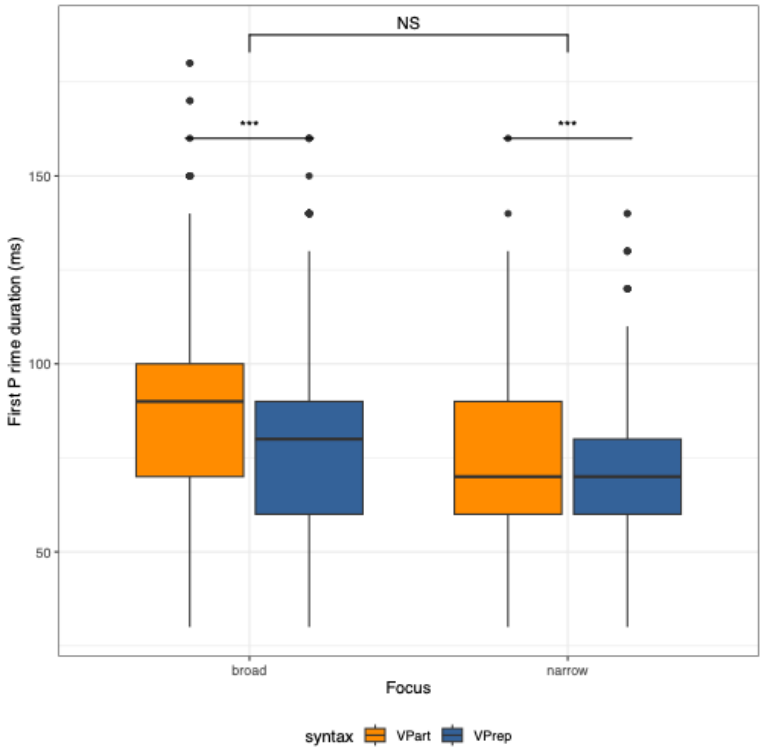

Figure 8: *Duration of P's stressed rime in different syntactic structures in broad focus and narrow focus.*

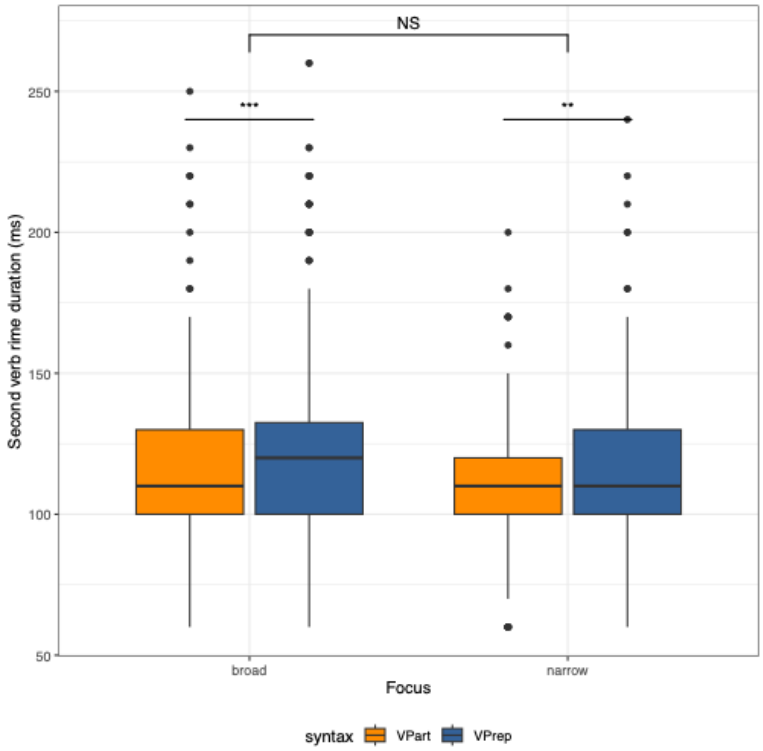

Figure 9: *Duration of the verb's final rime in different syntactic structures in broad focus and narrow focus.*

