Perception of Phi-Phrase boundaries in Hindi

Somnath Roy
Center for Linguistics
Jawaharlal Nehru University
New Delhi-110067
somnathroy86@gmail.com

Abstract

This paper proposes an algorithm for finding phonological phrase boundaries in sentences with neutral focus spoken in both normal and fast tempos. A perceptual experiment is designed using Praat's experiment MFC program to investigate the phonological phrase boundaries. Phonological phrasing and its relation to syntactic structure in the framework of the end-based rules proposed by (Selkirk, 1986), and relation to purely phonological rules, i.e., the principle of increasing units proposed by (Ghini, 1993) are investigated. In addition to that, this paper explores the acoustic cues signalling phonological phrase boundaries in both normal and fast tempos speech. It is found that phonological phrasing in Hindi follows both end-based rule (Selkirk, 1986) and the principle of increasing units (Ghini, 1993). The end-based rules are used for phonological phrasing and the principle of increasing units is used for phonological phrase restructuring.

1 Introduction

A phonological phrase (ϕ-phrase) is a short breathe group, which speakers sometimes shorten or elongate (Ohala, 1978). The phonological rules which apply across words are covered under the domain of ϕ-phrase. Two important theories of ϕ-phrasing—end-based theory (Selkirk, 1986) and relation-based theory (Nespor and Vogel, 1986) are the base works in this area. (Ghini, 1993) proposed purely phonological rules for ϕ-phrasing while the phrasing rules in (Hayes and Lahiri, 1991) are based on c-command relation.

The idea of Selkirk's end based rule is based on the syntactic constituents. According to which, the right edge of each syntactic XP coincides with the right edge of a ϕ-phrase. A syntactic XP is a phrase where X represents the head of that phrase. In short, end-based rule is written as Align(XP, R, ϕ, R), i.e., the right edge of each XP must be aligned to the right edge of ϕ-phrase (Truckenbrodt, 1995). See the following examples from (Truckenbrodt, 1995)

1.

a. [V NP $]_{VP} \rightarrow$ (V NP)ϕ
e.g., (ingile mtana:ni)ϕ /entered the room/
b. [V PP $]_{VP} \rightarrow$ (V PP)ϕ
e.g., (mapendo ya maski:ni)ϕ /the love of a poor man/
c.[N AP $]_{NP} \rightarrow$ (N AP)ϕ
e.g., (nthi:-khavu) / dry land/

2. [NP V] $\rightarrow$ (NP)ϕ (V)ϕ
e.g., (maski:ni ha:tali) /a poor man does not choose/

3. For complex sentence having more than one NP following phrasing pattern is applied.

a. [NP V NP] $\rightarrow$ (NP)ϕ (V NP)ϕ
b. [NP NP] $\rightarrow$ (NP)ϕ (NP)ϕ
c. [V NP NP] $\rightarrow$ (V NP)ϕ (NP)ϕ

(Ghini, 1993) proposed purely phonological rules for ϕ-phrasing in Italian. He preferred binary branching over n-ary branching as proposed in Nespor and Vogel. His claims are based on the following observations.

a. ϕ-phrasing is related to the concept of branching. The branching XPs are never restructured with its preceding branching phrases, because the branching XPs are longer and contain more phonological material than the non-branching XPs.

D S Sharma, R Sangal and A K Singh. Proc. of the 13th Intl. Conference on Natural Language Processing, pages 324–330,
Varanasi, India. December 2016.

b. The XPs use concept of weight, like a branching XPs are heavy and non-branchings are light. The heavy XPs contain more number of words than the light XPs.

c. The ϕ-phrasing is purely phonological. A string should be parsed into same length ϕs. By increasing and decreasing the tempo the ϕ-phrase is increased and decreased by one word respectively.

d. He has proposed the concept of phrasing based on increasing unit (prosodic word) in a phrase, which is called principle of increasing units. According to this principle, a preceding ϕ-phrase should not have higher weight than the following ϕ-phrase.

The motivation for this work lies in the following two points.

i. Phonological phrasing plays an important role in language comprehension. The information of ϕ-phrase boundaries is requisite in training data for developing a text-to-speech synthesis system with non-robotic voice quality. This work fulfills the requirement of ϕ-phrasing in Hindi text-to-speech synthesis system development.

ii. The role of spectral peaks in phonological phrasing as a predictor variable is investigated for both normal and fast tempo speech.

