SIMULTANEOUS ENCODING AND DECODING OF LARYNGEAL CONTRASTS AND PROSODIC STRUCTURE IN SEOUL KOREAN

A Dissertation Presented

by

SEUNG SUK LEE

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2025

Department of Linguistics

SIMULTANEOUS ENCODING AND DECODING OF LARYNGEAL CONTRASTS AND PROSODIC STRUCTURE IN SEOUL KOREAN

A Dissertation Presented

by

SEUNG SUK LEE

Approved as to style and content by:

Kristine M. Yu, Chair

John Kingston, Member

Mara Breen, Member

Seth Cable, Department Chair
Department of Linguistics

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to my chair, Kristine Yu, for her endless support and encouragement throughout my time at UMass. She taught me so much about academic writing and how to tackle difficult questions. She always believed in me and my sometimes too-ambitious ideas, and helped me turn them into something more concrete and feasible. She read every draft I ever wrote for anything so meticulously and carefully, and always returned it in a much better shape with so many constructive comments. I am so lucky to have an advisor like her, who is not only incredibly knowledgeable but also genuinely cares about her students' well-being and success. Her guidance is what made me always feel supported at UMass.

I can't thank John Kingston enough for his invaluable feedback and support throughout my time at UMass. He was in every single one of my committees (both my GPs and my dissertation). I can't imagine going through my PhD without his encouragement, insights, and expertise. He did everything he could to help me succeed, and I am so grateful for his mentorship. I remember that he offered to meet at evenings to accommodate 14 hours of time difference, when I was working on my GP1 in South Korea during COVID-19. He helped me get through some of the most challenging times in my PhD journey, and I will always be thankful for his generosity and kindness. He will always be my role model as a linguist, a mentor, and a researcher.

I am also very grateful to Mara Breen for her insightful comments and suggestions on earlier drafts of my dissertation. I will miss our zoom meetings and chatting about baseball!

Besides the three members of my committee, I am also grateful to all the faculty and staff in the Linguistics Department at UMass. Joe Pater has always been so encouraging and supportive as we worked on various projects on learning and representing stress typology. As he put it so well, he was the chair of my 'second dissertation'. I am also very grateful to Gaja Jarosz who was the other advisor on my GP1. Talking to her helped me get through the difficult times during the pandemic year. I am also grateful to other faculty members who helped me grow as a linguist, including Faruk Akkuş, Ana Arregui, Michael Becker, Rajesh Bhatt, María Biezma, Seth Cable, Brian Dillon, Lisa Green, Vincent Homer, Kyle Johnson, Shota Momma, and Magda Oiry.

Before my time at UMass, I was fortunate to have worked with great mentors, including Kelly Berkson and Isabelle Darcy at Indiana University Bloomington. My passion for phonetics started when I took 'Introduction to Phonetics' with Kelly Berkson, who taught me that lenis stops in Korean are not 'voiced' and it is realized with aspiration (it was an introductory course and she did not get into prosodic conditioning of lenis obstruents). Kelly, this dissertation started in that class you taught!

I met so many wonderful friends at UMass and I would also like to thank them for their friendship. I need to thank especially Alessa Farinella and Cerys Hughes for being my P-side cohort. We three were inseparable in our classes and various projects and I look forward to doing more collaborations in the future. I also want to thank my cohort members–Mariam Asatryan, Yosho Miyata, Yixiao Song, and Ayana Whitmal–for their unconditional support and friendship. We have shared so many good and difficult times together and I could not have survived graduate school without them.

In addition to these people, I would like to thank all other graduate students and staff members I have met at UMass: Narda Wakoluk, Michelle McBride, Brandon

Prickett, Andrew Lamont, Kaden Holladay, Anissa Neal, Jonathan Pesetsky, Angelica Hill, Jia Ren, Polina Kasyanova, Eva Neu, Özge Bakay, Mari Kugemoto, Ali Nirheche, Yi-Shih Helen Chen, Brynne Wilkinson, Mariana Caderón, Satoru Ozaki, Suet-Ying Lam, Hee Joong Choi, Andrea Matticchio, Shaunak Phadnis, Carla Spellerberg, Nawal Bahrani, Roger Liu, Yanran Mou, Duygu Demiray, Kenta Kakenami, and Nir Segal.

I would like to thank my friends and family in the US and in South Korea who have always been there for me. I am grateful to Khidong Kim, who has been a constant source of support since high school. Thanks to my older brother Donghoon (Dr. Lee No. 2) and my sister-in-law Bri for their support. Thanks to my other brother and sister-in-law, Jeeyoon and Ian, and my parents-in-law for their encouragement. I would like to thank my parents Semee Hong and Yongho Lee (Dr. Lee No. 1) for their unconditional love and support. I love you guys so much. I appreciate everything you have done for me.

My final thanks go to my loving wife, Shin Young Park. Shin Young, I quite literally could not have done this without you. You helped me enjoy life outside of academia and kept me sane during the most stressful times. While I've been working on this dissertation, you've been making something much more wonderful: our son. I love you (both) so much and I look forward to raising our son together with you.

ABSTRACT

SIMULTANEOUS ENCODING AND DECODING OF LARYNGEAL CONTRASTS AND PROSODIC STRUCTURE IN SEOUL KOREAN

SEPTEMBER 2025

SEUNG SUK LEE

B.A, INDIANA UNIVERSITY BLOOMINGTON

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Kristine M. Yu

This dissertation investigates how segmental contrast and prosodic structure are acoustically implemented in production and processed from the acoustic signal in perception.

In Seoul Korean, it has been widely accepted that the primary acoustic cue for signaling prosodic structure is the F0 contour (Jun, 1993). Prosodic structure has also been reported to affect the voicing of lenis obstruents in Seoul Korean (Jun, 1993) but due to the finding that this segmental process is optional (Jun, 1994; Han, 2000), it has been assumed that segmental realization is a less reliable cue for prosodic structure than the F0 contour (Jun, 1994).

This dissertation presents findings that suggest that there may be a greater role for segmental realization in signaling prosodic structure than previously thought. First, I report empirical evidence that suggests the realization of lenis obstruents is in fact more systematically affected by the prosodic context than previously reported. Second, contrary to previous assumptions that F0 contours more reliably signal prosodic structure than segmental realization, I report evidence that segmental realization may in fact be more reliable in signaling prosodic structure in certain contexts in Seoul Korean. Finally, these findings that the realization of lenis obstruent is a reliable cue for prosodic structure are supported by the results of a perception experiment, which shows that listeners are more affected by the segmental realization of lenis obstruents than by the F0 contour in making decisions about prosodic structure.

Based on these findings, I argue that acoustic cues simultaneously encode both segmental contrast (lenis vs. aspirated stops) and prosodic structure (phrase-initial vs. phrase-medial) in Seoul Korean, and listeners hearing these acoustic cues decode both types of information simultaneously.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1 Prosodic encoding and decoding

This dissertation investigates how segmental contrast and prosodic structure are acoustically implemented in production and processed from the acoustic signal in perception.

There is a long list of previous work reporting evidence that underlying phonological, morphological, syntactic, and informational structures are acoustically encoded in the speech signal in production (e.g., Selkirk, 1984; Nespor & Vogel, 1986; Cho, 2001), and that listeners pay attention to such fine phonetic details and make syntactic parsing decisions based on them in auditory sentence processing (e.g., Lehiste et al., 1976; Scott & Cutler, 1984; Shattuck-Hufnagel & Turk, 1996).

When speakers produce utterances, they encode at least three kinds of information in the acoustic signal: (1) what the sequence of phones is, (2) how the sequence is grouped or chunked, and (3) which of these chunks are prominent or salient relative to other chunks (Beckman, 1996; Shattuck-Hufnagel & Turk, 1996). The first pertains to producing the segments with the relevant acoustic cues that distinguish the phonological contrasts that exist in the language (e.g., /ban/ vs. /pan/). The second is related to encoding the underlying structure into the utterance, such that the linguistic units that are necessary to understand the message, e.g., morphemes or syntactic phrases, can be retrieved by the listener. In speech production, speakers 'encode' such linguistic structure in the acoustic signal, by producing the acoustic signal in 'chunks' delimited by various phonetic cues such as duration, F0, and intensity.

This chunking of the acoustic signal is referred to as the 'delimitative' function of prosody in the literature (Shattuck-Hufnagel & Turk, 1996). Finally, in an utterance, one of the syntactic constituents may be marked as signalling new information, or as contrastive, in the information structure. Such a constituent may be realized with a particular prosodic prominence, making it more prominent than other ones within the same utterance (Shattuck-Hufnagel & Turk, 1996; Truckenbrodt, 1995). This function of prosody is referred to as the 'culminative' function of prosody (Shattuck-Hufnagel & Turk, 1996).

In speech perception, this encoding process is reversed. Listeners perceive disruptions in the signal and 'decode' the acoustic signal into a phonological representation: segments that are grouped into hierarchically prosodic units such as syllables, feet, prosodic words, and prosodic phrases. These units are related to syntactic units that are needed to understand the message, i.e., morphemes, lexical words, and syntactic constituents.

From this 'structural' view, prosody includes 'both the higher level organization, with its constituent boundaries and prominences, and the phonetic reflexes of this organization in the pattern of F0, duration, amplitude and segment quality/reduction within an utterance' (Shattuck-Hufnagel & Turk, 1996, p. 196). This dissertation investigates the interdependency and simultaneity of the encoding and decoding of segmental contrast and delimitation of speech signal. Specifically, when a segment varies in its acoustic quality as a function of the prosodic context (Nespor & Vogel, 1986), its acoustic realization signals both the prosodic structure and the segmental contrast. I investigate this phenomenon from the lens of a particular model of speech perception proposed in Cho et al. (2007), which assumes that the 'prosodic structure is built in parallel to lexical/segmental analysis' in speech perception (Cho et al., 2007, p. 233).

I will explain this parallel analysis with a schematic example. Consider a hypothetical language with a word-initial phonemic contrast of X and Y: there exists a minimal pair of words in this language that differ in word meaning, depending on the identity of the word-initial segment. Now assume that X and Y also differ in the acoustic quality as a function of the prosodic context in speech production. X is realized as A in the prosodic context 'P_1', and as B in another prosodic context 'P_2'. In the same prosodic contexts, Y is realized as C and D, respectively. These realizations are summarized in (1).

(1) Possible realizations of segments in speech production

a. $X \rightarrow A / P_1$

b. $X \rightarrow B / P_2$

c. $Y \rightarrow C / P_1$

d. $Y \rightarrow D / P_2$

In this example, the hypothetical language has two linguistic contrasts: X vs. Y, and P_1 vs. P_2. The first (lexical contrast) differentiates a pair of words (e.g., 'ban' vs. 'pan'), while the second (prosodic contrast) can differentiate a pair of syntactically ambiguous constructions (e.g., the difference in meaning between: [Aaron and Bill] or [Charlie] and [Aaron], and [Bill or Charlie]) (Nespor & Vogel, 1986). The prosodically conditioned acoustic variation suggests that there are four possible contrasts that need to be resolved by the listener in speech perception, as summarized in (2).

(2) Four possible contrasts in speech perception

a. A vs. B (prosodic contrast (P_1 vs. P_2) when segment is X)

b. C vs. D (prosodic contrast (P_1 vs. P_2) when segment is Y)

c. A vs. C (segmental contrast (X vs. Y) when prosodic context is P_1)

d. B vs. D (segmental contrast (X vs. Y) when prosodic context is P_2)

A particular acoustic realization of a segment (e.g., A) conveys both the information about the segmental contrast (is it X or Y?) and about the prosodic context (is it in the prosodic context P_1 or P_2?). When listeners hear a sentence, they understand both what the sentence means and what words it contains, and these computations can occur simultaneously, considering a particular acoustic realization of a segment has information for both computations. In fact, in this set up, listeners might not be able to just process one computation without also processing the other. This is the prediction of the model proposed in Cho et al. (2007) as the computation of the prosodic structure (is it in the prosodic context P_1 or P_2?) and the lexical analysis (is it X or Y?) are argued to occur simultaneously.

In theory, any language with prosodically conditioned variation in acoustic realization of segments (e.g., Selkirk, 1984; Nespor & Vogel, 1986) may potentially exhibit the schenario described in this example. However, previous studies have reported that segments may only optionally be affected by the prosodic context (Jun, 1993, 1994; Han, 2000; Arvaniti & Baltazani, 2005), which would undermine the role of segmental realization in speech perception. For instance, if X is only optionally realized as B in P_2, and can be realized as A in P_2 sometimes, the realization A does not reliably convey information of the prosodic context.

This dissertation investigates Seoul Korean which potentially exhibits the test case scenario. Seoul Korean is particularly interesting as the culminative function of prosody may not be independent from the delimitative function of prosody. Seoul Korean is argued to be an 'Edge-prominence' language, as opposed to a 'Head-prominence' language like English (Jun, 2014). In a 'Head-prominence' language, some of the prosodic constituents are marked with a pitch accent, which is associated with word-level stress, marking the word more salient than others. On the other hand, in an Edge-prominence language, there is no word level prominence. Instead, a salient constituent is the one that is marked with a prosodic constituent edge, hence the name 'Edge-prominence'. Therefore, in an Edge-prominence language, such as Seoul Korean, it is assumed that prominence marking is never acoustically implemented independently from boundary marking. Recent proposals that make different assumptions about intonational phonology of Seoul Korean exist, where it is argued two prosodic words with the same prosodic phrasing may additionally differ in the location of prominence (Hatcher et al., 2024). However, in this dissertation, I follow the view that prominence is always marked with a prosodic constituent edge in Seoul Korean.

In Seoul Korean, the realization of lenis obstruents is conditioned by the prosodic context, and yet it has also been reported to that there is optionality in the prosodic conditioning of the realization of lenis obstruents (i.e., optional voicing of lenis stops) (Jun, 1993, 1994; Han, 2000). This optionality of segmental realization has led researchers to consider segmental realization as a less reliable cue than patterns in the F0 contour in Seoul Korean (Jun, 1994).

This dissertation presents findings that suggest that there may be a greater role for segmental realization in signaling prosodic structure than previously thought. First, contrary to previous findings that the realization of lenis obstruent is optionally affected by the prosodic context (Jun, 1993, 1994; Han, 2000), I report empirical evidence that suggests the realization of lenis obstruents is in fact more systematically affected by the prosodic context than previously reported. Second, contrary to previous assumptions that F0 contours more reliably signal prosodic structure than segmental realization, I report evidence that segmental realization may in fact be more reliable in signaling prosodic structure in certain contexts in Seoul Korean. Finally, these findings that the realization of lenis obstruent is a reliable cue for prosodic structure are supported by the results of a perception experiment, which shows that listeners are more affected by the segmental realization of lenis obstruents than by the F0 contour in making decisions about prosodic structure.

The rest of this chapter proceeds as follows. §1.2 reviews the intonational phonology and laryngeal contrasts of Seoul Korean. This review presents the possible realizations of the laryngeal contrast of lenis vs. aspirated obstruents in Seoul Korean in different prosodic contexts. §1.3 reviews how the decoding of the acoustic signal into prosodic structure and segmental contrasts is tested experimentally in general, and it also presents relevant previous perception experiments on Seoul Korean. §1.4 reviews Cho et al. (2007)'s model of perception. Finally, §1.5 ends this chapter with an overview of the dissertation.

1.2 Seoul Korean

In this section, I present background on Seoul Korean. I start with a review of the model of intonational phonology of Seoul Korean and its prosodic transcription system, K-ToBI (Jun, 1993, 2000, 2006, 2007), which I adopt in this dissertation. I then move on to the three-way laryngeal contrast of Seoul Korean and the phonetic correlates for the laryngeal categories which are dependent on the prosodic position.

1.2.1 K-ToBI: Intonational phonology of Seoul Korean

This dissertation adopts the autosegmental-metrical system of Seoul Korean intonation called the K-ToBI model (Jun, 1993, 2000, 2006, 2007). The full system of Seoul Korean intonation, as proposed in the K-ToBI model, is given in Figure 1.1. In this model, there are three levels of prosodic constituents above the Phonological Word (PW or PWd): the Accentual Phrase (AP), the intermediate phrase (ip), and the Intonational Phrase (IP) (Jun, 1998, 2000, 2006, 2007).

In Seoul Korean, the syllable is the tone-bearing unit in the intonational phonology, and feet are not included in the K-ToBI model (Jun, 2000). The Phonological Word (PW) is defined loosely as a lexical item with optional postpositions such as case markers and sentence-final particles. What constitutes a 'word' in Korean is not well defined in the literature, partially due to the absence of word-level prominence (i.e., lexical stress) in Korean. In the K-ToBI literature, researchers often assume that an orthographic word, or an 'Eojeol' (the sequence of syllables that are demarcated with a space in writing) is a PW in Korean (e.g., Jun, 2000; Jung et al., 2007a). I follow the same assumption in this dissertation and use Eojeols and PWs interchangeably.

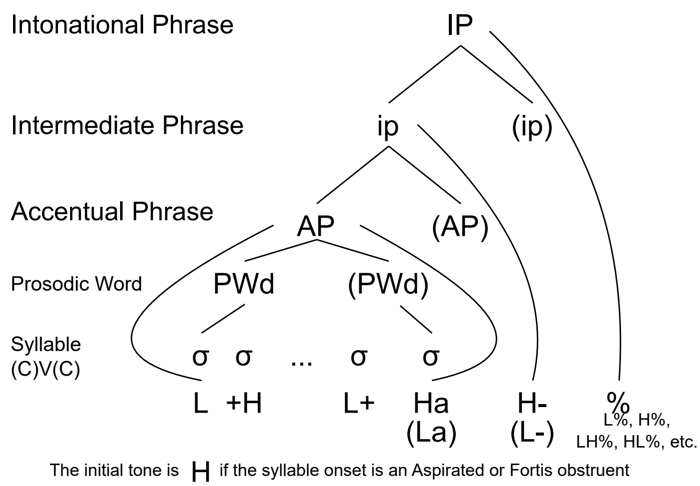


Figure 1.1: The K-ToBI model of Seoul Korean Intonational system (Jun, 1998, 2000, 2006, 2007). 'La' and 'L-' are in parentheses as they are reported to be less frequent than 'Ha' and 'H-', respectively (Jun, 2000, 2007).

Among the prosodic constituents above the PW, the AP has been the focus of past literature because it not only has a distinct tonal melody, but it also seems to be the domain for segmental processes such as lenis obstruent voicing (Jun, 1998). Since the AP is the first level of constituent above the syllable, that is well-studied in terms of its prosodic characteristics, I will discuss the AP in more detail here.

The AP starts with either an 'H(igh)' or an 'L(ow)' tone, depending on its first segment. If the first segment is an aspirated or a fortis obstruent, the AP-initial tone is an 'H', and when the first segment is anything other than an aspirated or fortis obstruent, e.g., a lenis obstruent, it is an 'L'. This is summarized in Table 1.1.

We will come back to this tonal distinction between the laryngeal categories of aspirated vs. lenis obstruents in particular, in §1.2.2. The tones '+H' and 'L+' are

AP-initial Tone	Consonant types	Phones
H	Aspirated	/p^h, t^h, k^h, $ʧ^h$/
	Fortis	/p*, t*, k*, ʧ*/
	Fricatives	/s, s*, h/
L	Lenis	/p, t, k, ʧ/
	Nasals	/n, m/
	Vowels	/a, e, i, o, u, ɨ, .../

Table 1.1: H/L inducing segments

said to be 'loosely associated with the second and the penultimate syllable, respectively. These tones are also optional and may not be realized on the surface if the AP is shorter than four syllables. The last tone of an AP is most frequently 'Ha' though it can also be 'La'. The '-a' is a diacritic that indicates that the tone is associated with the *A*ccentual Phrase edge. Interestingly, while the right edge of the prosodic constituents above the PW in other languages tend to be marked both with a boundary tone and pre-boundary lengthening (Lehiste et al., 1976; Wightman et al., 1992), the AP is argued to be marked solely with the boundary tone, but not with the pre-boundary lengthening (Jun, 1993).

An AP may consist of multiple PWs though Jun (2000) argues that an AP is typically a lexical item plus a post-positional marker, such as the case marker. However, the AP is defined by its acoustic characteristics rather than as a syntactic constituent (e.g., a noun with a case marker), and is argued to be tonally 'demarcated by a phrase-final High tone' (Jun, 1998, 190). Jun reports that the sequence of syllables that has distinct tonal melody (e.g., 'L +H L+ Ha'), i.e., the AP, often cannot be explained as a syntactically defined constituent. Moreover, various segmental processes in Seoul Korean seem to refer to the AP, rather than syntactic constituents. For instance, voiced realizations of lenis obstruent onsets were largely found at the medial (non-initial) position of an AP in Jun's production study Jun (1994). But due to frequent mismatches of the prosodic constituents (the AP) and the syntactic constituents, it was inadequate to characterize the distribution of voiced lenis obstruent onsets using syntactic constituents. Finally, Jun reported that the realization of lenis obstruents was affected by speech rate, that is, there were more instances of voiced lenis obstruents when the speech was faster (Jun, 1994, 1998). It would be hard to explain how prosodic constituents could be directly generated from the syntactic constituents, or to propose that segmental realization is directly conditioned by the syntactic constituents since it is unlikely that the syntactic structure would change by the speech rate.

The ip is the prosodic constituent above the AP level. The ip was added to the K-ToBI model later to account for the pitch reset due to focus and syntactic phrase structure (Jun, 2006). According to Jun (2006), the ip can be detected if the end of an AP has a higher AP-final tone than that of the preceding AP, in which case this higher AP-final tone is labeled as the ip-final tone; or if an AP has a higher initial pitch than that of the preceding AP, in which case this higher initial pitch is labeled as the ip-initial tone. Described this way, detecting the ip boundaries seems to assume that the AP boundaries are first known to the labeler. In other words, this assumes that the AP labeling has been finalized and then the ip is found on the basis of that. For the sake of simplicity, I generally adopt the version of K-ToBI without the ip, and only discuss the ip as a potential alternative prosodic structure, when I present the prosodic analysis of an acoustic signal.