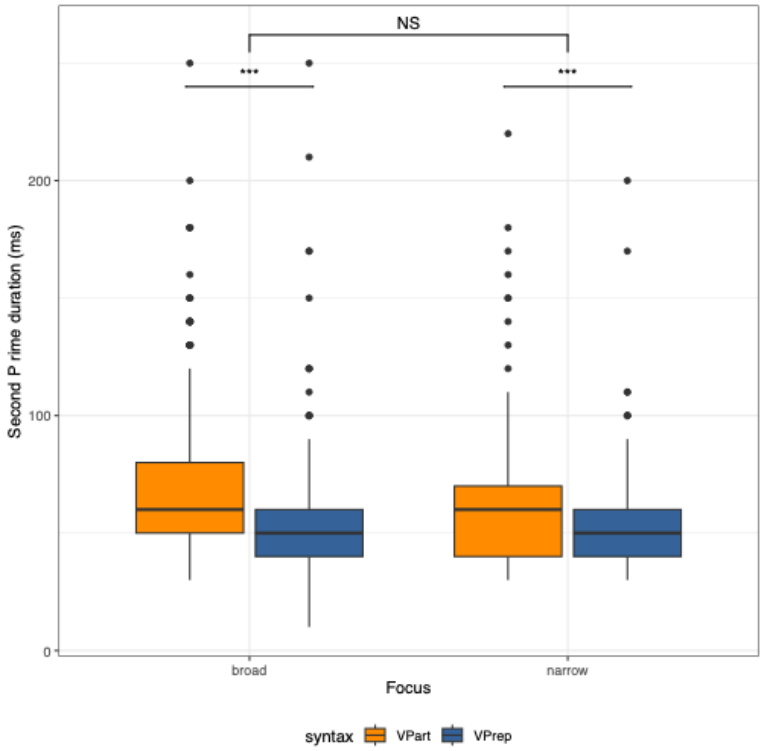

Figure 10: *Duration of P's final rime in different syntactic structures in broad focus and narrow focus.*

The primarily stressed rime of the subject is on average 6.47 ms longer in narrow focus than in broad focus ($p<0.001$) (Figure 11). The ratio of this rime duration to the average duration of V1 and P1 is greater in narrow focus than in broad focus ($p<0.001$) (Figure 12).

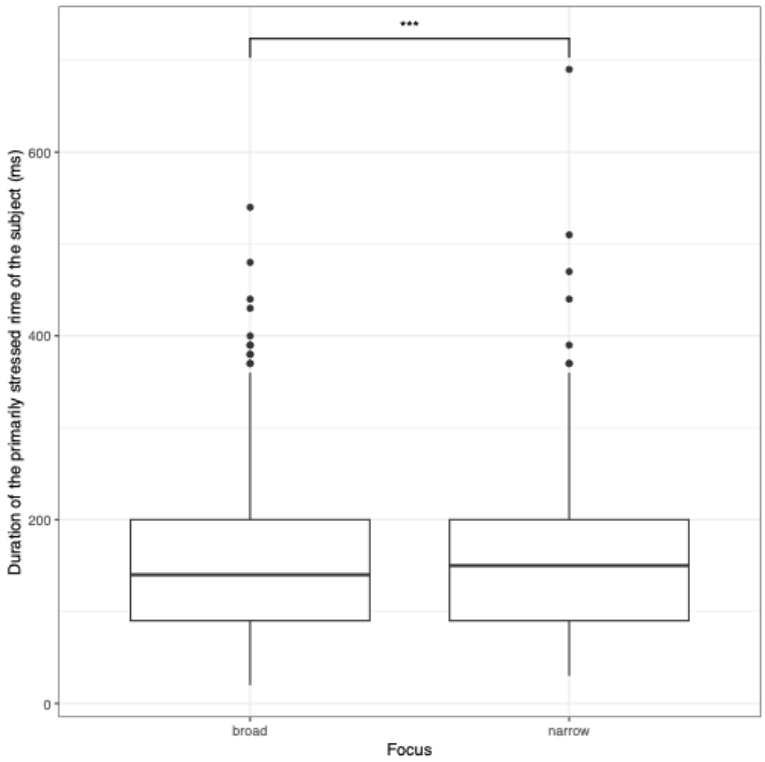

Figure 11: *Duration of the primarily stressed rime of the subject in broad focus and narrow focus.*