The rest of the paper is organized as follows. Section 2 describes the experimental details. Section 3 is the description of the algorithm for phonological phrasing in Hindi. Section 4 describes the significant acoustic cues which signal the phonological phrase boundary to the listeners. The conclusion is written in section 5.

2 Experimental Details

This section describes linguistic data collection and speech recording. It also describes the acoustic cues extracted for the analysis.

2.1 Linguistic Data

Fifty spoken sentences were selected for recording. The sentences were a combination of simple and complex declarative. The smallest sentence has four words, the largest sentence has fifteen words. The total count is 447 words in all 50 sentences. Ten speakers 5 males and 5 females were recorded, resulting into a total of 500 sentences. The speakers were either Delhite(born and brought up in Delhi) or one who has been studying in Delhi for the last 10 years. Each speaker is given few minutes to get familiar especially with the long sentences to mitigate hesitation while recording. If hesitation occurs for a sentence; speakers were advised to repeat for that sentence.

The sentences were recorded in two tempos—normal and fast. Both the tempos were used in neutral context only. The normal tempo for this study is the rate at which a speaker normally speaks. The recording was carried out in a noise proof sound recording studio. The speech was recorded at 44.1 kHz sampling rate and stored as 16 bits PCM data. The sentences for recording were selected from news websites covering news ranging from national to international, weather forecast, sports and entertainment. The sentences selected from these news websites were first corrected for Unicode rendering and then used for the recording.

2.2 Perceptual Experiment for ϕ-phrase boundary

A perception based experiment is performed to determine the ϕ-phrasing in Hindi. Ten persons participated in the experiment who are the native speakers of Hindi. Neither they have any formal background in linguistics nor they were taught about the phonological phrasing. The stimuli presented to the informants using the Praat's experiment MFC program. The response options were of four possible types of ϕ-phrases for each sentence. The reason for the limited number of possible response types are the following.

a. There can be many logical options depending upon the complexity of a sentence.

b. The best four plausible response choices are used for each experiment which in turn is decided by the expert.

The phrases were separated by square bracket. The participant were told that the square bracket denotes the short pause; click on the response which suitably shows the short pause in the stimuli. The participants were also asked to choose the confidence measure ranging from 1(for poor)—5(good). "Replay" button can be used for playing the same speech files again and again upto 10 times and "OOPS" button is for listening to the previous sound file. A test experiment having seven files is run first, and participants are asked to choose the suitable option. It is ensured that every participant understood the experiment clearly. At the end of the test experiment, they are asked two questions: 1—"Are you ok with the experiment?" and 2—"Do you have any question to ask?". The actual set of fifty sentences for both tempos are mixed randomly and run to each participant separately. The participants were advised to take a break of 15 minutes after continuously listening for 1 hour. The process completed in two days. The average time taken by informants is 11 hours and 42 minutes in two days. For clarity, see the Figure 1.

All utterances are annotated by all 10 annotators and the average pairwise kappa values are calculated for inter-informant agreement. The kappa values for inter-informant agreement for normal and fast tempo speech lie in the range of 0.7—0.9 and 0.6—0.7 respectively. The response option getting highest agreement is picked up for the analysis in both tempos. It is found that 2 sentences have different phrasing patterns for normal and fast tempos. Therefore, 2×10 (because the same sentence was recorded by all 10 speakers) are dropped and not analyzed in this study. The remaining 480 sentences show same phonological phrasing. The phonological phrasing rules based on the phrasing patterns selected by participants are presented in section 3.

2.3 Acoustic cues Extracted at ϕ-phrase Boundaries

Nine acoustic cues are extracted from the preboundary and the postboundary syllable of each ϕ-phrase, for clarity see the Figure 2 & 3.