The largest prosodic constituent is the IP, which is marked by boundary tones and final lengthening. There are at least nine attested IP boundary tones (L%, H%, LH%, HL%, LHL%, HLH%, LHLH%, HLHL%, LHLHL%). The boundary tones are used to convey pragmatic and semantic meanings of the utterance. For example, a WH-question can be marked by H%, LH% or HL% (Jun & Oh, 1996). When the right boundary of an AP or ip coincides with an IP right boundary, the AP (and/or ip) final tone is overridden by the IP boundary tone.

In the K-ToBI model, Jun (1993) takes the 'Intonation first approach', which argues that the prosodic constituents are defined by the acoustic characteristics and are independent from, though strongly related to, the syntactic constituents. This suggests that a given syntactic structure is not deterministically realized with a particular prosodic structure.

The K-ToBI model also assumes that the prosodic constituents in Seoul Korean are 'strictly layered' and properly bracketed (e.g., Selkirk et al., 1982; Selkirk, 1986). The Strict Layer Hypothesis states 'a constituent of category-level n in the prosodic hierarchy immediately dominates only constituents at category-level n-1 in the hierarchy' (Selkirk, 1984, p.26). This means that, according to the K-ToBI model, Seoul Korean prosodic constituents do not violate the constraints in (3). The constraint definitions are borrowed from Selkirk & Lee (2015), and they refer to a variable C^i, which stands for prosodic constituent of level i in the prosodic hierarchy. For Seoul Korean, the assumed prosodic hierarchy is: IP/ip/AP/PW/syllable (from the highest, IP, to the lowest level, syllable).

(3) Strict Layer Hypothesis constraints as defined in p. 190 in Selkirk (1996) :

a. Headedness: any C^i must dominate a C^{i-1} except if C^i is a syllable.

b. Non-recursivity: no C^i dominates C^j, where i = j.

c. Layeredness: no C^i dominates C^j, where j > i.

d. Exhaustivity (level-skipping): no C^i immediately dominates C^j, where j < i-1.

The Strict Layer Hypothesis assumes that smaller prosodic constituents are contained inside larger ones. This has been referred to as 'proper bracketing' (Bickel et al., 2009), who writes that 'proper bracketing' refers to the 'expectation that no language will exhibit non-stacking domains in its prosodic structure, e.g. the edge of a phonological phrase cannot be situated inside a phonological word'. These constraints are assumed to be universal for any prosodic hierarchy, though among these constraints, Exhaustivity and Non-recursivity are sometimes argued to be violable in some languages (e.g., Ito & Mester, 2012), and also in Seoul Korean (e.g., Jun, 2011; Lee, 2022; Baek & Yun, 2018). These alternative prosodic structures that violate some of the Strict Layer hypothesis constraints will be discussed further in Chapter 3 and Chapter 4.

Besides tonal labeling, K-ToBI has another component, which is Break Indices (BIs) labeling. The BIs are labeled subjectively based on listener's perception of juncture strength between PWs. Jun (2000) explains that there are 4 levels of BI in the K-ToBI model ranging from 0 to 3. The BI level 3 denotes the strongest prosodic juncture, usually instantiated by a notable pause or 'a strong subjective sense of pause' instantiated by the lengthening of the final syllable (i.e. final lengthening). This BI level is associated with IP boundaries. The BI level 2 corresponds to AP boundaries and is associated with a 'sense of phrase edge' usually instantiated by the tonal pattern at the right edge of the Accentual Phrase. In terms of the K-ToBI labels introduced above, this can be labeled as the phrase final 'Ha' tone (or less frequently as 'La'). Below the BI level 2, the BI level 1 denotes phrase-medial PW boundaries and the BI level 0 denotes cases for monosyllabic functional nouns that do not form a PW on their own (e.g., the nominalizer /kɔt/ or the negation marker /an/). These BI levels would indicate that the syllables that follow these junctures are the initial syllables of the constituents that the BI levels are associated with. For example, the syllable following the BI level 3 would be an IP-initial syllable, and the syllable following the BI level 2 would be an AP-initial syllable, and so on, as summarized in Table 1.2.

The first thing to note is that identifying these BI levels sometimes relies on in terms of 'the subjective sense' of the labelers/native speakers though in other cases they are described in terms of phonetic cues, such as final lengthening or identification of the phrase final 'Ha' tone (Jun, 2000). Marking the tonal labels on the AP bound-

BI level	$[_X\sigma$ (Initial syllable of the prosodic constituent X)
3	$[_{IP}\sigma$
2	$[_{AP}\sigma$
1	$[_{PW}\sigma$
0	$[_{PW}...\sigma$ (PW-medial)

Table 1.2: The BI levels and the associated positions in the prosodic constituents

aries then accompanies BI labels, which provides support for the found boundaries, by identifying the tonal events that are supposed to happen at the junctures, e.g., the boundary tones in the IP-final syllable, or the phrase final tone 'Ha' in the AP-final syllable. Interestingly, the tone labels in fact are indicative of prosodic boundaries, which would make the BIs completely redundant. For example, we would know that there is an AP boundary between two syllables if the second syllable marked with the initial tone 'L' is preceded by a syllable marked with the final tone 'Ha', without even needing to look at the BI tier. The K-ToBI manual mentions that the BI and the tones could mismatch sometimes which should be flagged on the BI tier with the diacritic of 'm' (Jun, 2000), meaning if a 2-like break does not coincide with an AP-like tone, this BI of 2 should be marked as '2m'. Also, the BI tier can have a diacritic of '-' which flags the labeler's uncertainty of the prosodic break strength. A BI of '2-' therefore is implying that the labeler knows that it is a 2-like break because of the tones, but perceives a weaker prosodic break. The fact that the uncertainties are marked on the BI tier, rather than the tonal tier, implies that the tonal pattern is prioritized over the BI. The preference for the tones and the partial redundancy of the BI labels could be one of the reasons why the BIs are 'less emphasized in the ToBI system' (Jun, 2022).

1.2.2 Laryngeal contrasts and their phonetic correlates

Next, we move on to the laryngeal contrast of Seoul Korean. There is a substantial amount of research done on the phonetic difference between these laryngeal categories.

The review in this section will only focus on how the realization of the laryngeal contrast may depend on the prosodic structure.

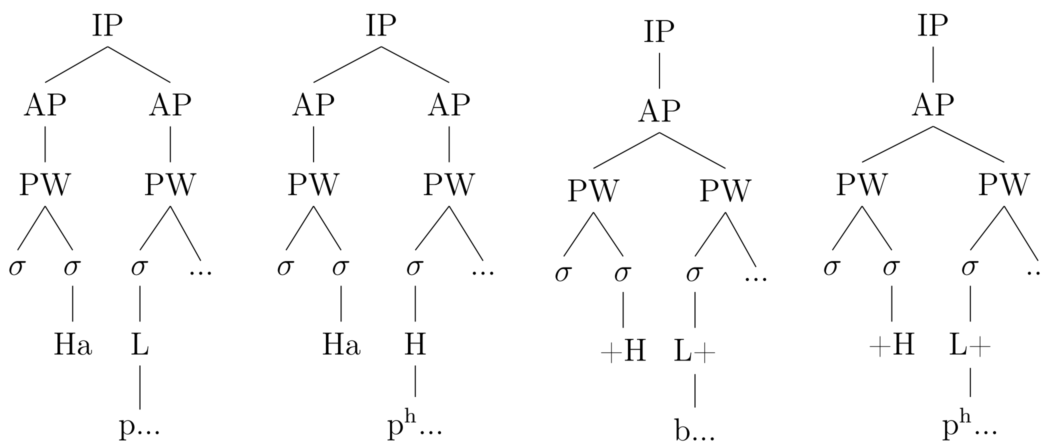
As already mentioned in §1.2.1, Korean has a three-way laryngeal contrast–fortis, aspirated, and lenis–in bilabial, alveolar, post-alveolar, and velar places of articulation. The post-alveolar categories are affricates and the others are stops. Throughout the dissertation, I will use 'obstruents' to refer to affricates and stops, though technically obstruents also include fricatives. I exclude fricatives because this dissertation investigates the lenis and aspirated contrast, but a fricative of lenis category arguably does not exist in Korean (Chang, 2008) because the non-fortis /s/ can be classified as aspirated, as it never becomes fully voiced, unlike other lenis categories.

In previous studies, lenis and aspirated obstruents have been investigated extensively due to a sound change involving them ('the lenis-aspirated VOT merger'), i.e., these two categories were previously distinguished by the difference in VOT, but now are distinguished by the difference in F0 (e.g., Silva, 2006; Beckman et al., 2014; Bang et al., 2018). The F0 difference between these two categories is also reflected in the K-ToBI model reviewed in §1.2.1. The AP starts with an H tone (a higher F0) when it starts with an aspirated obstruent (or a fortis category) and with an L tone when it starts with a lenis obstruent. However, the two categories only seem to differ in F0 in the AP initial position because their F0 difference is not significant in the AP medial position (Lee, 2018; Choi et al., 2020). In the AP medial position, lenis and aspirated obstruents can be differentiated in other ways, since only lenis obstruents, but not aspirated ones, are voiced in this context.

In AP-initial position, fortis obstruents are realized with a short lag VOT and an F0 value intermediate between lenis and aspirated obstruents, in the AP-initial position. AP-initially, the vowel following a fortis obstruent is also significantly longer, compared to AP-medially (Steffman et al., 2022). In AP-medial position, fortis and lenis obstruents are distinguished based on the degree of reduction because fortis obstruents are always realized with full closure, unlike lenis obstruents in the AP-medial position.

1.2.3 Four possible contrasts in speech perception of Seoul Korean

Having reviewed the intonational phonology of Seoul Korean and its laryngeal contrast between lenis and aspirated obstruents, the schematic example mentioned in §1.1 can be revisited as in Figure 1.2. In this figure, Figure 1.2a and Figure 1.2b have the same prosodic structure, but not the same segment in the initial position of the second PW. Figure 1.2a and Figure 1.2c have the same segment in the initial position of the second PW, but they have distinct prosodic structures.



(a) AP-initial Len. (b) AP-initial Asp. (c) AP-medial Len. (d) AP-medial Asp.

Figure 1.2: Four possible contrasts: lenis/aspirated obstruents in the AP-initial/AP-medial position. Lenis and aspirated obstruents are symbolized with /p/ and /p^h/.

The four realizations and contrasts schematized in (1) and (2) can be specified as in (4) and (5). In this particular example, the first PW is assumed to be disyllabic, and therefore the first syllable of the second PW bears an L+ tone, if the PW is AP-medial, as in Figure 1.2c and Figure 1.2d.

(4) Possible realizations of segments in speech production

a. Lenis → [Voiceless, L tone] / AP-initial

b. Aspirated → [Voiceless, H tone] / AP-initial

c. Lenis → [Voiced, L+ tone] / AP-medial

d. Aspirated → [Voiceless, L+ tone] / AP-medial

(5) Four possible contrasts in speech perception

a. [Voiceless, L tone] vs. [Voiced, L+ tone]

AP-initial vs. AP-medial when segment is lenis

b. [Voiceless, H tone] vs. [Voiceless, L+ tone]

AP-initial vs. AP-medial when segment is Aspirated

c. [Voiceless, L tone] vs. [Voiceless H tone]

Lenis vs. aspirated when prosodic context is AP-initial

d. [Voiced, L+ tone] vs. [Voiceless, L+ tone]

Lenis vs. aspirated when prosodic context is AP-medial

As reviewed in this section, the prosodic contrast when the segment is lenis obstruent (Figure 1.2a vs. Figure 1.2c) is acoustically encoded as a difference in the tonal transition between the second and the third syllables: 'Ha L' in Figure 1.2a and '+H L+' in Figure 1.2c. Previously, Lee & Lee (2013) showed that the F0 contours for the two similar sequences of tones, 'L Ha L Ha' and 'L +H L+ Ha', are measurably different. They investigated 400 utterances from a corpus containing speech by voice actors, and labeled the utterances using ToBI labels using a semi-automatized method provided by the Momel program (Hirst, 2007). They compared F0 excursions between the L and H targets in tetrasyllabic APs and two adjacent disyllabic APs. They showed that the tonal excursion between the two APs was significantly larger than the tonal excursion between the second and the third syllables within a tetrasyllabic AP (p.129). While their ToBI labeling system was not exactly the same as the K-ToBI model (Jun, 1993), their results can be interpreted as suggesting a phonetic difference between 'L Ha L Ha' and 'L +H L+ Ha'.

Additionally, these two APs in Figure 1.2a and Figure 1.2c also differ in the realization of the lenis obstruent in the initial position of the second PW (symbolized as /p/ in Figure 1.2a and Figure 1.2c). In the former, it is expected to be voiceless and in the latter, it is expected to be voiced. aspiration.

In the case of the contrast between Figure 1.2c and Figure 1.2d, the expected acoustic difference is the segmental realization of the initial segment of the second PW, as a lenis obstruent would be voiced in this prosodic context, where an aspirated obstruent would be voiceless. In terms of F0, they are not expected to differ, considering previous production experiments (e.g., Lee, 2018; Choi et al., 2020) that reported the F0 difference between laryngeal category types was only significant in the AP-initial position.

Chapter 4 presents a construction in Korean that exhibits the four-way contrast schematized in Figure 1.2, which involves a polysemous word in Korean, /kɨman/. /kɨman/ is an adverb meaning ‘up to that degree/amount’, but it can mean ‘stop’ or ‘start’, depending on how it is prosodically phrased with the following verb. When it is phrased in a separate AP from the following verb (as in Figure 1.2a and Figure 1.2b), it means ‘start’, and when it is prosodically phrased together with the following verb in the same AP (as in Figure 1.2c and Figure 1.2d), it means ‘stop’. See (6) for an example sentence with /kɨman/ followed by a verb starting with a lenis stop /p/, and see (7) for an example sentence with /kɨman/ followed by a verb starting with an aspirated stop /p^{h}/.

(6) /kɨman pul-mjɔn ʧoh-kes*-ta/

up to that degree/amount blow-COND be good-FUT-DECL

a. Grammatical in a context where a child has been blowing something, e.g., a balloon, to the extent that the balloon is about to pop (i.e., ‘stop blowing’)

b. Grammatical in a context where a child has not been blowing balloons, though the child has been asked to do so, and the speaker wants to say that it is about time to blow them (i.e., 'start blowing')

The verb following /kɨman/ in (7) starting with an aspirated stop /pʰ/ forms a minimal pair with the verb in (6). Listeners hearing a sentence with /kɨman/ followed by a verb starting with a lenis or an aspirated obstruent, therefore, can potentially use prosodic cues to disambiguate both the meaning of the sentence ('start' or 'stop') and the lexical identity of the verb, by identifying the verb-initial phoneme (lenis or aspirated).

(7) /kɨman pʰul-mjɔn ʧoh-kes*-ta/

up to that degree/amount solve-COND be good-FUT-DECL

a. Grammatical in a context where a child has been solving something, e.g., math problems, until very late (i.e., 'stop solving')

b. Grammatical in a context where a child has not been solving math problems, though the child has been asked to do so, and the speaker wants to say that it is about time to solve them (i.e., 'start solving')

For more details about this construction, see Chapter 4.

1.3 Perception of prosodic constituents

Thus far, I have reviewed how prosodic structure is implemented in the acoustic signal, focusing specifically on the implementation of the prosodic boundaries in Seoul Korean. In this section, I review the experimental work in the literature that has investigated how these cues affect speech perception.

I categorize the experimental work into four kinds: prosodic transcriptions (§1.3.1), word-spotting experiments (§1.3.2), prosodic disambiguation experiments (§1.3.3), and post-boundary segment identification experiments (§1.3.4). These experiments assume that prosodic boundaries are acoustically implemented and differences in prosodic structures are reflected in phonetic cues. They test how such cues are perceived and interpreted by listeners, and what the consequences of such interpretations are, in terms of lexical identification or sentence processing.

1.3.1 Prosodic transcriptions

I start with prosodic transcription, which is arguably the most direct way to test the role of acoustic cues in the perception of prosodic boundaries. The prosodic transcription process usually takes place by hiring multiple transcribers who are native speakers of the language and trained in using the convention (e.g., K-ToBI) and asking them to perform the transcription. The data collected from the transcribers are then analyzed in terms of the inter-transcriber agreement. If the agreement rate is higher than some conventional threshold, then it is deemed appropriate to use the transcriptions for further analysis. The prosodic structure of the data is then determined by the collection of transcriptions, and analyzed in terms of its acoustic properties (e.g., duration, F0, and post-segmental realization). In this very sense, the prosodic transcriptions are perceptual experiments on their own, testing the reliability of the acoustic cues in signaling the prosodic boundaries. I will briefly discuss some of the issues with the prosodic transcription process in this section.

First, prosodic transcription relies on subjective impressions of transcribers, and two transcription labels may not always be acoustically distinct. The acoustic cues for some utterances are clearer for the transcribers so they might unanimously agree on the label, but some may cause disagreements if they are ambiguous. For example, an AP juncture may be realized with a phrase final Ha tone, but not followed by a segmental strength that is expected (or vice versa). These cases may be analyzed post hoc by analyzing the acoustic measurements. However, other than the post hoc analysis, it will for instance remain a mystery why the transcribers might label a prosodic juncture with a BI of 2 rather than 3.

The post hoc analysis is also limited by the assumption that the researcher has measured the right phonetic properties. It is a common problem in phonetic research that researchers do not know what to measure to compare two contrastive phonological categories. I address this concern in Chapter 2, where I show that the previously conceived optionality of the prosodic conditioning of lenis obstruent voicing may be because there are other prosodic properties that are more reliably conditioned by the prosodic structure, that have been overlooked in previous studies. In other words, some of the acoustic boundaries where acoustic cues seem to mismatch, might not be analyzed as a mismatch, depending on how the acoustic cues are parameterized.

To make the grounds for transcriptional choices more transparent, Brugos et al. (2018) and Brugos et al. (2023) have proposed to explicitly label cues for prosodic labels. They report that the BI labels sometimes do not have all of the expected set of cues in the transcription of American English. To use the Seoul Korean AP-level juncture as an example, this can occur if only the Ha tone is identified but this syllable is followed by a voiced lenis obstruent contra expectation. This kind of cue-mismatch is expected to occur in Seoul Korean prosodic transcription due to the tone-first tendency where the prosodic constituent is defined by the tonal sequences, and the segmental realization is analyzed as processes that refer to the tonally-defined constituency. In the case of cue mismatch, i.e. when there is tonal evidence of boundary but no corresponding segmental realization expected at the boundary, it is often treated as an exception of the segmental process, where the segmental process is described as being 'optional'. However, if the transcribers are instructed to label the prosodic boundary primarily based on the tonal evidence, it is evident that the tonal process will not be 'optional' by definition. This may not be what listeners do in a naturalistic setting. For instance, when there is tonal evidence of an AP juncture followed by a voiced lenis obstruent (which contradicts the tonal evidence), it remains to be tested whether listeners process the signal as implementing an AP-level boundary or not, or how listeners might process such signal.

Moreover, once we get the categorical labels from the transcribers, we immediately lose the information about how much such categorical labels are supported by the actual data. However, in reality, the acoustic signal might only probabilistically indicate one label versus another.

Finally, the transcribers might have a strong bias coming from non-acoustic structures in the signal. This stems from two facts: first, they tend to be native speakers of the language, and second, they tend to work with orthographically transcribed utterances that are visually provided to them. In other words, reading the sentences, they might already have parsed the sentences syntactically, which would likely affect them so that they would find acoustic evidence for the boundaries that they expect to find based on the syntactic parse (Buxó-Lugo & Watson, 2016). Note that this is contrary to the proposal that the prosodic structure is independent of the syntactic structure (Selkirk, 1984; Nespor & Vogel, 1986) and that it is the prosodic boundaries that are acoustically implemented. It is known that listeners have expectations for upcoming syntactic structures before those words become available to them, and this might work as a Bayesian prior for the perception of prosodic boundaries (Buxó-Lugo & Watson, 2016).

To summarize, the prosodic transcription process is first 'mysterious' in the sense that we only know post hoc why a signal is labeled as such by measuring the phonetic properties. This is limited by our assumption of the phonetic properties that are manipulated in implementing the prosodic structure. Certain transcription conventions may prefer to use one phonetic property (tonal cues) over others (post-boundary segmental realization), which may not reflect what the listeners do in naturalistic settings, and potentially lead to undermining the other cues as being reliable or informative. The categorical labels also often miss the gradience or uncertainty involved in the perception of prosodic structures. Finally, the transcribers are heavily influenced by the other structures in the signal, and therefore their judgments cannot be said to be true reflections of what the phonetic properties do in the perception of the prosodic structure.

The Rapid Prosodic Transcription (RPT) experiment approach (Cole et al., 2010) addresses some of these concerns. In an RPT experiment, instead of prosodic transcribers who are trained in a particular transcription system, naive listeners participate in making the judgment of the prosodic structure. The listeners are asked to mark the prominent word and the boundaries between words. The judgments from multiple listeners are aggregated such that words are assigned a 'p-score' for prominence and a 'b-score' for the boundary, which are the proportions of the times that the participants chose the word as having a prominent or preceding a boundary. This approach addresses the concerns I've raised in that the labelers do not have a pre-imposed bias for prioritizing the tonal cues over the segmental cues in making the prosodic transcriptions, and the gradience in the signal can be captured by the transcription (p/b-scores).

However, it still does not directly model what the listener does in auditory sentence processing. The participants, though naive in prosodic transcriptions, but skilled in the language itself, would still be affected by the syntactic and semantic knowledge. Therefore their transcriptions may not be strictly modeling the salient 'acoustic' boundaries of the utterances. In most settings, participants are also given written transcriptions of the spoken materials. This would likely bias their perception of the boundary since most participants would know the orthographic rules involving punctuation marks.

1.3.2 Word-spotting experiments

Next, I will review previous work on the listener's use of acoustic cues in word-spotting experiments. In these experiments, listeners are auditorily presented with sequences of syllables, and they are asked to make a response when they hear the target word. The target sequence that is meant to exhibit the prosodic structure of interest is either phonetically modified or spliced from a different utterance where the target sequence is produced in the desired prosodic position.