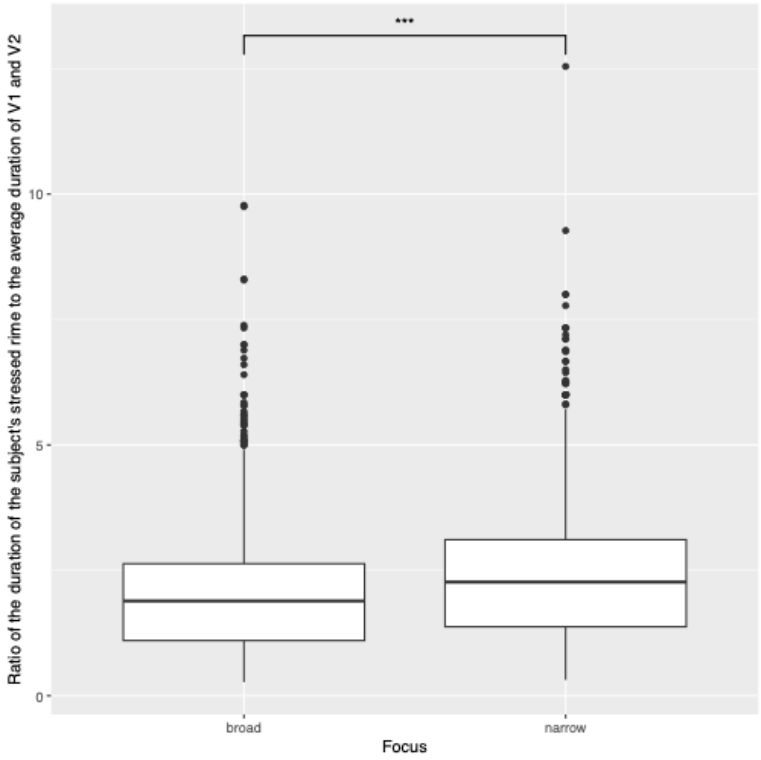

Figure 12: *Ratio of the duration of the subject's primarily stressed rime to the average duration of V1 and V2 in broad focus and narrow focus.*

4. Discussion

Our transcriptions and durational results of the broad-focus sentences were consistent with Price et al. (1991), who reported that prosodic boundary played a major role in distinguishing between V+Part and V+Prep structures, and prominence a supportive role. Under broad focus, prosodic phrasing reflects syntactic differences: V+Part is more likely to put a larger prosodic boundary following P than V+Prep, and V+Prep is more likely to put a larger prosodic boundary between V and P than V+Part. Assuming that a larger prosodic phrase triggers more phrase-final lengthening on the final rime before the boundary, the durational results suggest that under broad focus, the prosodic boundary between V and P is larger in V+Prep than in V+Part, and the prosodic boundary following P is larger in V+Part than in V+Prep.

Our transcriptions found that while the particle but not the preposition tends to be accented, the verb may be accented in both V+Prep and V+Part, suggesting that P's accentuation patterns were a more reliable cue to the syntactic differences between V+Part and V+Prep than V's accentuation patterns. This is consistent with Price et al.'s observations, who remarked that the particle was more likely to be prominent than the preposition, but did not discuss the stress patterns of the verb. Assuming that more prominent rimes tend to be longer, the durational results suggest that under broad focus, V is more prominent in V+Prep than in V+Part, and P is more prominent in V+Part than in V+Prep.