These acoustic cues include the variants of temporal and spectral cues. These are the minimum pitch at the start of a syllable (InF0Min), max pitch (F0Max), minimum pitch in the end of a syllable (FiF0Min), mean intensity(AvgInt), minimum intensity (MinInt), maximum intensity (MaxInt), three spectral peaks at first, second and third formants (SpecInt1, SpecInt2 and SpecInt3). The acoustic cues pitch, intensity and duration are the most accepted and investigated predictor variables at phrase boundaries (Oller, 1973)(Streeter, 1978). Duration is not included as an acoustic cue in this experiment. Because, there are ample studies which unanimously accepted it as a relevant cue for the prosodic phrasing and prominence (Klatt, 1975)(Oller, 1973) (Lehiste and Lass, 1976) (MacDonald, 1976) (Roy, 2014). Therefore, less established acoustic cues are included in this study. The objective is to investigate the relevance of new acoustic cues in phonological phrasing rather than reaffirming the importance of an already established acoustic cue. In this investigation spectral peaks are included as predictors because it signal perceptual stress(Sluijter et al., 1997) in an utterance. All nine cues are extracted using Praat software for speech signal analysis (Boersma and Weenink, 2013). The analysis of the importance of these cues in ϕ-phrasing is discussed in section 4.

3 Rules for ϕ-phrasing in Hindi

Hindi is a relatively free word order language. It follows SOV (subject followed by an object and then followed by a verb) structure. Based on the data collected from the performance and perception based experiment presented in section 2.2, the rules for ϕ are illustrated with examples. An intonation based analysis by (Patil et al., 2008) suggests that each content word in Hindi could be a prosodic phrase. The present analysis partially disagree to (Patil et al., 2008) as explained below (Parse tree is shown in Figure 4—7). See the following examples (the ϕ-phrases shown below for each sentence is the same chosen by the participants during the perception experiment).

a. raːm aːm kʰaːtaː hæː / Ram eats a mango/

[raːm]ϕ [aːm]ϕ [kʰaːtaː hæː]ϕ (normal & fast tempo)

b. siːtaː eːk gaːjikaː hæː. /Sita is a singer/

[sitaː]ϕ [eːk gaːjikaː hæː]ϕ (normal & fast tempo)

c. raːm neː siːtaː koː kitaːb diː /Ram gave a book to Sita/

[raːm -neː]ϕ [siːtaː koː]ϕ [kitaːb diː]ϕ (normal & fast tempo)

d. billiː məmmiː keː piːtʃheː bɦaːgiː / Cat ran behind the mother/

[billiː]ϕ [mə mmiː keː]ϕ [piːtʃʰeː bɦaːgiː] ϕ

In the above example (a), which is of type [NP NP V] for which informants choose the phrasing pattern [NP]ϕ [NP]ϕ [V]ϕ. The example (b) which is of the same pattern as of (a). However the phrasing rules of (a) if applied on (b), the phrases breaks into the length of 1+2+1. Such phrasing patterns violate the principle of increasing units and thus restructuring applies. And finally it becomes [NP]ϕ [NP V]ϕ; with phrase

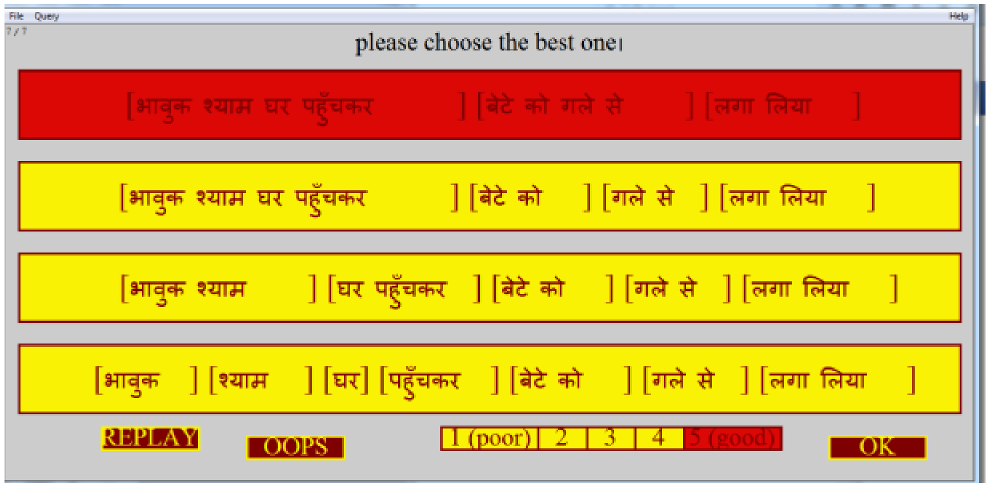

Figure 1: Praat's MFC experiment having four choices for the sentence "भावुक श्याम घर पहुँचकर बेटे को गले से लगा लिया", where square brackets denote short pauses or phonological phrase boundaries.