Christophe et al. (2004) conducted a series of lexical access experiments showing that listeners make use of prosodic boundary information in resolving local/temporary ambiguities in a sequence such as *chat grincheux* (*lit.* 'cat grumpy'), which might be temporarily ambiguous with *chagrin* ('sorrow'). In their lexical access experiments, participants were slower in reacting to the first word of the sequence ('chat') when there was a possibility that the first word might form another word (e.g., 'chagrin') with the following word, compared to when there was no such possibility (e.g., 'chat drogué'). This delay disappeared when 'chat' was realized in a phonological phrase-final position.

Similarly, Salverda et al. (2003) found listeners look longer at 'cap' and 'ham' when listening to 'captain' and 'hamster' in which their first syllable is replaced by 'cap' and 'ham' spoken in isolation. In both experiments, when the target syllable ('chat' or 'cap'/'ham') was produced with a longer duration (either produced in the phrase-final position or in isolation), it activated the lexical item and deactivated the alternatives. This was measured as a faster response time to the first word ('chat') in Christophe et al. (2004) and a longer fixation time on the first word ('cap'/'ham') in Salverda et al. (2003).

Christophe et al. (2004) found that the phrase-final realization exhibited pre-boundary lengthening and a different pitch pattern (falling, as opposed to phrase-medial rising). One way to interpret their result is that the lexical search was modulated by the prosodic boundary, implemented with lengthened duration and a falling pitch in their experiment, such that in the lexical competition process, the word straddling such a strong phrase-level boundary would not be activated (i.e. ‘chatgrin’ would not be compatible with the acoustic signal that is compatible with ‘chat # grin’ with a strong prosodic boundary).

Using the same logic, subsequent work has tested the role of other acoustic cues in implementing the prosodic structure and therefore in word-spotting tasks. For instance, Cho et al. (2007) tested the role of Domain-initial strengthening in lexical activation and Kim and collaborators tested the role of tonal sequences (2004, 2009, 2012).

Cho et al. (2007) conducted cross-modal identity-priming experiments to test the role of segmental strength in spoken-word recognition. Listeners heard sentences containing two-word sequences containing temporary lexical ambiguities (e.g., ‘bus tickets’ temporarily ambiguous with ‘bust’). When the second word-initial syllable was spliced in from an Intonational Phrase initial position, i.e. the first segment was realized as stronger, listeners were faster in making the lexical decision to the written letters ‘bus’ upon hearing ‘bus’ (their Experiment 2), compared to when the segment was spliced in from a word-initial (and IP-medial) position, i.e. the first segment of the second word was not as strong. The authors argued that the strong articulation of the initial segment of the second word (/t/) created a percept of a large prosodic boundary, such that the listener was not able to consider the word ‘bust’ as a competitor for ‘bus’. In other words, when the /t/ was spliced from the IP-initial position, the acoustic signal was likely interpreted as ‘bus][t’, which was incompatible with a word like ‘bust’. In the absence of strong competitors like /bust/, the lexical activation of ‘bus’ might be facilitated, resulting in a faster reaction time in the lexical decision task.

Following Cho et al. (2007), SH Kim and her collaborators have conducted a series of experiments that tested Korean listeners' use of tonal cues in various word segmentation experiments. Kim & Cho (2009) tested in their first experiment whether listeners pay more attention to the initial rise (#LH) or to the fall across the boundary (H#L) in identifying an attested word from a sequence of syllables surrounding the target word (e.g., /pa.mɔ.ɾo.*mu.ke*.nɔ.se/, where /mu.ke/ is the attested target word 'weight'). The strings of syllables were obtained from a natural production of the syllables, where there would not be any additional pre-boundary lengthening cue. The pitch of the syllables was manipulated in Praat. Participants were asked to spot an attested word in Korean, and as soon as they found it, they were asked to press a button and recite the word they found.

They were interested in investigating what the tonal cues are for the prosodic boundary. Kim (2004) found from a K-ToBI transcribed corpus that 88% of the transcribed APs ended with a 'Ha' tone, and 85% of the APs that did not start with an Aspirated/Tense consonant started with an 'L' tone followed by a '+H' tone. Put differently, there were two possible tonal transitions that listeners might be sensitive to: the fact that a tonal fall (Ha][L) frequently signals an end of a phrase or the fact that a tonal rise (L +H) frequently starts a phrase. Their key finding from the first experiment was that participants were more accurate in spotting the attested words when the pre-boundary tone was an 'H' *and* when the target word's initial tone was an 'L'. Crucially, they found that as long as the pre-boundary tone was an 'H' and the target word's initial tone was an 'L', there was no difference between what followed the target word's initial tone (i.e. the second tone in the AP), though at least numerically the trend was that participants were better at spotting a real word when the tonal transition was 'H#LH', meaning that the participants were most sensitive to the tonal transition across a boundary (i.e. Ha][L).

From this result, the authors claimed that a tonal fall from a phrase-final 'H' tone to a phrase-initial 'L' was what the listeners paid attention to in their lexical word spotting task. Their results are interesting because if listeners were simply paying attention to how the pitch changed, perhaps they would be good at spotting the word bearing 'HH' tones when the preceding and the following syllables all had 'L' tones (e.g., LLL#HH#LL). On the contrary, their results suggest that participants in their experiment were sensitive to the fact that the 'H' tone was not allowed as an initial tone of a word that did not start with a fortis or aspirated segment. The tone following the initial 'L' did not matter because once the fall from H to L signaled a boundary, the lexical search would be initiated and there was no need to prosodically mark the end of the word (e.g. with another H), because the stimuli in their experiment did not have any cases with a longer word containing a shorter word causing a temporary ambiguity (e.g., 'cap' in 'captain').

In their second experiment, they tested the tonal transition cue in Seoul Korean along with the presence or absence of a final lengthening (the syllable bearing the pre-boundary tone). The task in their perception experiment was the same as their first experiment. In their stimuli, syllables either had a L(ow) or a H(igh) tone, and had short or long duration, which were fixed values. Word initial consonants were spliced from the AP-medial position, and they were lenis or nasal stops, which are L-tone inducing segments (Jun, 1993). They measured mean error rate for identifying the target word correctly within 1500 ms to quantify how much word spotting was facilitated in each condition.

Here I review the results for four of their stimuli conditions in Kim & Cho (2009)'s second experiment. The four conditions are summarized in Table 1.3. They varied the tone and duration of the syllable preceding the target word (σ_0): L vs. H and Short vs. Long, resulting in four conditions. They found that pre-boundary lengthening facilitated word spotting (Condition A $<$ Condition B), but not when the word to

Cues	σ_0	$\# \sigma_1$	σ_2	Interpretation
Condition A: No Lengthening + LLH				
Duration	Short	Short	Short	
F0	L	L	H	$\sigma_0]_{PW}[\sigma_1\sigma_2$
Segment	-	AP-medial	-	$\sigma_0]_{PW}[\sigma_1\sigma_2$
Condition B: Lengthening + LLH				
Duration	Long	Short	Short	$\sigma_0]_{IP}[\sigma_1\sigma_2$
F0	L	L	H	
Segment	-	AP-medial	-	$\sigma_0]_{PW}[\sigma_1\sigma_2$
Condition C: Lengthening + HLH				
Duration	Long	Short	Short	$\sigma_0]_{IP}[\sigma_1\sigma_2$
F0	H	L	H	
Segment	-	AP-medial	-	$\sigma_0]_{PW}[\sigma_1\sigma_2$
Condition D: No lengthening + HLH				
Duration	Short	Short	Short	
F0	H	L	H	$\sigma_0]_{AP}[\sigma_1\sigma_2$
Segment	-	AP-medial	-	$\sigma_0]_{PW}[\sigma_1\sigma_2$

Table 1.3: Stimuli in Kim & Cho (2009). Word spotting was facilitated in the order of A < B = C < D, where 'X < Y' indicates that word spotting was facilitated more in Condition Y than Condition X.

be spotted had the tone 'L H' and the preceding syllable had an H tone (Condition C < Condition D). Word spotting was similarly facilitated in the two pre-boundary lengthened conditions (Condition A < Condition B = Condition C).

Below, I will walk through each of their conditions and explain how listeners might interpret the acoustic signal. They argued that 'when the computed prosodic boundary based on suprasegmental information is matched with prosodically-driven segmental information for that boundary, lexical segmentation is facilitated, whereas a mismatch between them is predicted to hinder lexical segmentation' (Kim & Cho, 2009, 3383). They suggested that different segmentation cues in the acoustic signal may be evaluated with differnt weights. Their results are interpretable if we assume that a long syllable is always decoded as an IP boundary, and in the absence of a long syllable, an HL sequence is always decoded as an AP boundary.

To start with Condition A, the short duration was decoded as signaling that there was no IP boundary, but both AP and PW boundaries were possible, since AP-final syllables are argued to exhibit no final lengthening (Jun, 1993). In Table 1.3, this is demonstrated by an empty cell in the 'Interpretation' column, for the 'Duration' row. Next, the F0 sequence is interpreted to signal no AP boundary because the sequence is L#L. In their first experiment, which varied the tonal sequences without varying pre-boundary lengthening, Kim & Cho (2009) found that the tonal sequence 'L#L' did not provide a strong lexical segmentation cue compared to 'H#L'. This finding was supported by the fact that APs often end with an Ha tone rather than an La tone Kim (2004). Finally, the segmental realization of the onset of the target word was from an AP-medial PW-initial syllable. The resulting phonological representation would be a PW boundary before the target word. This explains why Condition A was the condition where word spotting was the least facilitated, meaning the mean error rate for spotting the attested word was the highest.

In Condition B and C, the long duration was decoded as signaling an IP boundary. The L/H tones on the preboundary syllable were likely decoded as boundary tones since both L% and H% are attested boundary tones in Seoul Korean. As the prosodic boundary in these two conditions was larger than the one in Condition A, word spotting was more facilitated in these two conditions. However, as the segmental realization of the onset of the target word (onset of σ_1) mismatches with the computed phonological representation $(\sigma_0]_{IP}[\sigma_1\sigma_2)$ because the onset was spliced from an AP-medial position. Due to this mismatch, word-spotting was hindered in comparison to Condition D.

Finally, in Condition D, the short duration was decoded as signaling an absence of an IP boundary. The tonal sequence H#LH was decoded as signaling an AP boundary. Word spotting was the most facilitated in this condition, despite the fact that this AP boundary likely mismatched with the AP-medial realization of the onset of the target word. Kim & Cho (2009) argued that this was because segments could sometimes be realized without AP-initial strengthening (i.e., voiced lenis stop in AP-initial position) (Jun, 1993; Cho & Keating, 2001). However, it remains to be tested how listeners would perform when the segment was spliced from an AP-initial position, and therefore there was absolutely no mismatch. If AP-initial voicing was truly tolerated without a processing cost, as if there was no mismatch, there would be no difference in listeners' performance between two realizations of lenis stops. All in all, their result suggested that the mismatch of an AP boundary and an AP-medial segment realization was evaluated as introducing less hindrance to word spotting compared to the mismatch of an IP boundary and an AP-medial segment realization.

In addition to these four conditions, they also tested conditions where the initial tone of the target word was 'H' (not in Table 1.3). Recall that the onset of the target word was always a lenis or nasal stop, which are L tone inducing segments, and did not mismatch with the phonemic category of the onset of target word. Word spotting was hindered to a greater degree when the post-boundary tone mismatched with the expected 'L' tone. This mismatch was evaluated by listeners as being more critical than a more minor mismatch like AP-medial realization in the IP-initial position (Condition B, C).

Hypothetically, the relevant acoustic cues might simply accumulate to cause a perception of a boundary of some size, rather than signal prosodic representations compatible with the cues. For instance, if we hypothesize that the pre-boundary lengthening is equivalent in creating a percept of a prosodic boundary of some size, arbitrarily quantifiable as '2', and the tonal fall as '1', and finally the weak segmental size as '0'. We might also hypothesize that the word spotting performance is better when the boundary size is greater. When the tonal transition is something else, the participants were faster in the condition with the lengthened pre-boundary syllable, suggesting that the lengthening increases the boundary size, and therefore improves the word-spotting performance.

Under this hypothesis, however, the 'lengthening & H][LH' condition (Condition C) would yield a prosodic boundary size of 3 (2+1+0), and the 'no lengthening & H][LH' (Condition D) would yield a prosodic boundary of 1 (0+1+0). The fact that the participants were better in the latter condition suggests that the segmental information does not simply add to the boundary size, but needs to be interpreted based on the 'currently being considered' prosodic structure, which is computed with the durational and tonal cues. Kim and Cho write '[w]hen the computed prosodic boundary based on suprasegmental information is matched with prosodically-driven segmental information for that boundary, lexical segmentation is facilitated, whereas a mismatch between them is predicted to hinder lexical segmentation' (Kim & Cho, 2009, 3383).

However, this seems to conflict with the results of (Cho et al., 2007) since in their stimuli, the IP-initial /t/ would have also caused a mismatch with the syllables that are not pre-boundary lengthened. Therefore, it seems that how phonetic cues are integrated and interpreted given other cues in online speech processing needs further research. It is possible that a lengthened syllable is necessarily interpreted as an IP boundary whereas a non-lengthened syllable duration may be re-interpreted as an IP boundary once the segmental strength information becomes available.

In subsequent work, Kim et al. (2012) tested whether this tonal transition pattern of phrase-final H tone falling to a phrase-initial L was universally salient, or if it was something that Korean listeners were particularly sensitive to. They showed that while Korean listeners used the final F0 cue in their first exposure to learning Artificial language, Dutch listeners did not use the same cue. Interestingly, in their experiments, unlike the tonal transition, the pre-boundary lengthening cue was utilized by both Korean and Dutch speakers at the first exposure, adding further support to the hypothesis that the pre-boundary lengthening might be universally salient. Their results show that some acoustic cues for prosodic boundaries are universal (i.e., pre-boundary lengthening) while others need to be learned (i.e., the tonal fall).

Finally, Tremblay et al. (2019) showed that in an artificial language learning experiment, when the phrase-initial 'L' tone was manipulated such that it was not as canonically low, the participants' segmentation performance improved when the slope of the fall was steeper in word segmentation. Their results highlighted an important aspect of what listeners are doing that is quite different from the lab-controlled stimuli they were tested on in these experiments. In these experiments, the syllables either had an 'L' or 'H' tones, where the F0 values were fixed across stimuli. On the contrary, in actual speech, interpreting a given F0 value as a realization of a phonological 'L' or 'H' requires a complicated computation of what the speaker's range of F0 might be, what the global F0 declination is, and so on.

To summarize, the experiments I have reviewed here share the fact that they tested phonetic properties that cause a perception of prosodic boundary, and they have shown that word segmentation or spotting in online speech perception is either facilitated or inhibited by the perceived prosodic structure. In particular, listeners seem to compute the prosodic structure for the utterance using the acoustic cues and use the prosodic structure to either boost or penalize competitors in the lexical activation process. Some of these cues need to be learned from the native language, while others seem to be universal. How segmental cues are interpreted based on suprasegmental cues (tone and duration) is yet to be investigated.

1.3.3 Prosodic disambiguation experiments

Another popular type of experiment that is used to show the role of phonetic cues in prosodic boundary perception is what I call 'prosodic disambiguation of syntactic ambiguity' experiments (or 'prosodic disambiguation experiments' for short). In these experiments, the listeners are given a pair of sentences that are identical in string but can be prosodically disambiguated, for instance as in the difference in meaning between '[Aaron and Bill], or [Charlie]' and '[Aaron], and [Bill or Charlie]'.

In this section, I focus on the experiments that test the role of segmental realization in causing the prosodic boundary perception.

Scott & Cutler (1984) showed that American English listeners who produce a /ɾ/ across a word boundary in their production, can use this in parsing ambiguous sentences such as 'the day (that) we met Ann, (the day) was beautiful' (vs. the day we met, Ann was beautiful'). They found that listeners pay attention to the segmental realization, in particular, even in the potential presence of other prosodic cues to suggest an alternative prosodic structure. Listeners were more likely to assume 'beautiful' modified 'the day', rather than 'Ann' when the /t/ at the end of 'met' was produced as a flap. Their results suggest that a segmental realization that signals an absence of a prosodic boundary can potentially be a strong enough cue that can revert other prosodic cues for the presence of a prosodic boundary, such as pre-boundary lengthening.

More recently, Mitterer et al. (2021) investigated the role of the glottal stop in Maltese in parsing sentences where a conjunction word (/u/ 'and') was ambiguous between early and late closure: Noun1 or Noun2 and Noun3 ambiguous between '(Noun1) or (Noun2 and Noun3)' (early closure) and '(Noun1 or Noun2) and (Noun3)' (late closure).

Mitterer et al. (2021) manipulated two factors in their stimuli: the duration of Noun1/Noun2 and the presence or absence of a glottal stop before the word /u/ (/u/ or ʔu). Lengthening of Noun1 and an absence of a glottal stop would increase the likelihood of the early closure reading, while lengthening of Noun2 and the presence of a glottal stop would increase the likelihood of the late closure reading. Their results indicated that the glottal stop insertion always increased the likelihood of the late closure reading, and there was a smaller effect, in the predicted direction, of manipulating the duration of Noun1/Noun2. To make sure the effect of glottalization was robust, in their second experiment, they reversed the location of the conjunction word: '(Noun1) and (Noun2 or Noun3)' and '(Noun1 and Noun2) or (Noun3)', because, in the first experiment, the glottal stop insertion reinforced the late closure reading, which was hypothesized to be more preferred (Frazier & Rayner, 1982). They also exaggerated the pre-boundary lengthening manipulation in their second experiment, as their previous duration manipulation did not strongly affect the listeners in their first experiment. Their results showed that there were significant effects of duration and glottalization, but no interaction of the two. From their results, the authors argued that the prosodic structure was phonetically realized by the glottalization, which was argued to be segmental, rather than suprasegmental, since the glottal stop is a phoneme in Maltese. Maltese listeners seemed to process the phonetic signal of glottalization (which was inserted by lowering F0 and amplitude) as indicating a prosodic juncture, rather than as evidence that there was a glottal stop-initial word.

While the phonetic signal of glottalization is processed as evidence of a prosodic juncture, it does not seem to completely override the durational cue. For example, in Experiment 1, when Noun1 was lengthened and there was a glottal insertion on /u/ ('and'), the prosodic cues conflicted with each other: the duration cue suggested the parsing '(Noun1) or (Noun2 and Noun3) (i.e. early closure)', whereas the glottalization cue suggested the parsing '(Noun1 or Noun2) and (Noun3) (i.e. late closure)'. Comparing this with the stimuli where there was no glottal insertion on /u/, it wasn't the case that the absence of a glottal insertion completely made the participants choose the early closure reading; it only slightly decreased the proportion of choosing the late closure reading. A similar pattern emerged in Experiment 2. Again it wasn't the case that the participants chose the early closure reading completely in the case of glottal insertion, reverting the duration cue, but it only decreased the proportion of choosing the late closure reading, compared to the stimuli with no glottal insertion.

Put differently, these results suggest that the segmental realization that mismatches the prosodic structure that listeners construct based on other phonetic cues can assign some penalty to the prosodic structure, such that an alternative prosodic structure(s) that is viable gets a boost in the activation, relative to the one that otherwise matches with the other phonetic cues. This is in line with the discussion of Kim & Cho (2009)'s results in §1.3.2.

For instance, in Experiment 2, participants would consider the two possible parses: early vs. late closure, even before hearing the stimuli 'Noun1 ʔu Noun2 or Noun3', as they were instructed that the sentence they would hear could be interpreted in two ways. This prior awareness of syntactic ambiguity seems to be a common practice in prosodic disambiguation experiments. Among the two candidates, as mentioned, perhaps the listeners have a higher prior weight on the late closure reading than the early closure reading. Once they hear the glottal insertion on /ʔu/, they might first consider whether this was a lexical or an epenthetic glottal stop, but given that they knew what words they would hear, maybe they would skip this stage and just process it as an epenthetic one, suggesting a prosodic break. In this case, they might assign a penalty for the late closure candidate. However, as they hear the lengthening on Noun2, either the late closure candidate gets a reward, or the early closure candidate gets a critical amount of penalty. Notice that when they hear the stimuli without the glottal insertion, there won't be any initial penalty for the late closure candidate, and therefore the total probability for the late closure candidate would be higher than when there was a glottal insertion.

In Seoul Korean, previous work has shown that the acoustic realization of lenis obstruents in different prosodic positions can affect the prosodic parsing decisions. Yoo (2020) tested the prosodic parsing of string identical sentences involving ambiguity due to the silent case marker. In Korean, both the nominative (/ka/) and genitive morpheme (/e/) can be omitted, and therefore /ɔm.ma/ (‘Mom’) is compatible with both ‘Mom-NOM’ and ‘Mom-GEN’. While the noun after a genitive suffix tends to be phrased together with the possessor (e.g., mom), the noun after a nominative suffix tends to start a new AP. If this noun starts with a lenis obstruent, then that obstruent will be realized differently depending on the syntactic parse: voiceless if AP-initial, therefore nominative parsing, and voiced if AP-medial, therefore genitive parsing. Yoo showed a voiced realization of a lenis obstruent, which was cross-spliced from the AP-medial position (i.e. Genetive) into the AP-initial position (i.e. Nominative), increases the likelihood of the genitive reading. However, crucially, Yoo used stimuli that are artificially monotonized. Put differently, when there is no other acoustic information, the segmental realization seems to affect the syntactic parsing decision by providing evidence for the prosodic structure.

Yun & Lee (2022) tested how listeners disambiguated a syntactically ambiguous question in Korean. While they did not directly test the role of segmental realization in prosodic disambiguation, they argued that their results were interpretable if they assumed listeners were sensitive to how segmental realization was conditioned on prosodic structure.

Since WH-words in Korean are homophonous with indefinite pronouns (e.g., who and anyone), a WH-word followed by a verb may be ambiguous between a WH question and a Yes-No question (Yu Cho, 1990). Jun & Oh (1996) argued that the two question types differed in prosodic phrasing of the WH-word and the subsequent verb, though they also differed in the boundary tone type: ‘H%’ for Yes-No and ‘LH%’ for WH. In a Yes-No question, the WH-word and the verb are accentual phrased separately, whereas in a WH question, they are accentual phrased together.