Having confirmed that the syntactic differences between V+Part and V+Prep do lead to prosodic differences in phrasing and prominence in broad focus, we found these differences

persist in the post-focal context. Post-focally, the final rime of V is lengthened more in V+Prep than in V+Part, and the final rime of P is lengthened more in V+Part than in V+Prep, suggesting that the prosodic boundary between V and P is larger in V+Prep than in V+Part, and the prosodic boundary following P is larger in V+Part than in V+Prep. These durational differences do not differ significantly from those in broad focus, suggesting that the strength of the post-focal prosodic boundaries is comparable to that of the prosodic boundaries in broad focus.

Post-focally, the stressed rime of V is lengthened more in V+Prep than in V+Part, and the stressed rime of P is lengthened more in V+Part than in V+Prep, suggesting that V is more prominent in V+Prep than in V+Part, and P is more prominent in V+Part than in V+Prep. These durational differences do not differ significantly from those in broad focus, suggesting that the post-focal region has prominence distinctions that are comparable to those in broad focus. This supports Beaver et al.'s (2007) suggestion that pitch accent may be realized only by pitch, and phrasal stress may be realized by duration.

Since we made sure that the post-focal region had no pitch accent, the fact that we found phrasing in the post-focal region suggests that there can be accent-less phrases. There are reasons to think that those accent-less phrases are larger than a prosodic word. Price et al. (1991) observed that 60% of the time, under broad focus those phrases are an IP in V+Prep. Our transcriptions of a sample of the recordings indicate a weaker but still strong enough phrase: under broad focus 48% of those phrases have a ToBI break index of at least 2 in V+Prep, and 39% of those phrases in V+Part have a ToBI break index of at least 2. Often those phrases end with a phrase accent, a general indicator of an iP boundary (or at least break index 2, if there is mismatch of phrase accent and subjective judgment of boundary strength). Thus, we assume that the phrase of interest is often larger than a prosodic word, and between an iP and an IP. Since the durational differences in narrow focus do not differ significantly from those in broad focus, we infer that post-focally, those accent-less phrases are just as strong. The fact that there can be accent-less iPs challenges the accent-first hypothesis. The fact that there are prominence distinctions in the absence of pitch accent also challenges this hypothesis. These results are consistent with the stress-first hypothesis and the equal stresses hypothesis.

Since there may be accent-less iPs, we may further ask if they have a phrase accent. The unique speaker who sometimes produced H- and sometimes L- in broad focus always produced H* L- H-L% in narrow focus, suggesting that accent-less iPs do have a phrase accent.

Assuming that duration of a primarily stressed rime reflects its prominence, the fact that the primarily stressed rime of the subject is longer under narrow focus than under broad focus suggests that it is more prominent under narrow focus than under broad focus. Wagner & McAuliffe (2019) also found this effect, where the stressed syllable was longer under focus. This is consistent with the equal stresses hypothesis, which promotes the focused constituent to a higher level in the metrical grid, but not with the stress-first hypothesis, which assumes that narrowly-focused constituents have the same prominence as they are under broad focus. Furthermore, our finding that the ratio of the duration of the subject's primarily stressed rime to the average of V1 and P1 is larger under narrow focus than broad focus suggests that the subject is more prominent relative to V and P in narrow focus than broad focus. This is again consistent with the equal stresses hypothesis but not with the stress-first hypothesis.

Another result of our experiment is that V1, V2, P1 and P2 are all shorter post-focally than in broad focus. Assuming that rime duration reflects its prominence, and thus its position on the metrical grid, this suggests that all these rimes may project to a lower level post-focally than in broad focus. This is not the prediction of any of the three hypotheses we reviewed. It is worth mentioning that Wagner & McAuliffe (2019) found a small reduction in duration for some post-focal words but not others. We leave to future research whether there is actually a

general shortening effect on post-focal material, and if not, why some post-focal words may be shortened but not others.

A reviewer asked about the small effect size of the durational differences. One possible reason why they are small is that as our transcriptions of the broad-focus sentences suggest, speakers do not always distinguish between the two syntactic structures with prosody, at least not as clearly as we had hoped for. For example, the prosodic boundary following V in V+Prep under broad focus is a normal word boundary over 50% of the time, which is much smaller than Price et al.'s (1991) report of an IP boundary there 60% of the time. Also, V1 can sometimes be accented in both V+Prep and V+Part. It is worth mentioning that Price et al's experimental materials had no fillers, and also they corrected the speakers when they produced a tune that sounded different from what the experimenters had expected. In contrast, our speakers were likely to be completely unaware of the syntactic differences we wanted to test, and thus did not exaggerate the prosodic differences.