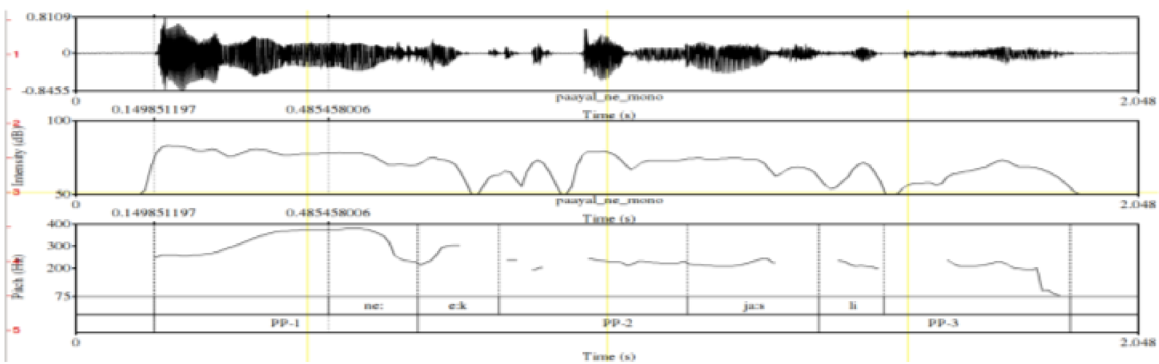

Figure 2: Preboundary and Postboundary syllables in the sentence, paːjəl neː eːk upənjaːs likhaː hei spoken in normal tempo. PP-1, PP-2 and PP-3 are three phonological phrase for which boundary is marked.

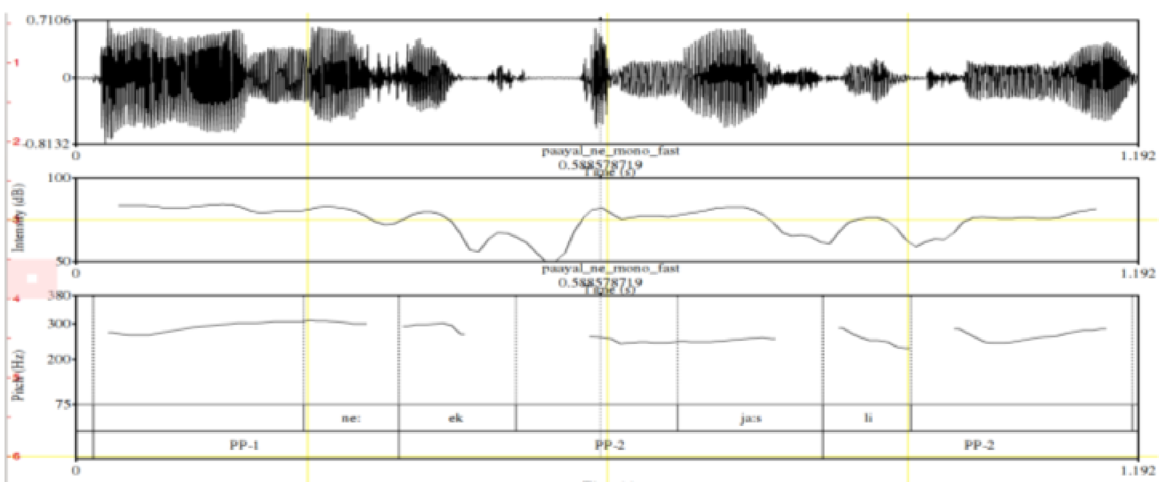

Figure 3: Preboundary and Postboundary syllables in the sentence, paːjəl neː eːk upənjaːs likhaː hei spoken in fast tempo. PP-1, PP-2 and PP-3 are three phonological phrase for which boundary is marked.