Yun & Lee (2022) predicted that the two question types may be similar to each other in terms of F0 when the WH-word is disyllabic and the subsequent verb is disyllabic as well, since two disyllabic APs and a single tetrasyllabic AP have similar tonal targets: L Ha L Ha vs. L +H L+ Ha. As verb-final syllables in both question types are sentence-final, those syllables carry the IP-final boundary tones, which override the AP-final 'Ha'. They tested their predictions in a production study, which replicated that the most frequent boundary tone for a Yes-No question is 'H%', while 'LH%' is most common for a WH question. In addition, their novel finding was that in a WH question, the penultimate L+ tone was not realized. The typical tonal contours for the sequence of disyllabic WH word followed by a disyllabic verb in their production experiment were therefore: L Ha L H% for Yes-No question and L +H ... LH% for a WH question, as in Table 1.4.

Question Type	WH-word		Verb	
Yes-No	L	Ha	L	H%
WH	L	+H	...	LH%

Table 1.4: Typical tonal sequences for two question types reported in Yun & Lee (2022)

Yun & Lee (2022) tested how listeners used the F0 patterns in prosodic disambiguation in a perception study. They parameterized the F0 contours of the two question types as linear interpolation between targets. In particular, they tested three parts of the F0 contours: F0 height of the H target on the second syllable of the WH-word, presence or absence of the L target on the penultimate syllable, and the boundary tone type on the last syllable. They created 8 F0 contours with an original utterance produced as a WH question by varying three binary factors, and created 8 counterparts with an original utterance produced as a Yes-No question. They summarized their findings as follows: (i) the boundary tone type affected the interpretation of questions, (ii) the absence of the post-WH L tone decreased the chance of interpreting Yes-No question as a Yes-No question, while the presence of the post-WH L tone did not decrease the chance of interpreting WH question as a WH question, and finally (iii) changing the F0 height of the H target on the second syllable did not affect the interpretation, when it was not combined with other factors like the boundary tone type.

Yun & Lee (2022)'s second finding is directly relevant for the current discussion, as it suggests a role for segmental realization in decoding the acoustic signal into prosodic structure. Deleting the post-WH L tone from a Yes-No question created an F0 contour that signalled an absence of an AP boundary before the verb, while the verb-initial segment remained to signal a presence of an AP boundary. In this condition, listeners were more likely to process the stimuli as a WH question, where there is no AP boundary before the verb.

On the other hand, inserting the post-WH L tone to a WH question created an F0 contour that conveyed a presence of an AP boundary before the verb, while the verb-initial segment signalled an absence of an AP boundary. In this condition, listeners were not affected by the F0 manipulation and interpreted the stimuli as a WH question, where there is no AP boundary before the verb.

Yun & Lee (2022) argued that the reason for the asymmetry between the role of the post-WH L tone might be due to an asymmetry in how listeners interpreted the verb-initial segment in the prosodic context. Since the segments were not manipulated in their stimuli, a Yes-No question audio had an AP-initial realization of verb-initial consonant, while a WH question audio had an AP-medial realization of verb-initial consonant. In their experiment, verbs always started with either a nasal or a lenis alveolar stop, both of which have AP-conditioned acoustic realization: denasalized vs. nasal and voiceless vs. lenited. They argued that their results indicate that AP-initial realization of consonants does not necessarily create the percept of an AP boundary, in the absence of an accompanying F0 cue (i.e., deleting L tone in an Yes-No question) while AP-medial realization of consonants may always create a percept of an absence of an AP boundary, even in the presence of an accompanying F0 cue (i.e., inserting L tone in an WH question). They further argued that this was expected given that lenis obstruents can optionally remain voiceless in the AP-medial position, while 'voicing in the AP-initial position is strictly prohibited in Korean' (p. 42).

These experiments all suggest that the segmental realization has an important role, either directly in building the prosodic structure or evaluating the multiple prosodic structures that are being considered, in online speech perception. Again, it remains to be investigated how exactly the segmental realization gets integrated into speech perception, how the prosodic structures are built, what kinds of structures can be considered in parallel in online processing, and how the prosodic structures map onto higher-order linguistic structures (e.g., syntactic parses).

1.3.4 Post-boundary segment identification experiments

So far, I've discussed the role of segmental realization in the perception of prosodic boundary, which modulates the lexical activation (§1.3.2) and the disambiguation of ambiguous syntactic structures (§1.3.3). In this section, I review yet another type of experiments, which tests the perceptual role of the prosodic boundary on the identification of the segment that follows the prosodic boundary. I will review two kinds of experiments which I will call the 'contrast-shift' experiment (Kim & Cho, 2013; Steffman, 2019; Steffman et al., 2022) and the 'phonological inferencing' experiment (Kuzla et al., 2010; Kim et al., 2018).

First, in the contrast-shift experiments, the participants were presented with a segment (e.g., a stop) manipulated along some phonetic dimension (e.g., VOT) and asked to make a phoneme category judgment, and the context before the segment is manipulated such that in one condition it signals a prosodic break larger than a word boundary (IP in (Kim & Cho, 2013), or AP in (Steffman et al., 2022)) while in another it does not. The participants showed a significant shift in their perceptual boundary depending on what phonetic values the preceding context had. The results suggest that the phonetic properties before the segment was interpreted by the listeners as signaling prosodic boundaries, and the listeners used such information in identifying the post-boundary segment.

For instance, Kim & Cho (2013) showed that when an English stop from a /p/-/b/ continuum varying on VOT was preceded by a lengthened syllable, English listeners required a longer VOT to categorize the stop as a /p/. Their stimuli were 'let's hear 'p/bah' again' where the word 'hear' was lengthened or not. They argue their results suggest that when listeners hear a lengthened 'hear', they parse the temporal cue as signaling an IP boundary, after which the aspirated stop /p/ is produced with a longer VOT in production. This would lead the listeners to expect a /p/ with a longer VOT, as they expect domain-initial strengthening, and consequently, they shift their perceptual boundary in the categorization task.

Steffman (2019) showed in two experiments also in English that listeners compute the speech rate differently, depending on their interpretation of the prosodic structure. Steffman's stimulus design was similar to the one in Kim & Cho (2013), except he varied two conditions on the word before the ambiguous stop (/p/-/b/) varying on VOT: the duration and the boundary tone. Steffman observed that when the word was short and had a boundary tone ('L-H%'), the listeners required 'shorter' not 'longer' VOT for a /p/. The logic used in interpreting the results is as follows. When listeners hear the boundary tone, they hypothesize that it is an IP boundary, after which they would require a longer VOT for a /p/. However, they also hear the word bearing the boundary as short, when they expect the word to be lengthened via pre-boundary lengthening. The listeners then assume the speaker has spoken at a faster speech rate. Therefore, it further modulates how the post-boundary segment is categorized. In other words, when the phonetic properties of the pre-boundary syllable conflict each other, since a boundary tone signals a boundary and a short duration does not, listeners find a way to resolve the conflict, by interpreting the short duration as speeding up the speech rate. On the contrary, Steffman did not observe the opposite effect where the long syllable with a 'flat' boundary tone was perceived as a local slow down, independent of the prosodic structure, because the 'flat' boundary tone was a viable pitch movement as a boundary tone when it was produced with a long syllable.

The results are interesting in two ways. First, when there is a pitch movement on a single syllable, it always suggests a prosodic boundary, such that no other information (a relatively short syllable) can revert the computation. Instead, the duration of the syllable is processed based on the expectation that it should be lengthened. Second, a longer syllable is always indicative of a prosodic boundary, but the lack of a pitch movement cannot always signal an absence of a boundary, since it may be perceived as a 'flat' boundary tone.

Steffman provided further support for this speech-rate normalization account in the second experiment, which showed that the participants required a shorter duration of the vowel to judge an ambiguous stop (/t/-/d/) as a /d/ in the word 'coat/d' when the preceding word 'say' was shorter but had a boundary tone. In other words, when the pre-boundary word had a boundary tone but a shorter duration, listeners perceived it as a boundary produced at a fast speech rate and considered a shorter duration of the vowel before the 't/d' coda sufficient for suggesting a cue for a voiced coda. This result supports the analysis that the boundary tone causes a speech rate normalization, rather than simply affecting the processing of the post-boundary segment. The vowel duration information becomes available after the onset (/k/ of 'coat/d') and the segment the listeners are asked to identify follows this vowel.

Following up on this work, Steffman et al. (2022) tested if a pitch variation alone (with no lengthening) could signal a boundary percept in Seoul Korean such that the post-boundary segment categorization could be modulated. This is because the AP boundary is argued to be implemented as a tonal transition but not with a pre-boundary lengthening (Jun, 1993, 2000). They tested the perception of an ambiguous stop between fortis and aspirated categories based on the F0 of the preceding syllable. When the preceding syllable had a higher F0, which indicated an AP boundary, listeners required a longer vowel duration to perceive a token as a fortis, because the vowel duration lengthens after a fortis stop in an AP-initial position compared to an AP-medial position.

Steffman's experiments showed that listeners shift their expectation of the segmental realization, based on the prior information they get about the prosodic structure and the speech rate at which the segment is realized. If the prior information suggests that the next segment will be spoken in a phrase-initial position, they would expect the segmental realization to have the acoustic characteristics that the segment tends to have in that position. These results seem to suggest that the listeners compute the prosodic structure and the speech rate based on the cues available to them, which they use to process the upcoming cues to make the segmental identification decision.

The second kind of experiment, the 'phonological inferencing' experiment (e.g., Kim et al. (2018)) can also be interpreted from this perspective. In these experiments, it was shown that the prosodic break (either produced naturally or imposed experimentally by manipulating the phonetic properties) can block the phonological inferencing process, which is 'undoing' a phonological process when there is a viable lexical competitor. For instance, in Korean, lenis obstruents are tensified (become fortis obstruents) when they follow an obstruent coda. This process is shown to be conditioned by the AP (Jun, 1993) such that when there is an AP boundary between the coda and the lenis obstruent, the lenis obstruent is not tensified.

Kim et al. (2018) showed that when there is no evidence of AP boundary between the coda and the lenis obstruent, participants gave significantly more looks to the lenis target, 'undoing' the tensification, compared to a distractor. However, they only did this when there was no lexical competitor with the fortis obstruent onset. Additionally, they demonstrated that when there was an F0 cue before the target word that signaled an AP-break, this phonological inferencing effect disappeared, which suggested that the participants did not consider a token that sounds like a fortis to be an instance of an underlying lenis obstruent token that was tensified, once they perceived the AP boundary, which blocks the tensification process.

The experiments showed that when the phonetic properties are manipulated to cause a percept of prosodic boundary, listeners seem to update their expectation of the phonetic values of the following segment, based on their computation of the prosodic representation.

1.3.5 Interim summary

This chapter started by explaining that when speakers produce an utterance, it is pronounced in 'chunks' which we call prosodic constituents, and marked with boundaries of the constituents with various acoustic cues, such as duration of the pre-boundary syllable, pitch contour near the boundary, and realization of the post-boundary segment.

I reviewed that in Seoul Korean, the combinations of these acoustic cues signal different types of boundaries, and the segmental and suprasegmental information are both important cues for the types of boundaries. I discussed a potential example (Figure 1.2) which involves the computation of the acoustic cues in the inference of both the prosodic boundary and the segmental identity, necessary for resolving a potentially four-way ambiguous set of sentences.

Finally, I discussed ways in which the role of the prosodic boundaries is tested in perception, which is summarized in (8).

(8) Summary of perceptual evidence for prosodic structures

a. Prosodic transcription: When we ask listeners, they can say they perceive different sizes of boundaries. We can make post hoc comparisons of these perceived boundaries, to figure out how the listeners made different judgments, assuming we know the right phonetic measurements to use in the comparisons. Still, listeners might be influenced by non-acoustic information in making such judgments, e.g., the syntactic parse of the utterance.

b. Word spotting experiments: When we ask the listeners to spot words, they do better if the perceived boundary is larger in general, but only if the segment after the boundary is pronounced as expected, given the prosodic structure hypothesized by earlier acoustic cues (Kim & Cho, 2009).

c. Prosodic disambiguation experiments: When we ask the listeners to disambiguate the meanings of syntactically ambiguous sentences which differ in the prosodic structure in production, listeners make use of acoustic cues that correlate with prosodic boundaries and interpret the sentence based on the perceived prosodic structure.

d. Segment identification: When we manipulate the stimuli to create a percept of a prosodic boundary and ask listeners to make the phoneme judgment for the segment after the boundary, they seem to process the acoustic cues and formulate a prosodic structure, which they use to update the expectations for how the segment after the boundary is pronounced.

These experiments provide evidence that the computation of the prosodic structure and the segmental identity are interdependent. Processing of the prosodic structure is affected by the processing of the post-boundary segment, and vice versa. Some of the results also provide clues about how the acoustic cues are integrated in real-time to make the judgments. For instance, when the listeners are asked to compute the prosodic structure (e.g., in a word-spotting experiment), they seem to use the post-boundary segmental realization to evaluate possible alternative prosodic structures that are currently being considered. When the post-boundary segment does not have the phonetic properties that match the strong candidate that they are considering, listeners are hindered more in the word spotting task (Kim & Cho, 2009). Similarly, when the listeners are asked to compute the post-boundary segment, they seem to update their expectation for the phonetic realization of the post-boundary segment, depending on their computation of the pre-boundary acoustic information.

1.4 Cho et al (2007) model of speech perception

The experimental results reviewed in §1.3 generally assume a specific model of speech perception, first proposed in Cho et al. (2007). More recently, McQueen & Dilley (2020) have proposed a Bayesian Prosody Recognizer, which formalizes the Prosody Analyzer model as a Bayesian inference procedure where the listener 'combines prior knowledge with signal-driven likelihoods to obtain an optimal interpretation of current input' (p. 5), where the 'interpretation of current input' includes both the prosodic structure as well as the segmental identity However, these models currently lack a specific computational implementation, in the sense that there is no formalization of this model to the extent that it can be readily used to model a result of a perception experiment. I will review the verbal descriptions of this model in the past literature in this section and discuss what is being proposed.

This model has two modules: the Segment Analyzer and the Prosody Analyzer, which interact with each other. The Segment Analyzer processes 'information in the speech signal which is relevant for phonemic contrasts'; while the Prosody Analyzer computes the prosodic structure of the incoming signal (Cho et al., 2007). For instance, when hearing the syllable 'pa' at the end of an utterance, the Segment Analyzer identifies that the segments in the signal are /p/ and /a/, while the Prosody Analyzer identifies that the segment is in IP-final position, by noticing that there is a pre-boundary lengthening as well as a boundary tone.

These two analyzers work in parallel and can affect each other. Once the Prosody Analyzer outputs the prosodic structure (e.g., the presence of an IP boundary), the

Segment Analyzer uses the information to update the expectations for the phonetic values of the upcoming segment.

As Cho et al. (2007) point out, the Prosody Analyzer must also rely on the Segment Analyzer because to compute whether the syllable was lengthened, it would first need to know how long /a/ tends to be ('intrinsic duration' of /a/). One way that this can be implemented is by assuming the listener stores the distributional information of the acoustic cues. For instance, when a syllable is perceived, the pitch contour of the syllable must be used to decide if it has a boundary tone, and if so, which boundary tone. This may be done by comparing this token with the distributions for each boundary tone type, and deciding that it matches the most closely with one type of boundary tone. Similarly, when a syllable is perceived, its duration is compared with other IP-final syllables that are stored. Note that there must be some speech-rate normalization before the comparison since the absolute duration measures cannot be compared directly if the syllables are spoken at different speech rates.

The prosodic disambiguation experiments and the segment identification experiments reviewed in §1.3.3 and §1.3.4 are set up such that one aspect of the acoustic signal becomes given information for making inference about another aspect. In the prosodic disambiguation experiments, post-boundary lexical items are generally chosen such that they would not be misperceived as a different lexical item, despite its acoustic variation. For instance, in Yoo (2020), a weak or strong realization of lenis obstruent would never be misperceived as aspirated or fortis, because the experiment did not include such lexical items in the list of stimuli words. Likewise, in the segmental identification experiments, while the presence or absence of a prosodic boundary is manipulated, it does not seem to change how the overall sentence meaning is processed. For instance, in Steffman (2019), the carrier sentence used was 'I'll say _____ again', where the duration and the pitch movement on the word before the target word was manipulated to vary the prosodic structure of the utterance. However, this manipulation arguably would not create a difference in the meaning of the sentence (i.e., an IP boundary between ‘say’ and the target word would mean the same as a PW boundary between ‘say’, except the post-boundary segment may be perceived as two different phonemes), in the same way that acoustic manipulations in a prosodic disambiguation experiment does.

However, as demonstrated in the quadruplet example in Figure 1.2 (§1.2.3), these two things could simultaneously be ambiguous in a naturalistic setting. In this context, processing of a particular segment immediately informs both the lexical contrast and the prosodic contrast. There needs to be a single joint inference for a phonological representation, which includes both the prosodic structure, and the segments therein, for a given acoustic signal. This is what I will attempt to show with the perception experiment in Chapter 6.

1.5 Overview of this dissertation

This dissertation has two components: Production and Perception. In the Production component, I demonstrate the role of the post-boundary segmental realization in signalling the size of the prosodic boundary. I provide empirical evidence that segmental realization can sometimes even be more informative than F0 patterns, for differentiating prosodic structures. These findings demonstrating the importance of segmental cues run counter to the prevailing trend has been to focus on the duration and the F0 pattern when studying the acoustic implementation of prosodic boundaries (Jun, 2000, 2022). Segmental realization has been assumed to be less informative by some researchers since segmental processes seem to be only optionally conditioned by prosodic structure (Jun, 1994; Arvaniti & Baltazani, 2005). One example of such segmental processes is how lenis obstruents are realized in the AP-medial position. While it has been found that lenis obstruents are voiced in the AP-medial position, it has also been reported that they can optionally remain fully voiceless (Jun, 1993, 1994; Han, 2000).

However, in Chapter 2, I show that the acoustic realization of lenis obstruents in a spontaneous speech corpus of Seoul Korean (Yun et al., 2015) is affected in multiple ways by the prosodic structure. I find that the change in intensity (a measure of phonetic reduction) and the change in duration is always (though gradiently) affected by the prosodic structure, and therefore there is no optionality in the effect of prosodic structure on the segmental realization contra what has been suggested by the previous literature (Jun, 1994; Arvaniti & Baltazani, 2005).

Chapter 3 presents another corpus study on the same spontaneous speech corpus. It shows that the segmental realization of lenis obstruents seems to better distinguish different prosodic boundary types, compared to the tonal transition between syllables. However, for Aspirated obstruents, the segmental realization does not differentiate different prosodic boundary types, and only the tonal transition is informative.

Chapter 4 is an interim chapter, where I present a construction in Seoul Korean involving the polysemous word /kɨman/ ('start/stop'), which has the four constrasts in the schematic example in Figure 1.2. It thus provides an ideal case study for how prosodic structure and phonemic contrast are simultaneously realized acoustically. I discuss the syntactic and prosodic structure of this construction and present a Maximum Entropy (Goldwater & Johnson, 2003) grammar analysis of how variable prosodic phrasing of a syntactic structure may be modeled. This construction will be the focus of the rest of the chapters.

Chapter 5 presents a detailed acoustic analysis of this construction and again shows that the tonal contour may not be informative in differentiating the prosodic structures, while the segmental realization of lenis obstruent does.

In Chapter 6, I test the role of realization of lenis obstruent in prosodic and lexical disambiguation in speech perception, using the same construction: /kɨman/

+ Verb ('start/stop verb-ing'). The results indicate that listeners are sensitive to the correlation between the prosodic context and the segmental realization.

Finally, in Chapter 7, I summarize the findings of this dissertation and discuss their implications for how we may model the interaction between phonetics, prosody and syntactic structure in speech production and perception.

CHAPTER 2

LENITION OF LENIS OBSTRUENTS IN SEOUL KOREAN SPONTANEOUS SPEECH

2.1 Introduction

This chapter revisits one of the oft-cited prosodically-conditioned segmental processes reported to be both optional and gradient: Seoul Korean lenis obstruent voicing. Previously, Jun (1993, 1994, 1995, 1998) has shown that lenis obstruents in Seoul Korean are intervocalically voiced only when they are within a prosodic constituent called the 'Accentual Phrase' (AP), which is defined by its tonal pattern, as in (1). The features [-continuant], [-constricted glottis], and [-spread glottis] are used to refer to lenis obstruents, as opposed to aspirated and fortis ones, and α is to indicate that the domain is the AP.

(1) [-cont, -const. glot., -spread glot.] $\rightarrow$ [+voice] / α(..[+voice] ___ [+voice] ..) (Jun, 1993, p.78)

Jun (1993) argued the lenis voicing process is optional and gradient because in her production study, Jun found AP-medial tokens that were fully voiceless, which were exceptions to the generalization in (1). Jun's results have been replicated in subsequent works (Han, 2000; Yun, 2000) which have also shown that lenis obstruents in AP-medial position are not always fully voiced and reported a range of possible phonetic outcomes from fully voiced to partially voiced to fully voiceless.

While the prosodically-conditioned realization of Seoul Korean lenis obstruents has often been described as a voicing phenomenon, lenis obstruents have also been reported to be shorter (Cho & Keating, 2001), fricativized, or approximantized (Silva, 1992; Lee, 1995), and even deleted (Kim et al., 2016) in AP-medial position. Throughout this study, fricativized or approximantized realization of lenis obstruents, as opposed to a full-stop realization, will be referred to as 'reduction'. While reduction may be phonetically measured in multiple ways, in this study, it is measured as the change in the velocity of intensity change following Kingston (2008) and Katz & Pitzanti (2019).

Previous work on other languages has noted that a reduced obstruent token is often characterized as more sonorant and vowel-like which is reflected in its high intensity, whereas a full stop token is produced with complete occlusion of airflow, thereby causing a rapid decrease in the intensity compared to the preceding vowel, followed by a rapid increase in the intensity going into the following vowel (Harris & Urua, 2001; Kingston, 2008). Therefore, the relative size of the change in intensity and the velocity of that change can be phonetic correlates that reflect the degree of reduction (see §2.3.3 for methods).