We compare our effect sizes with two previous works that tested post-focal prominence and phrasing in English respectively–Beaver et al. (2007) and Norcliffe & Jaeger (2005). Beaver et al.'s (2007) findings are similar to our effect sizes for V1 and P1: they found that post-focally, second-occurrence words were on average 6 ms longer than non-focused words. Note that they measured the duration of an entire word, while we did the duration of a rime. In a perception study they also found that subjects were significantly above chance at distinguishing the second-occurrence focus from non-focus in their recordings.

Our effect sizes for V2 and P2 are smaller than Norcliffe & Jaeger's (2005), and we suggest three possible reasons for this. With a sample size of N=48, Norcliffe & Jaeger's (2005) found that post-focally, the final rime of P was on average 28 ms longer in V+Part than in V+Prep, but half of those recorded sentences put a pitch accent in the post-focal region. They found that the final rime of V was lengthened in V+Prep, but it did not reach significance. Their finding that P was lengthened more significantly than V was consistent with our findings. We think that three factors contributed to their finding of a larger durational difference. First, they did not have filler items (E. Norcliffe, p.c.), and thus speakers may have been aware of the syntactic differences they wanted to test, and exaggerated their prosodic differences when they spoke. Second, the effect they found may have been a cumulative one of two types of lengthening effects–effects due to phrasing and effects due to prominence. Third, they did not exclude post-focal pitch accents, thus including non-post-focal Vs and Ps in their data. Since we found that post-focal material was generally shortened compared to material in broad focus, their failure to exclude post-focal pitch accents may have led to longer durations for V and P, and therefore greater durational differences than an analysis that does exclude post-focal pitch accents.

5. Conclusion

This paper has reported a production study of English verb phrases that are ambiguous between verb + particle and verb + preposition structures. We found prosodic cues in phrasing and prominence that differentiated between these two syntactic structures under broad focus, and found that post-focally, the iP boundaries and prominence distinctions are preserved. This suggests that there may be accent-less iPs. We further found that those accent-less iPs may end with a phrase accent. We also found that focused material is lengthened and therefore more prominent, while post-focal material is shortened and less prominent. Our results challenge the hypothesis that every iP must have a head, and this head is marked solely by pitch accent, and are compatible with hypotheses that do not obey the headedness hypothesis, or those that allow an iP head to be marked by something other than pitch accent, such as phrasal stress. Pitch

accent is realized primarily by pitch excursion rather than duration, making duration a measure of phrasal stress.

References

Beaver, D. I., Clark, B., Flemming, E. S., Jaeger, T. F., & Wolters, M. (2007). When semantics meets phonetics: Acoustical studies of second-occurrence focus. *Language*, *83*(2), 245–276. https://doi.org/10.1353/lan.2007.0053

Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes*, *11*(1–2), 17–68. https://doi.org/10.1080/016909696387213

Beckman, M. E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. A. Keating (Ed.), *Phonological Structure and Phonetic Form* (1st ed., pp. 7–33). Cambridge University Press. https://doi.org/10.1017/CBO9780511659461.002

Beckman, M. E., & Elam, G. A. (1997). *Guidelines for ToBI labelling (version 3.0)*.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, *3*, 255–309. https://doi.org/10.1017/S095267570000066X

Bolinger, D. L. (1958). A theory of pitch accent in english. *Word*, *14*(2–3), 109–149. https://doi.org/10.1080/00437956.1958.11659660

Büring, D. (2013). Syntax, information structure,and prosody. In M. Den Dikken (Ed.), *The Cambridge Handbook of Generative Syntax* (1st ed., pp. 860–896). Cambridge University Press. https://doi.org/10.1017/CBO9780511804571.029