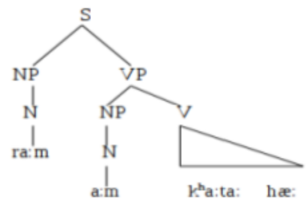

Figure 4: parse tree example1

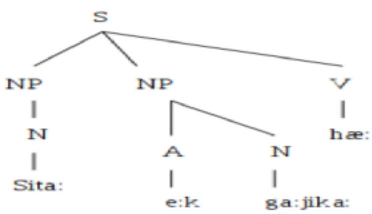

Figure 5: parse tree example2

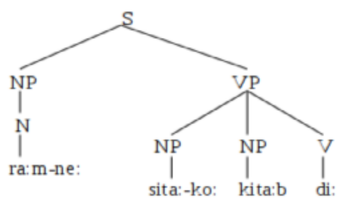

Figure 6: parse tree example3

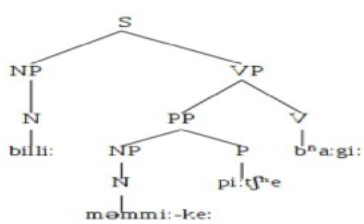

Figure 7: parse tree example4

length in increasing pattern i.e., 1+3. Example (b) also states that copula restructures itself with preceding ϕ-phrase. The example (c) is same as of the example (b), but differs on the category of verb at ultimate position. In this case it is a main verb and restructure itself with the preceding ϕ-phrase. The most important one is the fourth example (d), where the phrasing should be either [NP]ϕ [PP]ϕ [V]ϕ or [NP]ϕ [NP]ϕ [P]ϕ [V]ϕ but informants chose the phrasing pattern [NP]ϕ [NP]ϕ [P V]ϕ. It concludes that the restructuring takes place because both of the phrasing patterns i.e 1+3+1 and 1+2+1+1 violates principle of increasing units. Another important point need to be highlighted here is that in the ϕ-phrase [P V]ϕ, the postposition (P) modifies the verb. Therefore P is no longer a postposition rather an adverb. Thus the word piːtʃ^{h}eː /behind/ in example (d), is a content word which along with another content word constitutes a ϕ-phrase. There are many such examples in Hindi where a single content word does not form a ϕ-phrase by itself rather restructure with some other content word. This implies that the process of ϕ-restructuring is of central significance in Hindi, and the claim that each content word could be a ϕ-phrase seems feeble.

Algorithm: ϕ-phrasing

i. Parse the input sentence.

ii. Apply end based rule(Selkirk, 1986).

iii. If phrase are not in increasing units, restructure it using principle of increasing units (Ghini, 1993)

4 Significance of Acoustic cues in ϕ-phrasing

The acoustic cues described in section 2.3 are analyzed. One way ANOVA modeling is applied at α=0.01. The ϕ-phrases are categorized into three types like initial ϕ-phrase (PhInit), medial ϕ-phrase (PhMed) and final ϕ-phrase (PhFin). The acoustic cues become factors and three phrases work as the level for ANOVA modeling.

The Table 1 and Table 2 describe the result of ANOVA Modeling for both normal and fast tempos speech respectively. The result indicates that the different type of acoustic cues are responsible for the ϕ-phrase break in normal and fast tempos, even though their ϕ-phrases have same boundaries for a sentence.

For normal tempo F0Max is the most significant parameter and other two significant acoustic cues are MaxInt and SpecInt1. The average difference of 10.1 Hz—22.4 Hz (approx) is found for F0Max at the ϕ-phrase boundaries. The comparison of MaxInt at consecutive ϕ-phrase boundaries shows that there is an average difference of 3dB—13.4 dB (approx) between the syllables at preceding and succeeding ϕ-phrase at boundaries. The noticed range of difference is significant, which provides listeners a cue for break. The significance of spectral peak at first formant (i.e., SpecInt1) states that intensity at higher frequency too is an important factor in determining ϕ-phrase boundary. SpecInt1 shows a significant difference at the boundary and difference lies in the range of 3.8 dB to 22.6dB (approx). Such pattern infers that each of these acoustic cues show declination between two consecutive ϕ-phrase boundaries as can be seen in Figure 2. The effect of declination in these acoustic cues are perceived as a short break or ϕ-phrase boundary to the listeners.

For Fast tempo F0Max is the most significant parameter and the other significant parameter is FiF0Min. MaxInt and SpecInt1 are not the significant parameters in fast tempo speech. F0Max shows significant change at phrase boundaries as in normal tempo, however it is not declination always. In other words, the next ϕ-phrase may have higher F0Max then the preceding one. The average difference in F0Max at consecutive ϕ-phrase boundaries lies in the range of 5.8 Hz to 16.9 Hz. FiF0Min in fast tempo shows pitch reset phenomena as can be seen in Figure 3. Each ϕ-phrase will have different level of pitch range from the beginning. The average difference for FiF0Min at consecutive ϕ-phrase boundaries are in the range of 8.6Hz—17.9Hz.