Cross-linguistically, reduction processes, i.e., spirantization, fricativization, or approximantization, tend to target obstruents in phrase-medial intervocalic positions (Lewis, 2001; Kingston, 2008; Torreira & Ernestus, 2011; Bouavichith & Davidson, 2013; Ennever et al., 2017; Katz, 2016; Katz & Fricke, 2018; Katz & Pitzanti, 2019; Katz, 2021; DiCanio et al., 2022). As a result of reduction, obstruents become more vowel-like (Szigetvári, 2008; Katz & Fricke, 2018), shorter (Lavoie, 2001), and more sonorous (Smith, 2008; Katz & Fricke, 2018). The fact that reduction typically occurs within a prosodic constituent and that it results in a more sonorous phonetic realization of obstruents has led some researchers to argue that the lenited realization of these obstruents signals that 'the current prosodic constituent is continuing rather than ending or a new one beginning' (Kingston, 2008, p.1). Katz (2016) calls this 'continuity lenition', and the term has been used to describe similar lenition processes in Campidanese Sardinian (Katz & Pitzanti, 2019) and Gran Canaria Spanish (Broś et al., 2021).

This study characterizes the realization of Seoul Korean lenis obstruents as continuity lenition and investigates their acoustic realization in all three of the prosodically conditioned phonetic processes, i.e., voicing, reduction, and shortening. I will use the term 'lenition' as an umbrella term that includes all three of these phonetic processes.

This study has two research questions summarized in (2), which will be referred to as 'Optionality of voicing' and 'Optionality of lenition'.

(2) Research questions

a. Optionality of voicing: is AP-medial voicing of Seoul Korean lenis obstruents optional?

b. Optionality of reduction and shortening: Are AP-medial shortening and reduction of Seoul Korean lenis obstruents optional?

To address these research questions, this study investigates a spontaneous speech corpus of Seoul Korean (Yun et al., 2015). The three phonetic processes will be measured respectively with the following phonetic variables: the proportion of voiceless interval (§2.3.2) for voicing, the maximum falling velocity in the transition from the vowel to the obstruent closure (§2.3.3) for reduction, and the duration which includes both the closure and the release (§2.3.3) for shortening.

Most previous work on Seoul Korean lenis obstruents has used lab speech (Jun, 1993, 1994, 1995; Han, 2000; Yun, 2000; Choi et al., 2006; Yoo, 2020), and almost exclusively focused on the voicing while neglecting the reduction of lenis obstruents (cf. Lee, 1995; Yoo, 2020). This is partially due to the methodological challenge of dealing with reduced (fricativized and approximantized) lenis obstruent tokens. When Seoul Korean lenis obstruents are reduced, it can be hard to identify where they start and end from manually looking at spectrograms and waveforms. Plus, it is hard to quantify the degree of reduction which has been manually marked token by token, with a binary label (e.g., 'fricativized' and 'not fricativized') (Lee, 1995; Yoo, 2020), while empirically, obstruent reduction tends to be more gradient rather than categorical (De Jong, 1998; Simonet et al., 2012; Parrell & Narayanan, 2018). Having to code the tokens manually also limits the number of tokens that are investigated.

This study overcomes these methodological limitations by using an automatic algorithm that uses the annotation and audio files of a spontaneous speech corpus and automatically measures the duration (closure and release) and the degree of reduction of a lenis obstruent token.

Variants of this algorithm have been applied in research on lenition patterns of Peruvian and Ecuadorian Spanish (Kingston, 2008), Gurindji (Ennever et al., 2017), and Campidanese Sardinian (Katz & Pitzanti, 2019). The number of lenis obstruent tokens investigated in this study (91,112) far exceeds the number of tokens investigated in previous studies (on the order of hundreds for Kingston (2008) and Ennever et al. (2017) and about 4,000 in Katz & Pitzanti (2019)). Compared to data collected from production experiments, corpora generally provide much more data, which increases the statistical power in modeling the pronunciations of lenis obstruents. Spontaneous speech is ideal for lenition research since lenition tends to occur more frequently when speakers speak casually and spontaneously (Warner & Tucker, 2011; DiCanio et al., 2022).

To preview the results, this study replicates previous findings that voicing of lenis obstruents is optional (Jun, 1993, 1994; Han, 2000; Yun, 2000), as there are fully voiceless PW-medial lenis obstruents tokens in the spontaneous speech corpus which are preceded by a fully voiced vowel. However, this study finds that these fully voiceless PW-medial lenis obstruents are nonetheless realized as more reduced and shorter in duration compared to PW-initial lenis obstruents that have the same voicing profile, i.e., fully voiceless after a fully voiced vowel.

The rest of the paper is structured as follows. §2.2 presents background on lenis obstruents and Accentual Phrases in Seoul Korean, and presents the two research questions in more detail. §2.3 describes the corpus and the algorithms used to measure the phonetic variables to investigate the three acoustic characteristics of lenis obstruents: voicing, reduction, and shortening. §2.4 presents the exploratory results, showing the effect of prosodic position on the realization of lenis obstruents in voicing, reduction, and duration. It presents results relevant for the first research question: optionality of voicing. §2.5 presents inferential statistical analyses to address the second research question: optionality of reduction and shortening. §2.6 discusses the theoretical implication of the findings for the relationship between the prosodic structure and segmental realization. §2.7 concludes the chapter.

2.2 Background and research questions

Lenis obstruents in this study refer to lenis stops in three places of articulation–bilabial, alveolar, and velar (/p, t, k/), plus the lenis postalveolar affricate (/ʧ/). These may be most accurately referred to as [-continuant, -nasal] sounds because the term 'obstruents' technically includes fricatives as well. Fricatives /s, s*, h/ in Korean are not investigated as they are not of 'lenis' category: /s*/ is fortis, /h/ is aspirated, and /s/ does not seem to be straightforwardly categorized as lenis either (Chang, 2008), primarily because /s/ does not become fully voiced, unlike other lenis categories.

All three lenition processes make reference to the prosodic constituent called the Accentual Phrase (AP). The AP is one of the prosodic constituents proposed for Seoul Korean in the autosegmental-metrical system of Seoul Korean intonation called the K-ToBI model (Jun, 1993, 2000, 2006, 2007). I briefly review the K-ToBI model here. For a more thorough discussion of the K-ToBI model, see §1.2.1. In this model, there are three levels of prosodic constituents above the Phonological Word (PW) which are assumed to be strictly layered: Accentual Phrase (AP), intermediate phrase (ip), and Intonational Phrase (IP) (Jun, 1998, 2000, 2006, 2007). The highest two constituents in the K-ToBI model, the ip and the IP, are not the focus of the current study since the processes studied in this study seem to be conditioned by the AP level.

The AP starts with either a 'H(igh)' or a 'L(ow)' tone, depending on its first segment. If the first segment is an aspirated or a fortis obstruent, the AP-initial tone is an 'H', and when the first segment is anything other than an aspirated or fortis obstruent, e.g., a lenis obstruent, it is an 'L'. The last tone of an AP is most frequently 'Ha' though it can also be 'La'. The '-a' is a diacritic that indicates that the tone is associated with the *A*ccentual Phrase edge. An AP may consist of multiple PWs though Jun (2000) argues that an AP is typically a lexical item plus a post-positional marker, such as the case marker. However, the AP is an intonational constituent as opposed to a syntactic one, and is defined as tonally 'demarcated by a phrase-final High tone' (Jun, 1998, p.190).

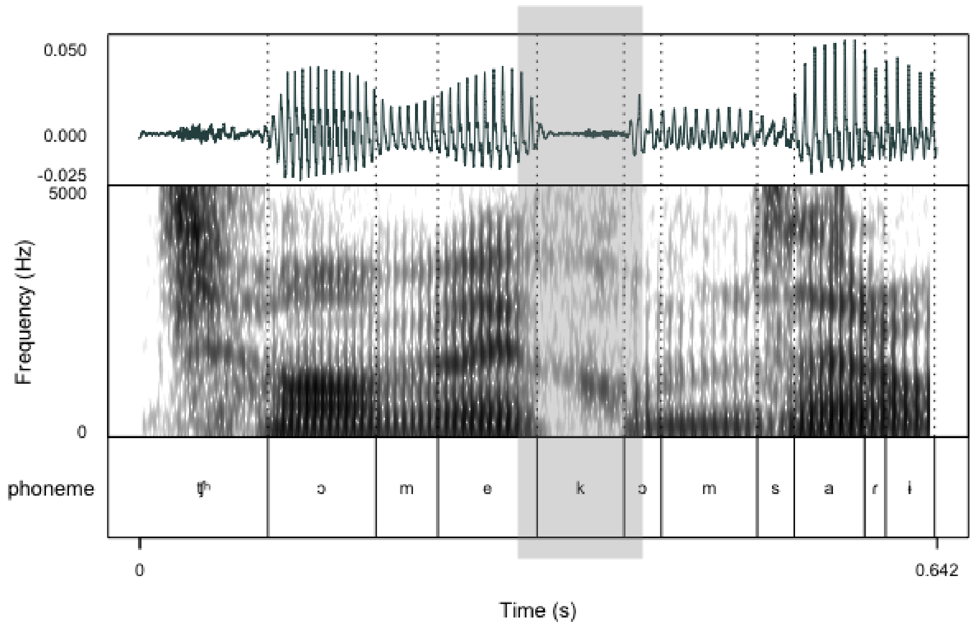


Figure 2.1: Waveform, spectrogram, and phone-level annotation of an AP-initial PW-initial voiceless /k/, taken from the Seoul Corpus (Yun et al., 2015). InputFile: s01m16f6, /k/ starts at t = 162.30530. /ʧʰɔme/ ('at first') /kɔmsa-ɾɨ/ ('examination-ACC').

Figure 2.1 plots the waveform and spectrogram of a typical voiceless and aspirated realization of a PW-initial lenis velar stop (the region highlighted with a gray box), taken from the Seoul Corpus (Yun et al., 2015).[1] In contrast, Figure 2.2 shows a voiced realization of a PW-medial /k/ from the corpus. This particular token is clearly approximantized and fricativized, as well as being voiced, as there seems to be no sign of stop closure. Throughout this chapter, waveforms plotted in red are taken from the PW-medial position.

[1]Throughout this chapter, token start time ('t') indicates where the lenis token in the figure starts in the original sound file. On the figures, the time axis always starts at 0.

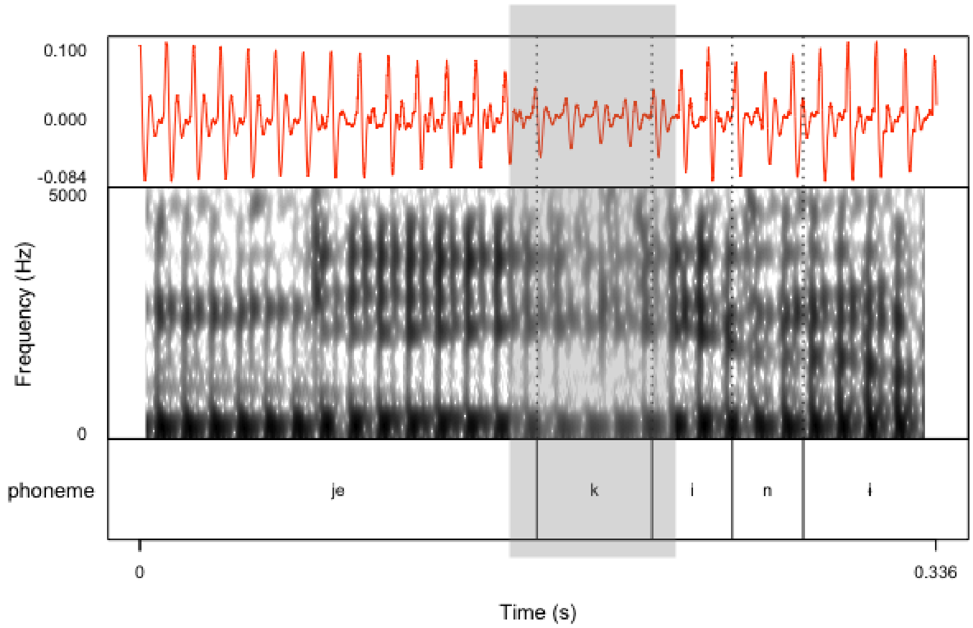


Figure 2.2: Waveform, spectrogram, and phone-level annotation of a PW-medial voiced /k/, taken from the Seoul Corpus (Yun et al., 2015). InputFile: s02m16f1, /k/ starts at t = 175.29585. /jeki-nɨ/ ('story-TOP').

2.2.1 Optionality of voicing

Recall that the generalization is that lenis obstruents are voiced between voiced segments within an AP, which is repeated in (3).

(3) [-cont, -const. glot., -spread glot.] $\rightarrow$ [+voice] / α(..[+voice] ___ [+voice] ..) (Jun, 1993, p.78)

There are two possible ways shown in (4) that this generalization may be broken. It may be called an exception if a lenis obstruent is voiced in the AP-initial position as in (a), or if a lenis obstruent is voiceless in the AP-medial position as in (b).

(4) Two possible exceptions to the generalization in (3).

a. AP-initial voicing: [[+voice] $]_{AP}$[**[+voice]**[+voice]]

b. AP-medial non-voicing: $_{AP}$[[+voice]**[-voice]**[+voice]]

(i) PW-medial non-voicing: $_{AP}[_{PW}[$ [+voice]**[-voice]**[+voice]]]

For a corpus investigation such as the present study, it is challenging to identify the boundaries of the AP, since it requires rigorous prosodic transcriptions, preferably done by multiple trained transcribers. Moreover, as Yoo (2020) points out, depending on how the AP boundaries are annotated, an 'AP-initial' lenis token in one study might be analyzed as 'AP-medial' in a different study (p. 84). Therefore, the first kind of exception–AP-initial voicing–may be harder to investigate in a corpus study such as the current study.

On the contrary, the second kind of exception–AP-medial non-voicing, can be addressed, as we can investigate lenis obstruents in PW-medial position (4b-i). Unlike APs, PWs are not defined prosodically since there is no word-level prominence (word stress) in Korean. They are widely assumed to be the lexical words (or 'Eojeol's in Korean, which is defined as a white space delimited sequence of letters in orthography) both in the K-ToBI convention (Jun, 2000) and in other works (e.g., Han, 2000; Choi et al., 2006). I follow this convention and use the term 'Eojeol' interchangeably with the word 'PW', throughout this dissertation. Eojeols are content words combined with optional suffixes, such as verb-endings and case markers. Since the corpus I investigate provides an Eojeol-level segmentation but no prosodic transcriptions including AP boundaries, I specifically use 'Eojeol' to refer to the intervals marked in the corpus. In the K-ToBI model, the levels of the prosodic constituents are assumed to be properly bracketed and strictly layered (Jun, 1993, 2000). This indicates that a PW-initial syllable can be AP-initial or AP-medial, depending on its alignment with the AP, while a PW-medial syllable is always AP-medial since an AP cannot start in the middle of a PW. Therefore, the lenis obstruents in the onset of the PW-medial syllables provide a good test case for the optionality of the voicing process, since the PW-medial lenis obstruent onsets are expected to be always voiced if the voicing process is not optional.

Jun (1994) models the AP-medial Lenis voicing generalization in (3) as blending and reduction of glottal opening gestures, using Browman & Goldstein (1989)'s gestural score model of phonetic representation. In this proposal, lenis obstruents are AP-medially voiced because the glottal opening gestures for the lenis obstruent are blended with those for the preceding vowel. As Jun (1994) points out, if this is the right interpretation of the AP-medial lenis obstruent voicing, then lenis obstruents may be fully voiceless in the PW-medial position if the preceding vowel is also devoiced. Indeed, there were lenis obstruent tokens in Jun's study that were fully voiceless in the PW-medial position and were preceded by fully devoiced vowels. These tokens have not been considered as true exceptions to the AP-medial voicing generalization as the lenis token is not preceded by a voiced segment. Jun (1994) showed however that not every instance of PW-medial fully voiceless lenis obstruent was preceded by a fully devoiced vowel, and some PW-medial tokens in her production data were fully voiceless despite having a regular, not devoiced, vowel before them. These PW-medial fully voiceless realizations of lenis obstruents after a fully voiced vowel were also observed in a later study (Han, 2000).

The first research question of the study seeks to replicate both these previous findings that some PW-medial voiceless lenis tokens are due to being preceded by a preceding devoiced vowel (Jun, 1994) but some PW-medial lenis tokens may be fully voiceless after a voiced vowel (Jun, 1993, 1994; Han, 2000; Yun, 2000). Such empirical evidence suggests AP-medial voicing of lenis obstruents is optional. This amounts to finding tokens in (5), i.e., PW-medial voiceless tokens that are preceded by a devoiced vowel ('v̥') or a regular vowel ('v'). The result of this research question is presented in §2.4.

(5) $_{AP}[_{PW}[$ ...Lenis$_{[-voice]}$...]]

a. Not an exception of (3): $_{AP}[_{PW}[$ v̥ Lenis$_{[-voice]}$ v]]

b. True exception of (3): $_{AP}[_{PW}[$ v Lenis$_{[-voice]}$ v]]

2.2.2 Optionality of lenition

If investigating the first research question leads to a successful replication of previous findings (Jun, 1993, 1994; Han, 2000; Yun, 2000), then the second research question is called for: given the optionality of AP-medial voicing, are AP-medial shortening and reduction of Seoul Korean lenis obstruents also optional? Investigating the optionality of prosodically conditioned segmental processes has theoretical importance since it has raised doubts about the reliability of segmental realization in signaling prosodic constituency (Arvaniti & Baltazani, 2005). The second research question is summarized in (6).

(6) Optionality of lenition: Are AP-medial shortening and reduction of Seoul Korean lenis obstruents optional?

a. Optional prosodic effect: fully voiceless AP-medial lenis obstruents can be produced without reduction and with long duration.

b. Obligatory prosodic effect: fully voiceless AP-medial lenis obstruents are produced as reduced and with short duration.

When lenis obstruents are realized as fully voiceless in the AP-medial position, can they also be produced without reduction, and with long duration, as if they are produced in the AP-initial position? This would mean that a particular realization of segment does not cue the prosodic position reliably. Alternatively, when a lenis obstruent is fully voiceless in the AP-medial position, it may be produced as reduced and shorter, despite the voicelessness, thus reliably signaling the continuation of the current prosodic constituent, as 'continuity lenition' account predicts (Kingston, 2008; Katz, 2016; Katz & Pitzanti, 2019). The former will be referred to as the 'optional prosodic effect' hypothesis, while the latter will be referred to as the 'obligatory prosodic effect' hypothesis, as in (6). The result of this question is presented in §2.5.

2.3 Methods

This section describes methods to parameterize lenis obstruents in terms of degree of voicing, degree of reduction, and duration. It first introduces the spontaneous speech corpus (Yun et al., 2015) used in this study (§2.3.1), then phonetic measurement for voicing in (§2.3.2), and phonetic measurements for duration and degree of reduction (§2.3.3). Methods for inferential statistics appear later in §2.5.

The complete workflow of the methods in this study is also described in a document named 'Doing phonetic research on the Seoul Corpus' in the OSF repository[2] that hosts the supplementary materials, which also provides documentation for all Praat, Python, and R scripts for processing the corpus annotation files and audio files, as well as for generating the figures used in this study.

2.3.1 Corpus

Data for this study came from a spontaneous speech corpus of 40 Seoul Korean speakers, the 'Seoul Corpus' (Yun et al., 2015). Five male and five female speakers were recorded in each of four age groups: teens, twenties, thirties, and forties. The youngest of the forty speakers was a fifteen-year-old male, the oldest was a forty-seven-year-old female. Speakers and their parents were raised in or near Seoul. Speakers conversed with the interviewers on predetermined topics, but the conversation was not limited to those topics. Each speaker's recording lasted roughly one hour. The interviewer's speech and non-speech intervals such as laughter were removed which reduced the total duration of the recordings from 42.8 to 24.2 hours.