Carlson, K. (2009). How prosody influences sentence comprehension. *Language and Linguistics Compass*, *3*(5), 1188–1200. https://doi.org/10.1111/j.1749-818X.2009.00150.x

Cohan, J. (2000). *The Realization and Function of Focus in Spoken English*. University of Texas, Austin.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and Speech*. Harvard University Press. https://doi.org/10.4159/harvard.9780674283947

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. https://doi.org/10.1016/j.wocn.2004.05.002

Dimitrova, S., & Turk, A. (2012). Patterns of accentual lengthening in English four-syllable words. *Journal of Phonetics*, *40*(3), 403–418. https://doi.org/10.1016/j.wocn.2012.02.008

Féry, C. (2011). German sentence accents and embedded prosodic phrases. *Lingua*, *121*(13), 1906–1922. https://doi.org/10.1016/j.lingua.2011.07.005

Gussenhoven, C. (1983). Focus, mode and the nucleus. *Journal of Linguistics*, *19*(2), 377–417. https://doi.org/10.1017/S0022226700007799

Halliday, M. A. K. (1967). *Intonation and Grammar in British English:* De Gruyter. https://doi.org/10.1515/9783111357447

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. Univ. of Chicago Press.

Hayes, B., & Lahiri, A. (1991). Bengali intonational phonology. *Natural Language and Linguistic Theory*, *9*(1), 47–96. https://doi.org/10.1007/BF00133326

Horne, M. (1993). The phonetics and phonology of the rhythm rule in post-focal position: Data from English. In *Studies Presented to Claes-Christian Elert on the Occasion of His Seventieth Birthday* (Vol. 2, pp. 69–77). Dept. of Phonetics, Univ. of Umeå.

Ishihara, S. (2003). *Intonation and Interface Conditions*. Massachusetts Institute of Technology.

Ishihara, S. (2007). Major phrase, focus intonation, multiple spell-out (MaP, FI, MSO). *The Linguistic Review*, *24*(2–3). https://doi.org/10.1515/TLR.2007.006

Ishihara, S. (2016). Japanese downstep revisited. *Natural Language & Linguistic Theory*, *34*(4), 1389–1443. https://doi.org/10.1007/s11049-015-9322-8

Ito, J., & Mester, A. (2003). Weak layering and word binarity. In T. Honma, M. Okazaki, T. Tabata, & S. Tanaka (Eds.), *A New Century of Phonology and Phonological Theory. A Festschriftfor Profssor Shosuke Haraguchi on the Occasion of His Sixtieth Birthday* (pp. 26–65). https://people.ucsc.edu/~ito/papers/1992_ito_mester_weaklayering.pdf

Jun, S.-A. (2011). Prosodic markings of complex NP focus, syntax, and the pre-/post-focus string. In M. B. Washburn, K. McKinney-Bock, E. Varis, A. Sawyer, & B. Tomaszewicz (Eds.), *Proceedings of the 28th West Coast Conference on Formal Linguistics*. Cascadilla Proceedings Project.

Jun, S.-A., & Fougeron, C. (2000). A phonological model of French intonation. In A. Botinis (Ed.), *Intonation* (Vol. 15, pp. 209–242). Springer Netherlands. https://doi.org/10.1007/978-94-011-4317-2_10

Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology*, *50*(2), 194–231. https://doi.org/10.1016/j.cogpsych.2004.08.002

Kügler, F., & Féry, C. (2017). Postfocal downstep in German. *Language and Speech*, *60*(2), 260–288. https://doi.org/10.1177/0023830916647204

Ladd, D. R. (1980). *The Structure of Intonational Meaning: Evidence from English*. E-CONTENT GENERIC VENDOR.

Ladd, D. R. (1996). *Intonational Phonology*. Cambridge University Press.