These results can be analyzed from different perspective. One perspective is that both normal and fast tempos show significant change in F0Max at ϕ-phrase boundaries. In other words, phrase break is signaled by the declining value in F0Max in normal tempo speech, but both rise and decline in F0Max are the factor signalling for ϕ-phrase boundaries in fast tempo speech. Also in normal tempo speech at phrasal boundaries speakers max intensity and spectral peak at first formant show declination but in fast tempo this effect is not found. Hence in normal tempo, speakers form a negative slope for these intensities between two adjacent boundaries. But in fast tempo speakers keep their intensity value almost constant between the adjacent boundaries.

5 Conclusion

The phonological phrasing in Hindi for a given sentence is same irrespective of the rate of speech. That is, both tempos yield same phonological phrasing. The phonological phrase restructuring rule in normal tempo follows the the principle of increasing units by (Ghini, 1993). The acoustic cues signaling phonological phrase boundary vary for both tempos. The acoustic parameters F0Max, MaxInt and SpecInt1 are the significant contributors for phonological phrasing in normal tempo speech, but for fast tempo F0Max and FiF0Min are the significant contributors for the phonological phrasing.

cues	**F-Value**	**Pr>(F)**	**Significance**
F0Max	11.584	1.93e-05	0.001
MaxInt	5.673	0.00412	0.01
SpecInt1	4.866	0.00882	0.01

Table 1: Anova Modeling for Significant Acoustic Cues in Normal Tempo Speech

cues	**F-Value**	**Pr>(F)**	**Significance**
F0Max	8.115	0.000491	0.001
FiF0Min	8.3	00.000417	0.001

Table 2: Anova Modeling for Significant Acoustic Cues in Fast Tempo Speech

References

Paul Boersma and David Weenink. 2013. Praat software. *Amsterdam: University of Amsterdam.*

Mirco Ghini. 1993. Phi-formation in italian: a new proposal. *Toronto Working Papers in Linguistics*, 12(2).

Bruce Hayes and Aditi Lahiri. 1991. Bengali intonational phonology. *Natural Language & Linguistic Theory*, 9(1):47–96.

Dennis H Klatt. 1975. Vowel lengthening is syntactically determined in a connected discourse. *Journal of phonetics*, 3(3):129–140.

Ilse Lehiste and Norman J Lass. 1976. Suprasegmental features of speech. *Contemporary issues in experimental phonetics*, 225:239.

Nina H MacDonald. 1976. Duration as a syntactic boundary cue in ambiguous sentences. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'76.*, volume 1, pages 569–572. IEEE.

Marina Nespor and Irene Vogel. 1986. *Prosodic phonology*, volume 28. Walter de Gruyter.

John J Ohala. 1978. Production of tone. *Tone: A linguistic survey*, pages 5–39.

D Kimbrough Oller. 1973. The effect of position in utterance on speech segment duration in english. *The journal of the Acoustical Society of America*, 54(5):1235–1247.

Umesh Patil, Gerrit Kentner, Anja Gollrad, Frank Kügler, Caroline Féry, and Shravan Vasishth. 2008. Focus, word order and intonation in hindi. *Journal of South Asian Linguistics*, 1(1).

Somnath Roy. 2014. Prominence detection in hindi: A mathematical perspective. In *Computational Science and Computational Intelligence (CSCI), 2014 International Conference on*, volume 2, pages 119–124. IEEE.

Elisabeth O Selkirk. 1986. *Phonology and syntax: the relationship between sound and structure*. MIT press.

Agaath MC Sluijter, Vincent J Van Heuven, and Jos JA Pacilly. 1997. Spectral balance as a cue in the perception of linguistic stress. *The Journal of the Acoustical Society of America*, 101(1):503–513.

Lynn A Streeter. 1978. Acoustic determinants of phrase boundary perception. *The Journal of the Acoustical Society of America*, 64(6):1582–1592.

Hubert Truckenbrodt. 1995. *Phonological phrases–their relation to syntax, focus, and prominance*. Ph.D. thesis, Massachusetts Institute of Technology.