A screenshot of a sample corpus annotation file is in Figure 2.3. The corpus annotation files provide three kinds of segmentation: Utterance, Eojeol (lexical word), and phone. The Utterance tier contains chunks of running speech delimited by 'silence' intervals. In the Eojeol tier, the utterances are segmented into Eojeols. There are two

[2] `https://osf.io/ukh6d/?view_only=d9bda726aebe4512830c6996f2ae4cae`

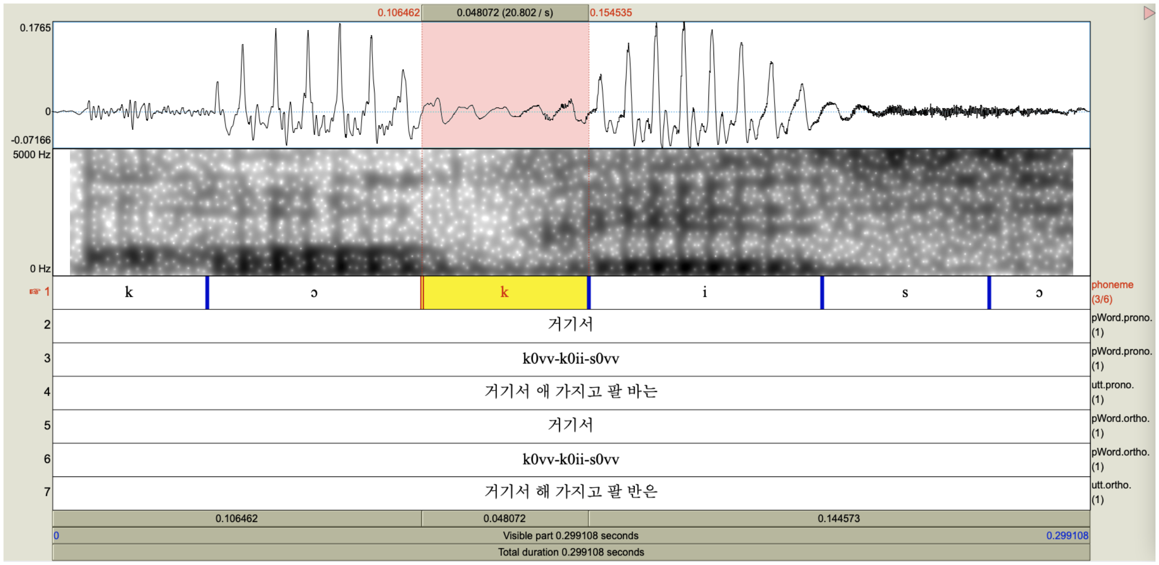


Figure 2.3: Screenshot of a corpus wave file and annotations. Tiers: phone, Eojeol pronunciation (Korean), Eojeol pronunciation (Romanized), Utterance pronunciation (Korean), Eojeol orthography (Korean), Eojeol orthography (Romanized), Utterance orthography (Korean).

kinds of Eojeol tiers: orthographically transcribed and phonetically transcribed. The phonetically transcribed Eojeol tiers contain the broad transcription of the pronunciation of the Eojeols, which reflects any alternation of phonemes (lenis obstruents realized as fortis or aspirated)[3] but not if there is an 'allophonic' change (lenis obstruents realized as voiceless or voice). I focus on the lenis obstruents that are transcribed as lenis in the phonetically transcribed tier, to exclude tokens that are underlyingly lenis but realized as fortis or aspirated obstruents. I found relevant lenis obstruents in the corpus that were specifically between two vowels to control for the neighboring segment environment. The description in (3) defines the phonological environment as between two [+voice] segments. Since Korean phonotactics do not allow consonant clusters, this means that lenis tokens between two [+voice] segments mean they may appear after a sonorant coda and before a vowel. I limited the tokens I investigated to

[3]For instance, lenis obstruents alternate to fortis when they're produced after an obstruent coda as in [sok.t*o] /sok.to/ ('speed'), and they alternate to aspirated following a certain morpheme as in [sil.t^ha] /sil-ta/ ('hate-DECL').

the lenis obstruent onsets in the intervocalic position only, excluding tokens that are between a sonorant coda and a vowel. This also excluded lenis tokens that appeared after a pause interval (i.e., post-pausal utterance-initial tokens). In the Eojeol tiers, the syllable boundaries were marked with '-', which was used to identify the lenis obstruents in the onset position, but not in a coda position. When a lenis obstruent was in a coda position, and resyllabified to the onset position of the next syllable, this change was only reflected in the phonetically transcribed Eojeol tier, but not in the orthographically transcribed Eojeol tier. For instance, /mok.i/ ('neck-NOM') would be transcribed as 'mo-ki' in the phonetic Eojeol tier, but as 'mok-i' in the orthographic Eojeol-tier. An Eojeol-final coda could in theory be resyllabified to an onset of the following Eojeol-initial syllable, but there were only 40 instances of such cases, and I excluded these tokens from analyses.

Finally, in the Phone tier, the phonetically transcribed Eojeols are segmented into phones. The intervals for obstruents contain both the closure interval and the aspiration interval, if there is any, meaning they included the speech from the end of the preceding segment (i.e., the preceding vowel) to the start of the following segment (i.e., the following vowel). The segmentation and transcription (Eojeol-level and phone-level) were first done with an automatic forced-aligner and manually corrected by 9 labelers. The authors report an impressively high agreement rate of 98.1% among the 9 labelers (Yun et al., 2015).

For this study, the Phone intervals were searched using the '`textgrid`' package (Gorman, 2014) in Python to select the obstruents in the syllable onset position (those preceded by the syllable boundary symbol in the phonetic Eojeol tier). This resulted in 150,088 obstruent onset tokens in the intervocalic position in total. These tokens were labeled 'Eojeol initial' and 'Eojeol medial', depending on where the phone interval was with respect to the Eojeol interval boundaries. Not every token was included in the analyses however because some had to be removed in the process of parameterization, which we turn to next. The final number of tokens that were analyzed is summarized in Table 2.2 in §2.3.3. The other two obstruent categories (Aspirated and Fortis) were also investigated so as to demonstrate that the acoustic realization of lenis obstruents are clearly conditioned on the prosodic position while the other two categories remain relatively unchanged in terms of the phonetic dimensions that I measured.

2.3.2 Parameterizing the voicing: Proportion of voiceless interval

The degree of voicing measure was measured using a Praat script, which uses 'the fraction of locally unvoiced frames' contained in the Voice Report (Boersma & Weenink, 2023). The minimum and maximum F0 were set differently for male and female speakers to get the best result since the Voice Report function in Praat works by measuring whether there is detectable f0 in a given interval (see Praat manual for more explanation (Boersma & Weenink, 2023)). Other values were set to Praat's default. This measure was used in previous work (e.g., Davidson, 2016, 2018) to quantify the degree of stop voicing in a gradient way, rather than transcribing the tokens as 'voiced' or 'voiceless'. Note that I used the proportion of voiceless interval (the original fraction of locally unvoiced frames), as opposed to converting it to the proportion of voiced interval. This is to have the measure of voicing point in the same direction as duration: a lower proportion of voiceless interval (i.e. more voiced) and a shorter duration both indicate a more lenited token.

When a lenis obstruent token had a value of 0 in the proportion of voiceless interval, it indicated that it was fully voiced; and when it had a value of 1, it indicated that it was fully voiceless. Therefore, with respect to the 'Optionality' question, if there is a PW-medial lenis obstruent that is 100% voiceless after a vowel that is 0% voiceless (i.e., fully voiced), then such a token would be a clear exception of the generalization in (3). The results for these measurements are presented in §2.4.2.

2.3.3 Parameterizing the degree of reduction and duration

While it is common to use continuous measures for duration (e.g., Cho & Keating, 2001) and voicing (Jun, 1993; Han, 2000), qualitative measures have been more frequently used for reduction (Lee, 1995; Yoo, 2020; Colantoni & Marinescu, 2010; Mansfield, 2015). In past literature, visual criteria have been applied to identify whether an obstruent is reduced (fricativized or approximantized), which includes the presence or absence of certain phonetic features that characterize reduction, such as burst, voicing, visible formants above F1 (Bouavichith & Davidson, 2013; Mansfield, 2015). Manual inspection requires a great deal of human hand annotation, meaning it becomes quickly expensive to get the data annotated as the size of the data gets bigger.

However, besides this practical issue, Ennever et al. (2017) has summarized three potential problems of taking measurements by hand. First, consistently discerning the boundaries of phones is nearly impossible without an automatic algorithm. As Ladefoged (2003) notes, segments often 'don't have clear beginnings and ends' and as Turk et al. (2006) argues, perhaps in this sense the segmentation of spontaneous speech into segments is an 'artificial task'. The quantification of lenition and the segmentation of the obstruent are bound to be interdependent since any attempt to quantify the lenition of a segment presupposes that the researcher knows where to measure from, i.e. the segment boundaries. Second, different types of segments often require using different phonetic correlates for segmentation which causes difficulties in trying to investigate lenition across different segments. Finally, qualitative measures, in specific, are hard to reproduce, since the observation of the segments depends on the specific details in the software settings for displaying the waveform and the spectrograms. Ennever et al. (2017) has shown that the same token can be considered lenited or not, depending on the Praat setting. As Katz & Pitzanti (2019) notes, labeling using qualitative criteria necessarily introduces a lot of subjectivity of the labelers, and they state that it is almost like conducting a binary forced-choice experiment on the labelers.

In contrast, Kingston (2008) proposed an algorithm of performing the segmentation and lenition quantification using the intensity, which was later modified to be used for investigating larger data sets of other languages: Gurindji (Ennever et al., 2017) and Campidanese (Katz & Pitzanti, 2019). This study uses a blended version of Kingston's original procedure and the improvements made in the later work, which we will call the 'Smoothed Kingston' algorithm, to increase the robustness of the method. This version is almost identical to the one used in Katz (2021) except for the choice of frequency band to bandpass filter the signal. The Smoothed Kingston is designed to measure the degree of lenition of an obstruent token, by measuring the velocity of intensity change from the interval between the preceding vowel to the obstruent. In the intervocalic position, a full stop token is produced with complete blocking of airflow, which causes a rapid fall in the intensity compared to the preceding vowel, followed by a rapid increase in the intensity as the following vowel starts. In contrast, when an obstruent is realized with an incomplete closure (i.e. reduced), such changes in the velocity of intensity change would be less extreme.

To demonstrate how the Smoothed Kingston algorithm quantifies the degree of obstruent lenition as well as discover the segmental boundaries, a token of lenis stop /k/ in the data is shown in Figure 2.5 with the parameterization results of the algorithm. The original waveform, spectrogram, and annotation of this token is presented in Figure 2.4.

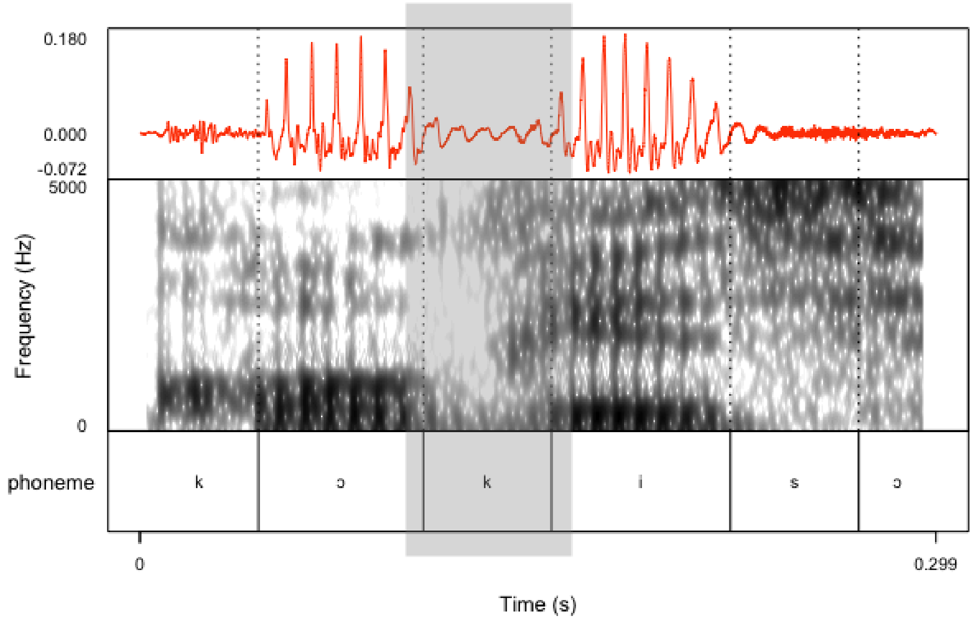


Figure 2.4: Waveform, spectrogram, and phone-level annotation of a PW-medial voiced /k/, taken from the Seoul Corpus (Yun et al., 2015). InputFile: s04m15m5, /k/ starts at t = 179.57818. /kɔkisɔ/ ('there').

The token was taken from the Eojeol: /kɔ.**ki**.sɔ/ ('there'), and it is the second /k/ in bold. Since it was an Eojeol medial token, it was expected to be voiced, reduced, and shorter. The solid lines labeled as 'A' and 'B' are boundaries marked in the corpus annotation file, whereas the dashed lines labeled as 'C' and 'D' are the boundaries identified by the algorithm. The solid line labeled as '*Pit*' is where the intensity is the lowest within the consonant boundaries.

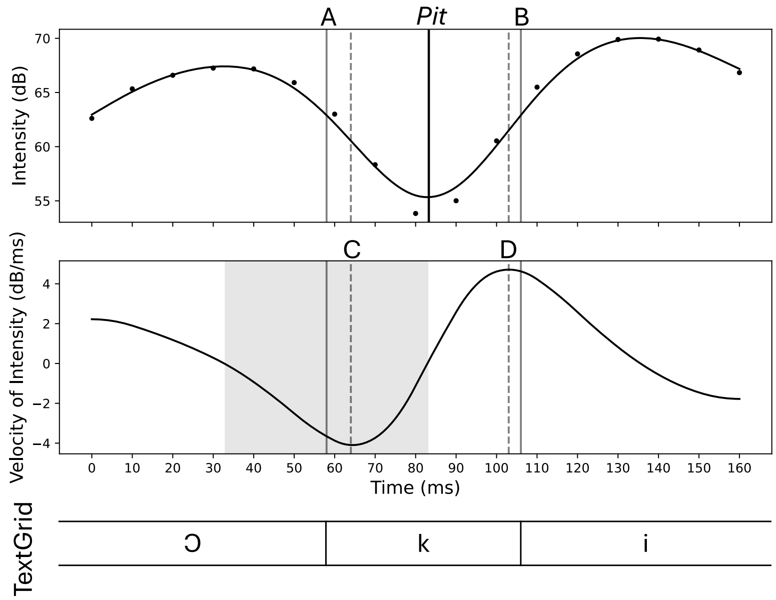


Figure 2.5: A demonstration of how a PW-medial /k/ token in between /ɔ/ and /i/ vowels is parameterized using the Smoothed Kingston algorithm. The consonantal interval marked as 'A'-'B' is based on the consonant boundaries in the corpus annotation file. The consonantal interval marked as 'C'-'D' is from the consonant boundaries identified by the algorithm. The degree of reduction is the velocity of intensity change at point 'C'. This point is searched for within the interval marked with dark gray. The *Pit* refers to the point where the intensity is the lowest within the consonant.

First, the acoustic signal of the entire sound file was bandpass filtered into different frequency bands. This study experimented with three bands: 0-400 Hz ('Low band'), 400-1200 Hz ('Mid band'), and 1200-3200 Hz ('High band') and found the low band yields the optimal parameterization. For how the parameters of this procedure were evaluated, including this choice of frequency band, see Supplementary Material 'Choosing the right method and parameters to quantify the degree of reduction' in the the OSF repository. From the bandpass filtered acoustic signal, the intensity of the energy was extracted using the *ToIntensity* function in Praat. Then, the algorithm looked for the obstruent tokens in the intervocalic position using the annotation file and for each token, it looked for the manually marked consonant boundaries on the Phone tier on each side of the segment. These boundaries from the corpus annotation are shown as solid vertical lines ('A' and 'B'), across all panes in Figure 2.5. In contrast to Kingston's algorithm, the algorithm used in Ennever et al. (2017) and Katz & Pitzanti (2019) used only the left boundary of the segment. This choice of using only the left boundary affected the robustness of the algorithm to consistently find the boundaries for the segment, and therefore this study used both boundaries (see Supplementary Material 'Choosing the right method and parameters to quantify the degree of reduction' for a demonstration of how this decision might matter and see also Katz (2021) for similar opinion (p. 667)).

A user-specified small margin was added on both sides of the segment interval, such that the intensity measurements from the end of the preceding vowel and the beginning of the following vowels could be included. After some experimentation, I used 25 ms as the margin which is half of the 50 ms margin used in (Kingston, 2008), as 50 ms was found to be too wide (see Supplementary Material 'Choosing the right method and parameters to quantify the degree of reduction' in the OSF repository for more details). This margin-padded segmental interval worked as the search window over which we found relevant acoustic correlates that quantify the degree of lenition as well as the boundaries of the given obstruent token. This margin on the left boundary of the segment is marked as the gray interval in Figure 2.5.

The intensity measurements taken from the search window associated with each obstruent token were used to compute the velocity of intensity over time. While Kingston (2008) computed the velocity by first differencing the intensity measurements over time, later variants in Ennever et al. (2017) and Katz & Pitzanti (2019) used a spline smoothing to get a smoothed intensity contour prior to computing the derivative. This study took the second approach of smoothing the intensity contour before taking its derivative (to compute the velocity), hence 'Smoothed' in the name of the algorithm 'Smoothed Kingston'. The spline smoothing was done using the '*smooth.spline*' function in the R Stats package (version 3.6.2) (R Core Team, 2023). A user-specified smoothing parameter '*spar*' was used to decide how smooth the curve will be. A higher smoothing parameter means heavier smoothing, which tends to misrepresent the contour if the contour changes abruptly. A lower smoothing parameter on the other hand means possibly overrepresenting the contour, which may contain measurement outliers. The velocity contour was then computed by taking the derivative of the smoothed intensity contour.

From these contours, the following inflection points were found. First, from the intensity contour (the top pane of Figure 2.5), the *Pit* was identified, by finding the minimum intensity value within the search interval. The assumption that intervocalic obstruents were produced with a closure (either complete or incomplete if it is lenited) implied that the intensity contour measured from the 'VCV' interval would be U-shaped, as illustrated in Figure 2.5. In the typical case where the intensity contour was U-shaped, the *Pit* could be identified somewhere in the middle of the segment, as in Figure 2.5. However, in a case where the segment was lenited to the extent that it has virtually the same intensity energy compared to the neighboring vowels, then this *Pit* might not be found at all. This latter case was the failure condition of the algorithm used in (Ennever et al., 2017; Katz & Pitzanti, 2019) and was also used in this study to exclude tokens that were too lenited to determine their boundaries. A very conservative criterion was used to decide which tokens were excluded in this study: a token was discarded if the algorithm failed to find its *Pit* for one of the three frequency bands (high, mid, and low band) I experimented on. This made sure that the remaining tokens had the expected U-shape intensity contour. This strict criterion is likely to exclude heavily lenited obstruents which, as a result, are disproportionately left out of the analysis. However, it is not a problem for addressing my research questions. Recall that the first research question asks if reduction is optional. If it turns out that reduction seems to have happened for tokens that are not heavily lenited, then it should provide evidence that reduction is not optional.

From the velocity contour (the middle panel of Figure 2.5), two inflection points were identified: the most extreme falling velocity (point 'C' on Figure 2.5) near the left boundary and the most extreme rising velocity (point 'D' on Figure 2.5) near the right boundary. Crucially these extrema were found within the user-specified margin (the dark gray interval in Figure 2.5) near the manually marked boundaries. When a token was realized as extremely shortened, a large margin would accidentally cause the most extreme rising velocity to be found near the left boundary, instead of the right boundary. Kingston (2008) reported that for around 9-10% of his data, the extreme rising velocity preceded the extreme rising velocity in time, and discarded those tokens. The same criterion was applied to the present data.

Finally, the tokens that had a non-negative minimum velocity value were removed because it meant that there was no decrease in intensity compared to the preceding vowel. For these two filtering criteria, the conservative method was used again, that is if the algorithm failed on one of the three frequency bands, then the token was discarded.

Table 2.1 summarizes the number of tokens that were removed for each step during this parameterization procedure. Notice that a single token could potentially have both the problem of having no *Pit* and having the wrong order of the maximum velocity and the minimum velocity. The table shows the counts of the tokens that were removed from the data set by the time that the filtering criterion was applied and therefore the numbers in Table 2.1 do not double count the number of tokens that had more than one reason to be removed. There were considerably more lenis obstruent tokens that were excluded compared to the other two categories. A total of 16,449 obstruent tokens were removed. While this seems to be a large number, it amounts to only 11% of the initial number of 150,088. This rate is comparable to the percentage of tokens removed from the final analysis in previous studies: around 10% in Kingston (2008), 8% in Katz & Pitzanti (2019), and 7% in Ennever et al. (2017), despite applying a more strict criterion for a token to be deemed unparameterizable. Some of the reasons why the parameterization might fail for some tokens are discussed in depth in the Supplementary Material 'Choosing the right method and parameters to quantify the degree of reduction'.

Out of the tokens that were removed, 15,489 of them were lenis, 396 were fortis, and 564 were aspirated. Lenis obstruents were disproportionately excluded partly because there were a lot more lenis obstruents than the other two categories to begin with, but also because only lenis obstruents were expected to be realized as reduced (fricativized and approximantized).

Filtering criterion	Number of excluded tokens
No identifiable *Pit*	16,382
Non-negative value of CVE	51
RVE preceding CVE	12
Proportion of voiceless intervals error	4
Total number of tokens removed	16,449

Table 2.1: Table summarizing the number of tokens that are removed from the data because of the parameterization issue. For each step, the number of tokens is counted from the data that resulted from the previous filtering step.

The extreme falling velocity value (the Y-axis value of point 'C' in Figure 2.5) is named the 'Closure Velocity Extremum' (CVE) because it is expected to align with the consonantal constriction, as the energy falls fast from the preceding vowel. Likewise, the extreme rising velocity value (the Y-axis value of point 'D' in Figure 2.5) is named the 'Release velocity Extremum' (RVE) because is expected to align with the end of the constriction and the beginning of the following vowel, as the energy rises fast as the vowel gets produced. The time points of CVE and RVE are then used as the updated, phonetically grounded left and right segment boundaries of the obstruent token. This resulted in a slightly shorter duration (39 ms, the interval between 'C' and 'D') compared to the corpus duration (48 ms, the interval between

'A' and 'B') for the given token in Figure 2.5, but it was not always the case that the corpus duration was longer than the found duration.

Kingston (2008) used the CVE, the duration, and the RVE to quantify the degree of lenition as more extreme values of these indicated that the token was less lenited because an actual stop (non-lenited) would have a rapid decrease in intensity (*CVE*) with a full occlusion, which would be sustained for a longer period (the updated duration), followed by a burst of airflow causing a rapid increase in intensity (*RVE*). Parameters that could potentially cause differences in the parameterization of the lenis obstruents were set after some experimentation (see Supplementary Material 'Choosing the right method and parameters to quantify the degree of reduction' for details). The bandpass filter was set to the low-frequency band (0-400 Hz), and the smoothing parameter (*spar*) was set to 0.4.

Finally, 4 additional tokens were removed from the dataset since the proportion of voicing measurement returned *NA* values for their neighboring vowels, as these vowels were completely devoiced.

	Lenis	Aspirated	Fortis
Initial	36,801	4,575	5,341
Medial	54,311	14,644	17,967
Total	91,112 (68.2%)	19,219 (14.4%)	23,308 (17.4%)

Table 2.2: Number of lenis, aspirated and fortis onsets from Eojeol initial and Eojeol medial positions

The remaining 133,639 tokens are summarized for each of the three laryngeal categories and two Eojeol positions (Initial or Medial) in Table 2.2.[4] These obstruent tokens were all measured in terms of their proportion of voiceless intervals, the proportion of voiceless intervals of their preceding vowels, the CVE, and duration.

[4]Note these were all in the onset position, therefore there were no Eojeol-final obstruents.