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *The Journal of the Acoustical Society of America*, *53*(1_Supplement), 380–380. https://doi.org/10.1121/1.1982702

Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. T. Oehrle (Eds.), *Language Sound Structure* (pp. 157–233). MIT Press.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal forced aligner: Trainable text-speech alignment using Kaldi. *Interspeech 2017*, 498–502. https://doi.org/10.21437/Interspeech.2017-1386

McAuliffe, M., & Sonderegger, M. (2024). *English (US) ARPA acoustic model* (3.0.0) [Computer software]. https://mfa-models.readthedocs.io/acoustic/English/English (US) ARPA acoustic model v3_0_0.html

Newman, S. S. (1946). On the Stress System of English. *Word*, *2*(3), 171–187. https://doi.org/10.1080/00437956.1946.11659290

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

Okobi, A. O. (2006). *Acoustic Correlates of Word Stress in American English*. Massachusetts Institute of Technology.

Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *The Journal of the Acoustical Society of America*, *66*(2), 363–369. https://doi.org/10.1121/1.383670

Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. Massachusetts Institute of Technology.

Pierrehumbert, J., & Hirschberg, J. (1990). The Meaning of Intonational Contours in the Interpretation of Discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in Communication* (pp. 271–312). The MIT Press. https://doi.org/10.7551/mitpress/3839.003.0016

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America*, *90*(6), 2956–2970. https://doi.org/10.1121/1.401770

Schafer, A. J., Warren, P., Speer, S. R., & White, S. D. (2000). Prosodic indication of syntactic structure in quasispontaneous speech. *The Journal of the Acoustical Society of America*, *108*(5_Supplement), 2466–2466. https://doi.org/10.1121/1.4743091

Schmerling, S. F. (1976). *Aspects of English Sentence Stress*. University of Texas Press. https://doi.org/10.7560/703124

Selkirk, E. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry*, *11*(3), 563–605.

Selkirk, E. (1984). *Phonology and Syntax: The Relation Between Sound and Structure*. MIT press.

Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology Yearbook*, *3*, 371–405. https://doi.org/10.1017/S0952675700000695

Selkirk, E. (1995). Sentence prosody: Intonation, stress and phrasing. In J. A. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *The Handbook of Phonological Theory* (pp. 550–569). Blackwell. https://www.semanticscholar.org/paper/Sentence-Prosody%3A-Intonation%2C-Stress-and-Phrasing-Selkirk/f8ebaf1db9b2708eec2fcb8b45d0726b05fffa15

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*(2), 193–247. https://doi.org/10.1007/BF01708572

Silverman, K. E. A., & Pierrehumbert, J. B. (1987). The timing of prenuclear high accents in English. *The Journal of the Acoustical Society of America*, *82*(S1), S19–S19. https://doi.org/10.1121/1.2024693

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*(4), 2471–2485. https://doi.org/10.1121/1.417955

Sugahara, M. (2003). *Downtrends and post-FOCUS intonation in Tokyo Japanese*. University of Massachusetts, Amherst.

Truckenbrodt, H. (1995). *Phonological Phrases—Their Relation to Syntax, Focus, and Prominance*. Massachusetts Institute of Technology.

Truckenbrodt, H. (2006). Phrasal stress. In *Encyclopedia of Language & Linguistics* (pp. 572–579). Elsevier. https://doi.org/10.1016/B0-08-044854-2/04447-3

Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, *35*(4), 445–472. https://doi.org/10.1016/j.wocn.2006.12.001

Vanderslice, R., & Ladefoged, P. (1972). Binary Suprasegmental Features and Transformational Word-Accentuation Rules. *Language*, *48*(4), 819. https://doi.org/10.2307/411990

Wagner, M. (2005). *Prosody and Recursion*. Massachusetts Institute of Technology.

Wagner, M. (2010). Prosody and recursion in coordinate structures and beyond. *Natural Language & Linguistic Theory*, *28*(1), 183–237. https://doi.org/10.1007/s11049-009-9086-0

Wagner, M., & McAuliffe, M. (2019). The effect of focus prominence on phrasing. *Journal of Phonetics*, *77*, 100930. https://doi.org/10.1016/j.wocn.2019.100930

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, *91*(3), 1707–1717. https://doi.org/10.1121/1.402450

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, *33*(2), 159–197. https://doi.org/10.1016/j.wocn.2004.11.001