There were notably more lenis onsets (68.2%) across the two positions in Eojeol compared to the other two categories (14.4% for aspirated and 17.4% for fortis). The number of lenis obstruent tokens that are of principal interest in this study far exceeds the number of tokens investigated in previous work that uses a similar method to acoustically quantify the degree of lenition (Kingston, 2008; Ennever et al., 2017; Katz & Pitzanti, 2019), which was in the order of hundreds for the former two and was about 4,000 in Katz & Pitzanti (2019). As noted, this is an advantage of investigating a speech corpus, rather than speech collected from a production experiment, as a larger number of tokens provide higher statistical power.

2.4 Distribution of phonetic variables in different prosodic positions

This section presents the distribution of obstruent tokens in the phonetic variables explained in §2.3: proportion of voicing, CVE, and duration. I explore the distribution of lenis obstruents in these phonetic variables to address the research questions: Are AP-medial voicing, shortening and reduction of Seoul Korean lenis obstruents optional?

While the other two laryngeal categories are not expected to have different values across different Eojeol positions, lenis obstruents are expected to have different values for these phonetic variables, across the two Eojeol (PW) positions: initial and medial.

As discussed in §2.2, it is expected that there would be two kinds of Eojeol initial tokens: AP-initial and AP-medial. On the other hand, the Eojeol-medial lenis obstruents are uniformly AP-medial and hence expected to be voiced, reduced, and shortened, compared to the Eojeol-initial tokens. Yet, there might be exceptional Eojeol-medial lenis obstruents that might be fully voiceless, less reduced, and longer, compared to other Eojeol-medial tokens, if voicing, reduction, and shortening are optional in the AP-medial position. If these exceptional Eojeol-medial lenis obstruent tokens exist, they would be visible from the distribution.

2.4.1 Prosodic conditioning of voicing

We start with the distribution of the proportion of voiceless interval. As mentioned in §2.3.2, I defined the variable as the proportion of 'voiceless' interval, and therefore a value of 0 would indicate a fully voiced realization, whereas a value of 1 would indicate a fully voiceless one.

The data measured from the corpus replicate the optionality and gradience of AP-medial lenis obstruent voicing found previously in lab speech (a.o., Jun, 1994, 1998; Han, 2000).

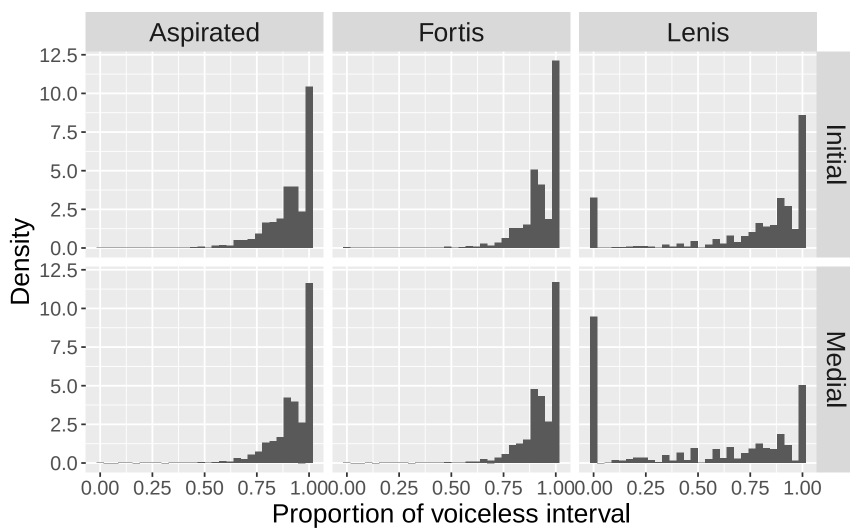


Figure 2.6: Histograms showing the density of the proportion of voiceless interval measured from the obstruent interval. The top row shows the measurements from Eojeol initial position, and the bottom row shows the measurements from Eojeol medial position. The columns separate out the three laryngeal categories.

Figure 2.6 plots the density of the proportion of voiceless interval, for each of the three laryngeal categories in Eojeol initial and medial positions. It is quite noticeable that while the aspirated and the fortis obstruents are either fully voiceless (i.e., the proportion of voiceless interval is 1 or the right end of the continuum on the figure) or voiceless for more than half of the interval in both Eojeol initial and medial positions, the lenis obstruents vary greatly by Eojeol position. There are many more fully voiceless tokens than fully voiced tokens in the initial position while the opposite is true in the medial position. There were lenis obstruent tokens with intermediate values in both positions.

The distribution for lenis obstruents in Figure 2.6 in both Eojeol initial and medial positions can be characterized as a near-categorical bimodal distribution with some noise (the intermediate values). There was a clear tendency that they were more likely to be fully voiceless in the Eojeol initial position, compared to the Eojeol medial position. The bimodality that Eojeol-initial tokens exhibits for voicing likely represents that some of these Eojeol-initial tokens are AP-initial while some of them are AP-medial. However, it was also found that Eojeol medial lenis obstruents were sometimes fully voiceless, replicating previous findings that argued for the optionality of voicing (Yun, 2000; Han, 2000).

For example, Yun (2000) showed that the variation of this percentage of voicing variable was huge for the PW-medial lenis obstruents, and could range from 0% (completely voiceless) to 100% (fully voiced) across places of articulations. Han (2000) reported that one of her participants produced PW-medial lenis as fully voiced (Voicing percentage = 100%) only 34% of the time. However, the same speaker produced PW-medial lenis as at least partially voiced (i.e. in between 0% and 100%) 70% of the time.

2.4.2 Voicing of lenis obstruent influenced by voicing of the preceding vowel

As noted in §2.3.2, the same voicing measure was obtained from the preceding vowel to see if the voicelessness of an Eojeol medial lenis obstruent could be attributed to the fact that the previous vowel is devoiced (Jun, 1994).

	Fully voiced lenis obstruents	
Preceding Vowel	Eojeol initial	Eojeol medial
Fully voiced	3,916 (94%)	16,057 (91%)
Intermediate	209 (6%)	1,625 (9%)
Fully devoiced	6 (0%)	11 (0%)
Total	4,131	17,693
	Fully voiceless lenis obstruents	
Preceding Vowel	Eojeol initial	Eojeol medial
Fully devoiced	969 (9%)	1,556 (17%)
Intermediate	6,616 (61%)	4,915 (52%)
Fully voiced	3,331 (31%)	2,931 (31%)
Total	10,916	9,402

Table 2.3: Fully voiced/voiceless lenis obstruents (N = 42,142) by Eojeol position and by the voicing of the preceding vowel.

Table 2.3 presents the number of lenis obstruent tokens for each position in the Eojeol (initial or medial). To simplify the description, tokens with intermediate voicing values were not included, i.e., the ones with intermediate values in Figure 2.6. These fully voiced or fully voiceless lenis obstruents are divided according to the voicing status of the preceding vowel: fully voiced, intermediate or fully voiceless. Just as lenis obstruent tokens, if the preceding vowel had a value of 0 in the proportion of voiceless interval, it was considered as 'Fully voiced' and if it had a value of 1, it was considered as 'Fully voiceless'. All other vowel tokens were labeled as 'Intermediate'. The fully voiced lenis obstruent tokens are presented on the top, and the fully voiceless lenis obstruent tokens are presented on the bottom. Rounded percentages were calculated for each column in the sub-table and presented in parentheses.

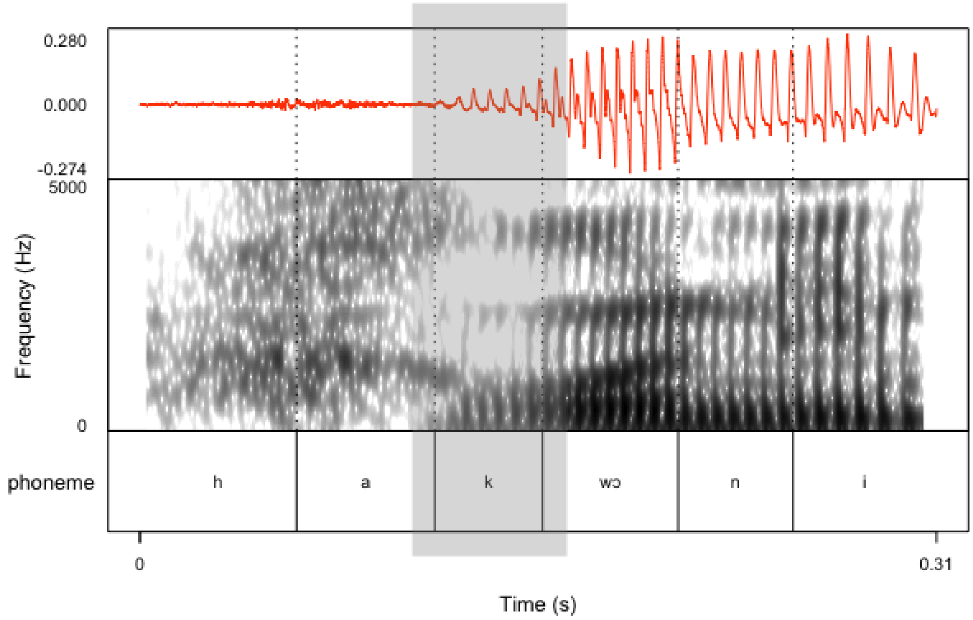


Figure 2.7: Waveform, spectrogram, and phone-level annotation of a **PW-medial voiced /k/ after a devoiced vowel**, taken from the Seoul Corpus (Yun et al., 2015). InputFile: s02m16f2, /k/ starts at t = 423.8264. /hakwɔn-i/ ('academy-NOM').

The fully voiced lenis obstruents were predominantly preceded by fully voiced vowels across Eojeol positions, and vanishingly few tokens were fully voiced after a fully devoiced vowel (17 tokens) across the two Eojeol positions. Figure 2.7 shows an example of such a case. After the PW-initial segment /h/, the vowel /a/ is devoiced. Nonetheless, the subsequent /k/ is realized as voiced and reduced. While these tokens suggest an interesting phonological interaction between vowel devoicing and lenis voicing, I leave this topic for future studies.

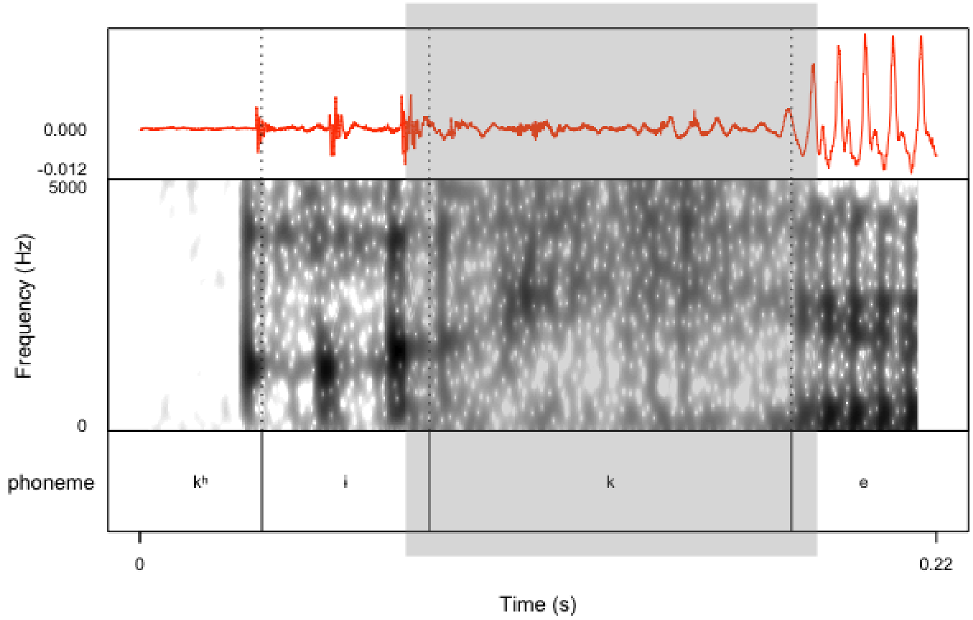


Figure 2.8: Waveform, spectrogram, and phone-level annotation of a **PW-medial voiceless /k/ after a devoiced vowel**, taken from the Seoul Corpus (Yun et al., 2015). InputFile: s01m16f2, /k/ starts at t = 167.45604. /k^hɨke/ ('greatly').

In contrast, a higher proportion of lenis obstruents were fully voiceless after fully devoiced vowels in the Eojeol medial position (17%) than in the Eojeol initial position (9%). In other words, close to 17% of the fully voiceless Eojeol medial lenis obstruents were predictably voiceless, due to the devoicing of the preceding vowel. An example of such tokens is in Figure 2.8. The velar stop (highlighted) in this example is preceded by a clearly devoiced vowel, which itself is preceded by an aspirated stop.

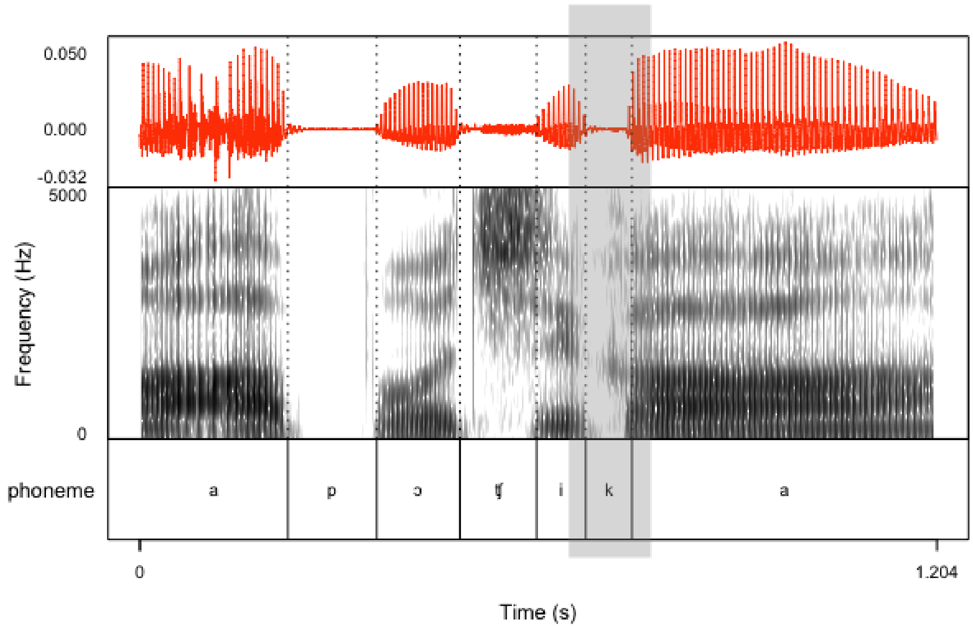


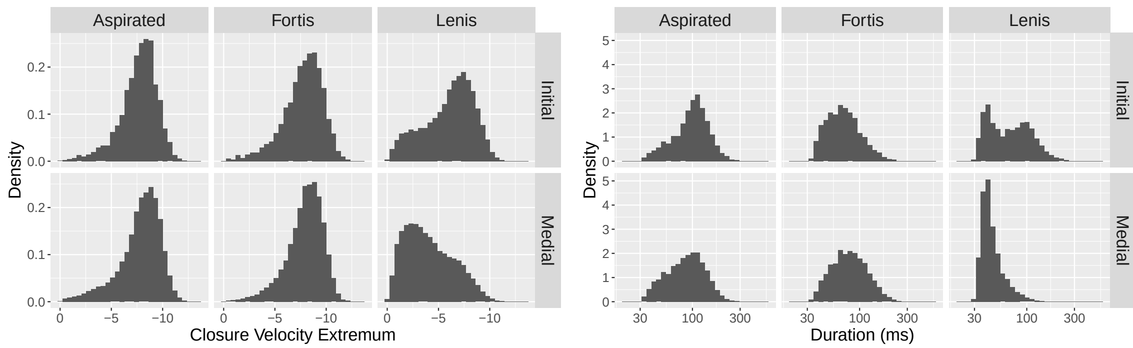
Figure 2.9: Waveform, spectrogram, and phone-level annotation of a **PW-medial voiceless /k/ after a voiced vowel**, taken from the Seoul Corpus (Yun et al., 2015). InputFile: s01m16f1, /k/ starts at t = 65.52562. /apɔʧi-ka/ ('fater-NOM').

Nonetheless, there were lenis obstruents tokens that were realized as fully voiceless, despite being preceded by a fully voiced vowel across Eojeol positions. Therefore, there seems to be optionality in voicing lenis obstruents, even after the influence of the preceding vowel is considered. An example of this kind of token is in Figure 2.9, where the PW-medial lenis velar stop /k/ is completely voiceless after a fully voiced vowel /i/.

The results here replicated the findings reported in Jun (1994), on a much larger scale, and using data taken from a spontaneous speech corpus. However, it remains to be seen if these exceptionally voiceless tokens are prosodically affected in other ways than voicing. This will be addressed in §2.5.

2.4.3 Prosodic conditioning of reduction and shortening

Now we turn to the other two phonetic measurements: CVE and duration, and see if such exceptional tokens also exist in the distribution of these two variables.



(a) CVE (b) Duration

Figure 2.10: Histograms showing the density of the absolute value of CVE and the log-transformed duration measured from the obstruent interval. The top row shows the measurements from Eojeol initial position, and the bottom row shows the measurements from Eojeol medial position. The columns separate out the three laryngeal categories.

First, Figure 2.10 plots CVE (Figure 2.10a) and the duration (Figure 2.10b), for each of the three laryngeal categories in Eojeol initial and medial positions. Notice that the horizontal axis for CVE was flipped, so that visually the 'stronger' realization of lenis obstruents (voiceless, less reduced, and longer) would be on the right end of the continuum, like for the proportion of voiceless interval in Figure 2.6. The plots for duration show the original values on a log-scaled horizontal axis since the raw duration was heavily skewed.

As we've seen for voicing, Figure 2.10 shows that again only the distribution of lenis obstruents changes across two Eojeol positions (top vs. bottom rows), while the other two laryngeal categories remain unchanged. Figure 2.10 shows that lenis obstruents in the Eojeol medial position were more reduced and shorter compared to the ones in the Eojeol initial position, i.e. the mode of the distribution is closer to the left of the continuum.

As for the distribution of Eojeol initial lenis obstruents, it was not as clearly bimodal for CVE as it was for the duration. As with voicing, the bimodality for duration likely represents that some of these Eojeol-initial tokens are AP-initial while some of them are AP-medial. In contrast, as for the distribution of Eojeol medial tokens, there is similar bimodality in voicing (Figure 2.6), but not in CVE (Figure 2.10a) or duration (Figure 2.10b).

2.4.4 Interim summary of results

In this section, I have presented the distribution of the three phonetic variables for each of the three laryngeal categories in the Eojeol-initial and Eojeol-medial positions. The distributions suggest that the prosodic conditioning of voicing, reduction, and shortening seems to be mainly restricted to the lenis obstruents since the other two laryngeal categories remained relatively unchanged across Eojeol positions.

In §2.4.2, I have replicated the previous findings that lenis obstruent voicing is optional even in the Eojeol medial position, which is by definition AP-medial (Jun, 1993; Han, 2000). I have also replicated that some of these seemingly optionally voiceless tokens are predictably voiceless, due to the preceding devoiced vowels, but also that there are Eojeol-medially fully voiceless tokens after fully voiced vowels (Jun, 1994).

These results suggest that regarding the 'Optionality of lenition' research question, while lenis obstruents are optionally voiced in the PW medial position, some of the lack of voicing is explainable from the previous vowel devoicing. In contrast to voicing, I have shown that optionality of reduction and shortening of lenis obstruents in the PW-medial position is less clear since the distributions of the medial lenis obstruents in Figure 2.10a and Figure 2.10b were not as notably bimodal, compared to the distribution of proportion of voiceless interval of lenis obstruents in the PW-medial position.

2.5 Lenition of 'exceptional' Eojeol medial voiceless lenis obstruents

This section provides statistical analysis that addresses the 'optionality of lenition' research question, about the optionality of reduction and shortening. The second research question is restated in different terms, in (7).

For this research question, I test whether the Eojeol-medial lenis obstruents with exceptional voicing differ phonetically from the Eojoel-initial tokens with the same voicing profile, namely fully voiceless after a fully voiced vowel. Fully voiceless Eojeol-initial lenis obstruents after a fully voiced vowel are possibly also AP-initial. On the other hand, Eojeol-medial lenis obstruents are AP-medial, despite having the same voicing profile.

The optional prosodic effect hypothesis predicts a null hypothesis which states that these two groups do not differ significantly in degree of lenition. In contrast, the obligatory prosodic effect hypothesis predicts that Eojeol-medial lenis obstruents are more lenited than Eojeol-initial ones.

(7) Is there a difference in degree of lenition between PW-initial and PW-medial lenis obstruents with the same voicing profile?

a. V $]_{PW}[$ Lenis$_{[-voice]}$ V: PW-initial, but also possibly AP-initial

b. $_{PW}[$ V Lenis$_{[-voice]}$ V]: Exceptional PW-medial

c. Optional prosodic effect: (a) and (b) do not differ in degree of lenition.

d. Obligatory prosodic effect: (b) is more lenited than (a) because (b) is AP-medial PW-medial

To test these hypotheses, the data set was filtered so that it contained only the tokens that were fully voiceless after a fully voiced vowel. These are the tokens labeled as 'fully voiceless' and preceded by 'fully voiced' vowels in Table 2.3. This resulted in a data set of 6,262 lenis obstruents, where 3,331 tokens were in the Eojeol initial position and 2,931 were in the Eojeol medial position (tokens like the one in Figure 2.9).

Two linear mixed-effects models were fit using the `lme4` package in R (Bates et al., 2015): one for the degree of reduction and one for duration. The p-values were calculated using the '`lmerTest`' package (Kuznetsova et al., 2017). I used an alpha level of .05 for all statistical tests.

§2.5.1 discusses other predictors included in the model. §2.5.2 explains the model structure and the process for model selection. §2.5.3 presents the results. I show that the 'unpredictably voiceless' lenis obstruents, i.e. fully voiceless Eojeol-medial lenis obstruents preceded by a fully voiced vowel, are nonetheless reduced and shorter than fully voiceless Eojeol-initial tokens. This suggests that unlike voicing, reduction and shortening are not optional within the AP, and therefore the segmental realization can reliably signal the prosodic position.

2.5.1 Predictors

The response variables were the negative Closure Velocity Extremum (henceforth `CVE`), and the log-transformed duration (henceforth `Dur`). Negative CVE values were used instead of the original ones so as to keep the direction of effect consistent with `Dur`: lower CVE means more lenited, just as lower `Dur` means more lenited. The duration variable (`Dur`) was log-transformed because it was heavily skewed. `CVE` was included as a predictor for the model predicting `Dur`, and vice versa.

The prosodic position (`EojeolPosition`), was contrast-coded as a binary variable with 'Eojeol-initial' being 0.5 and 'Eojeol Non-initial' (i.e., Eojeol-medial) being -0.5. As we've seen in §2.4, I expected more lenited realization of lenis obstruents (lower values in `CVE` and `Dur`) in Eojeol medial position.

The speaker's gender (`Gender`) was contrast-coded as a binary variable with 'female' being 0.5 and 'male' being -0.5. Previous work on Korean has not specifically reported variation in lenis realization due to extralinguistic factors like age and sex. It has been shown for other languages that male speakers are more likely to produce fully voiced stops than female speakers (Swartz, 1992; Ryalls et al., 1997; Puggaard-Rode et al., 2022). Puggaard-Rode et al. further explained that the differences due to the speaker's sex could be due to biological and/or sociolinguistic factors. First, it might be due to the biological nature of males' larger supralaryngeal cavities, which allows for the closure voicing to last over a longer period of time (Puggaard-Rode et al., 2022). Alternatively, this might be related to the sociolinguistic tendency that male-gendered speakers tend to speak less carefully, which would cause the stops to be more lenited (voiced). I expected more lenited realization of lenis obstruents (lower values in `CVE`, and `Dur`) for male speakers based on these results.

The speech rate (`SpeechRate`) was measured as the number of syllables uttered per second in the utterance that each token was taken from. Recall that the utterance-level segmentation was provided in the corpus annotation files as pause-delimited chunks. A higher value in this speech rate variable indicated a faster speech rate because it indicated that the speaker uttered more syllables per second. This speech rate variable was log-transformed. I expected more lenited realization of lenis obstruents (lower values in `CVE` and `Dur`) in faster speech. `CVE`, `Dur`, and `SpeechRate` were z-scored across speakers.

I expected the lenis obstruents to vary in their phonetic realization as a function of the place of articulation. `Place` was coded as a custom-coded fixed effect term (velar vs. others, postalveolar vs. alveolars and bilabials, and alveolars vs. bilabials). Little work has looked at the effect of the place of articulation on the realization of lenis obstruents partially because of the methodological obstacle of determining the consonant boundaries. However, Lee (1995) found that in her production experiment, the velar lenis /k/ was realized as an approximant significantly more often than the other stops (/p, t/). Lee argued that velar lenis obstruents were more reduced because unlike other consonants, /k/ shares the primary articulatory gestural variable (i.e., Tongue body) with the flanking vowels and therefore is more likely to be affected by articulatory gestures of the flanking vowels. There is even less work on the lenition of postalveolar affricate, except it was previously noted that they were the least consistently voiced within the AP (Jun, 1995).

Other than `EojeolPosition`, investigating the variation due to other predictors (`Gender`, `SpeechRate`, and `Place`) is not the central topic of the present paper, but these factors are included in the statistical models as controls. Interaction terms of each of these predictors with `EojeolPosition` were included.

As for the random effect structure, random intercepts for Eojeol and speaker were included, and random slopes for each of the other predictors, except `Gender`, were included. The resulting model formula is in (8).

(8) Model formula predicting `X`. When `X` is `CVE`, `Y` is `Dur`, and vice versa.

```
X ~ EojeolPosition + SpeechRate + Place + Gender + Y
+ EojeolPosition:SpeechRate
+ EojeolPosition:Place
+ EojeolPosition:Gender
+ EojeolPosition:Y
+ (1|Eojeol)
+ (1|Speaker)
+ 0 + (EojeolPosition|Speaker)
+ 0 + (Place|Speaker)
+ 0 + (SpeechRate|Speaker)
+ 0 + (Y|Speaker)
```

2.5.2 Model construction and selection

The `CVE` and `Dur` models were fit with the formula in (8). The models converged with a singular fit, which can indicate the random effect structure is overly complicated (Brauer & Curtin, 2018). The random effect terms were checked with `ranova()` function in the `lmerTest` package (Kuznetsova et al., 2017), which returns likelihood ratio tests to assess the significance of including each random effect term, compared to a null model without any random effect terms. The results are in Table 2.4 and Table 2.5.

Term	LogLikelihood	AIC	p-value
Baseline (no random effects)	-7265.5	14595	
`(1|Eojeol)`	-7310.7	14683	< 0.0001 ***
`(1|Speaker)`	-7265.5	14593	0.9991
`0 + (EojeolPosition|Speaker)`	-7308.1	14674	< 0.0001 ***
`0 + (Place|Speaker)`	-7289.5	14623	< 0.0001 ***
`0 + (SpeechRate|Speaker)`	-7265.5	14593	0.9977
`0 + (Y|Speaker)`	-7273.2	14608	< 0.0001 ***

Table 2.4: The results of 'Ranova' test for the `CVE` model

Term	LogLikelihood	AIC	p-value
Baseline (no random effects)	-7567.9	15200	
`(1|Eojeol)`	-7613.5	15289	< 0.0001 ***
`(1|Speaker)`	-7567.8	15198	1.0000
`0 + (EojeolPosition|Speaker)`	-7584.5	15227	< 0.0001 ***
`0 + (Place|Speaker)`	-7576.6	15197	0.0666 .
`0 + (SpeechRate|Speaker)`	-7569.1	15200	0.1236
`0 + (Y|Speaker)`	-7575.9	15214	< 0.0001 ***

Table 2.5: The results of 'Ranova' test for the `Dur` model

For the `CVE` model, `(1|Speaker)` and `0 + (SpeechRate|Speaker)` did not improve the model fit. This indicated that there were no speaker-specific difference in the `CVE` on average, and no speaker-specific difference in the speech rate effect on `CVE`.

For the `Dur` model, `(1|Speaker)` and `0 + (SpeechRate|Speaker)`, as well as `(1|Speaker)` and `0 + (Place|Speaker)` did not improve the model fit. It indicated that there were no speaker-specific difference in the `Dur` on average, and no speaker-specific difference in the speech rate or the place of articulation effect on `Dur`. These terms were thus excluded from the final models.

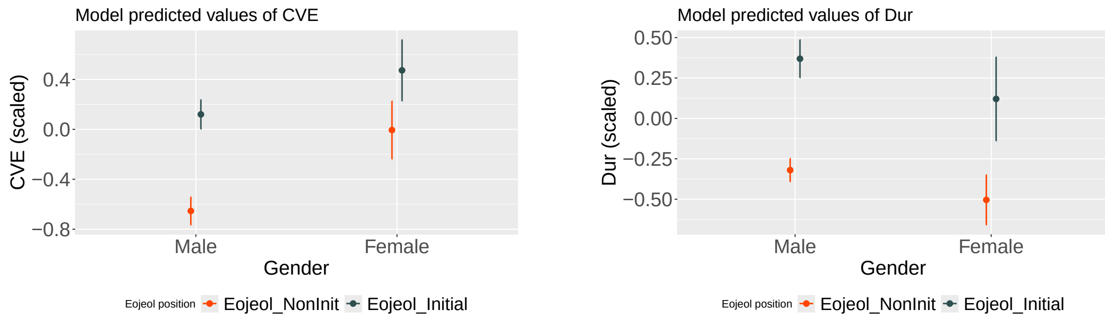
2.5.3 Result

This section presents the model results. As noted before, the only relevant predictors are `EojeolPosition` and the interaction terms involving that predictor. In both models, as predicted, `EojeolPosition` were significant (`CVE` model: $\beta = 0.403$, Std. Error = 0.055, t value = 7.319, p-value < 0.0001; `Dur` model: $\beta = 0.652$, Std. Error = 0.047, t value = 13.934, p-value < 0.0001). Fully voiceless Lenis obstruents after a fully voiced vowel were less reduced and longer in the Eojeol-initial position, compared to the Eojeol-medial position.

In what follows, the results of the interaction terms are reported. To interprete the model results, I use partial effect plots, which show the predicted effect of varying the predictor in the interaction on the response variable, while all other predictors are held constant. Model predictions in the plots were computed using the fixed efffect coefficient estimates, and the error bars represent 95% confidence intervals. The full results of the models are reported in §2.9.

These interaction terms were also tested with a post-hoc pairwise comparison test using the '`emmeans`' package (Lenth, 2018), which tested whether the difference between the two levels of `EojeolPosition` was significant across different values of the other predictor in the interaction. Estimates and standard errors of the model predicted difference between the two levels of `EojeolPosition` (Eojeol Non-initial - Eojeol Initial) are reported, while varying the other predictor in the interaction. P-values were corrected using the Tukey Honestly Significant Difference method.

2.5.3.1 Gender and EojeolPosition



(a) Model of CVE (b) Model of Dur

Figure 2.11: Partial effect plots showing estimated means of the response variables, while varying Gender.

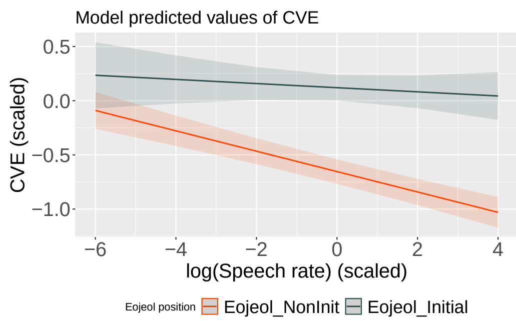
Post-hoc pairwise comparison results Contrast: Eojeol Non-initial - Eojeol Initial				
		CVE model		
Gender	Estimate	Standard Error	z ratio	p-value
Male	-0.550	0.0738	-7.455	< 0.0001
Female	-0.255	0.0751	-3.397	0.0007
		Dur model		
Gender	Estimate	Standard Error	z ratio	p-value
Male	-0.706	0.059	-11.973	< 0.0001
Female	-0.598	0.059	-10.116	< 0.0001

Table 2.6: Post-hoc pairwise comparison results for the interaction of Gender and EojeolPosition. Estimates show the model predicted difference between the two levels of EojeolPosition (Eojeol Non-initial - Eojeol Initial).

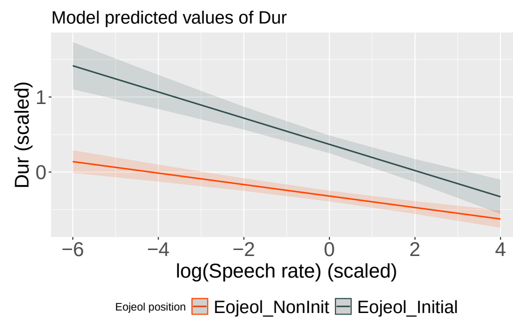
We start with Gender:EojeolPosition. A main effect of Gender was significant in both models (CVE model: β = 0.501, Std. Error = 0.104, t value = 4.818, p-value $<$ 0.0001; Dur model: β = -0.217 , Std. Error = 0.0697, t value = -3.108, p-value = 0.004) but the effects were in the opposite direction. Overall, female speakers produced less reduced lenis obstruents but their lenis obstruents were shorter. Partial effect plots in Figure 2.11 show the interaction of EojeolPosition and Gender. Both CVE and Dur were lower in the Eojeol-medial position than in the Eojeol-initial position, for both genders.

This was confirmed in the post-hoc pairwise comparison tests, which are reported in Table 2.6. The results show that the difference between the two levels of `EojeolPosition` was significant for both genders, and the estimates show that the Eojeol-initial lenis obstruents were less reduced and longer than the Eojeol-medial lenis obstruents.

2.5.3.2 Speech rate and EojeolPosition



(a) Model of `CVE`



(b) Model of `Dur`

Figure 2.12: Partial effect plots showing estimated means of the response variables, while varying `SpeechRate`.

The next interaction term is `SpeechRate:EojeolPosition`. A main effect of `SpeechRate` was significant in both models (`CVE` model: β = -0.057, Std. Error = 0.011, t value = -5.172, p-value $<$ 0.001; `Dur` model: β = -0.126, Std. Error = 0.011, t value = -10.979, p-value $<$ 0.001) indicating that lenis obstruents were significantly less reduced and longer in slower speech. Partial effect plots in Figure 2.12 show the interaction of `EojeolPosition` and `SpeechRate`. Both `CVE` and `Dur` were always lower in the Eojeol-medial position than in the Eojeol-initial position, regardless of the speech rate.

The interaction of `SpeechRate` and `EojeolPosition` was different in the two models. In the `CVE` model, the difference between the two Eojeol positions was greater when the speech rate was faster, while in the `Dur` model, it was greater when the speech rate was slower. The post-hoc pairwise comparison tests are reported in Table 2.7. Since SpeechRate was a continuous variable, the estimates and standard errors are reported for the two levels of EojeolPosition at the speech rate at the 5th and 95th percentiles of SpeechRate. The two percentiles were chosen to show the effect of EojeolPosition at the two extremes of the speech rate.

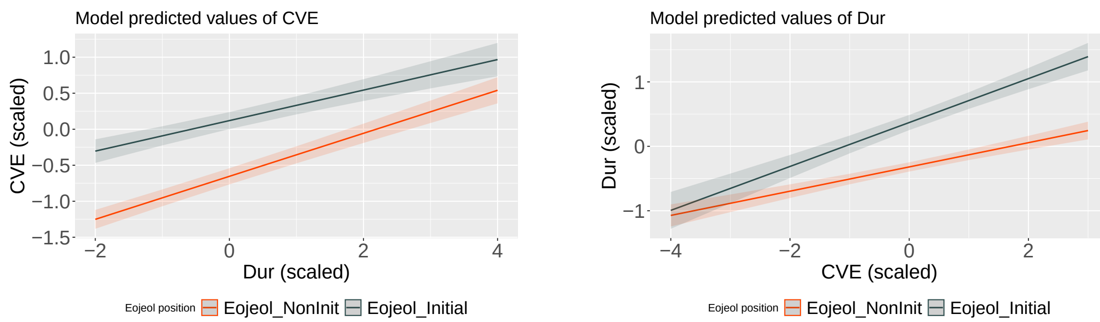
Post-hoc pairwise comparison results				
Contrast: Eojeol Non-initial - Eojeol Initial				
CVE model				
SpeechRate	Estimate	Standard Error	z ratio	p-value
5th Percentile	-0.246	0.0664	-3.703	0.0002
95th Percentile	-0.490	0.0662	-7.400	<.0001
Dur model				
SpeechRate	Estimate	Standard Error	z ratio	p-value
5th Percentile	-0.793	0.0607	-13.074	<.0001
95th Percentile	-0.474	0.0584	-8.120	<.0001

Table 2.7: Post-hoc pairwise comparison results for the interaction of SpeechRate and EojeolPosition. Estimates show the model predicted difference between the two levels of EojeolPosition (Eojeol Non-initial - Eojeol Initial).

The results confirmed that difference between the Eojeol positions was significant regardless the speech rate. The estimates indicated that in faster speech (95th Percentile), Eojeol-medial lenis obstruents were much more reduced than Eojeol-initial ones, while in slower speech (5th Percentile), the difference between the two levels of EojeolPosition was smaller. In contrast, in slower speech, Eojeol-initial lenis obstruents were much longer than Eojeol-medial ones, while this difference was smaller in faster speech.

2.5.3.3 Dur/CVE and EojeolPosition

The next interaction terms are Dur:EojeolPosition and CVE:EojeolPosition, which are in the CVE and Dur models, respectively. Main effect of Dur/CVE was significant in the respective model (CVE model: Dur β = -0.057, Std. Error = 0.011, t value = -5.172, p-value $<$ 0.001; Dur model: CVE β = -0.126, Std. Error = 0.011, t value = -10.979, p-value $<$ 0.001) indicating that these two measures are significant predictors of each other. Partial effect plots in Figure 2.13 show the interaction of `EojeolPosition` and `Dur`/`CVE`. The interaction of `CVE` and `EojeolPosition` in the `Dur` model was different from the interaction of `Dur` and `EojeolPosition` in the `CVE` model. The effect of `EojeolPosition` on `CVE` and `Dur` was significant, regardless of the other measure of lenition in general, except when `CVE` was very low in the `Dur` model, the difference between the two levels of `EojeolPosition` was very small. On the other hand, when `Dur` was high in the `CVE` model, the difference between the two levels of `EojeolPosition` was smaller.



(a) Model of `CVE` (b) Model of `Dur`

Figure 2.13: Partial effect plots showing estimated means of the response variables, while varying the other measure of lenition: `Dur` or `CVE`

This trend was confirmed in the post-hoc pairwise comparison tests, which are reported in Table 2.8. As before, the estimates and standard errors are reported for the two levels of `EojeolPosition` at the speech rate at the 5th and 95th percentiles of `Dur`/`CVE`.

The results mirrored the interaction between `SpeechRate` and `EojeolPosition`. In a faster speech rate, duration in general becomes shorter. This corresponds to the case where `Dur` is low (5th Percentile) in the `CVE` model, and the difference between the two levels of `EojeolPosition` was larger, as it was the case for faster speech. In a slower speech rate, where `Dur` is high (95th Percentile), the difference between the two levels of `EojeolPosition` was smaller, as it was the case for slower speech. Putting these results together, in slower speech, Eojeol-initial and medial tokens are similarly

Post-hoc pairwise comparison results				
Contrast: Eojeol Non-initial - Eojeol Initial				
CVE model				
Dur	Estimate	Standard Error	z ratio	p-value
5th Percentile	-0.499	0.0613	-8.133	<.0001
95th Percentile	-0.237	0.0756	-3.136	0.0017
Dur model				
CVE	Estimate	Standard Error	z ratio	p-value
5th Percentile	-0.379	0.0655	-5.784	<.0001
95th Percentile	-0.888	0.0609	-14.585	<.0001

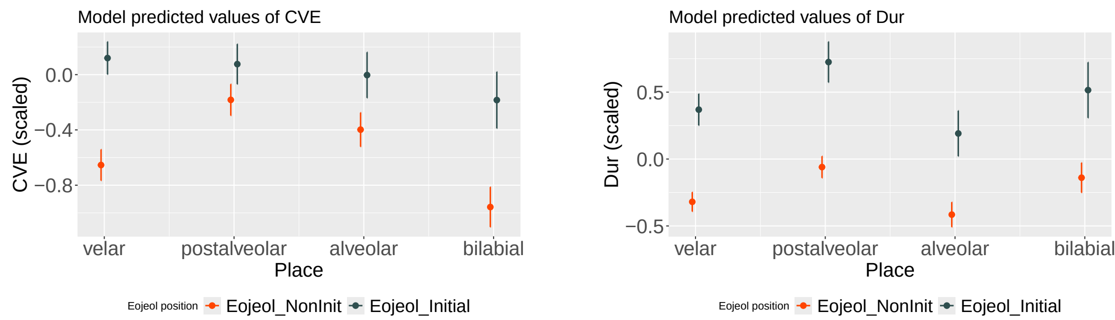
Table 2.8: Post-hoc pairwise comparison results for the interaction of `Dur` and `EojeolPosition` for the `CVE` model, and the interaction of `CVE` and `EojeolPosition` for the `Dur` model. Estimates show the model predicted difference between the two levels of `EojeolPosition` (Eojeol Non-initial - Eojeol Initial).

pronounced without reduction (high `CVE`), but they differ in duration: Eojeol-initial tokens are longer. In faster speech, they are similarly short, but they are distinct in degree of reduction: Eojeol-initial tokens are reduced less. The results are discussed further in §2.6.

2.5.3.4 Place and EojeolPosition

Finally, we move onto the interaction between `Place` and `EojeolPosition`. Since `Place` had multiple levels, interpreting the effects of `Place` would be more convenient with the partial effect plots. Partial effect plots in Figure 2.14 show the interaction of `EojeolPosition` and `Place`. Most importantly, in both models, Eojeol-initial lenis obstruents were always less reduced and longer than medial ones, regardless of the place of articulation. The only exception was the postalveolar lenis affricate in the `CVE` model. Though it was not significant, the numerical trend was in the same direction as the other places of articulation: Eojeol-medial tokens tended to be more reduced than Eojeol-initial ones.

This was confirmed in the post-hoc pairwise comparison tests, which are reported in Table 2.9. The effect of `Place` was different in the two models, as well as the interactions. The full results are available in the Appendix in §2.9. Here, only the



(a) Model of `CVE` (b) Model of `Dur`

Figure 2.14: Partial effect plots showing estimated means of the response variables, while varying `Place`

interaction of `Place` and `EojeolPosition` is discussed, that is, how the difference between the two Eojeol positions varies as a function of the place of articulation. The difference between the two Eojeol positions in `CVE` was the greater for velar and bilabial stops, while it was smaller for postalveolar affricate and alveolar stops. In contrast, the difference between the two Eojeol positions in `Dur` was the greater for velar stops and postalveolar affricates, while it was smaller for alveloar and bilabial stops. These results suggest that the effect of prosodic position on the lenition of lenis obstruents may not be uniform across the place of articulation. However, for the purpose of addressing the research question, the important finding is that the Eojeol-initial lenis obstruents were almost always less reduced and longer than the Eojeol-medial ones, regardless of the place of articulation, with the exception of the postalveolar lenis affricate in the `CVE` model.

2.6 Discussion

This study investigated three prosodically conditioned phonetic properties of lenis obstruents in spontaneous speech–voicing, reduction, and duration. There were two research questions: (1) whether previous results finding fully voiceless PW-medial lenis obstruents can be replicated in a spontaneous speech corpus (2) whether these exceptionally voiceless tokens are nonetheless more reduced and shorter in comparison

Post-hoc pairwise comparison results				
Contrast: Eojeol Non-initial - Eojeol Initial				
		`CVE` model		
`Place`	Estimate	Standard Error	z ratio	p-value
Velar	-0.626	0.0586	-10.692	<.0001
Postalveolar	-0.111	0.0649	-1.709	0.0874
Alveolar	-0.247	0.0647	-3.822	0.0001
Bilabial	-0.626	0.1032	-6.069	< 0.0001
		`Dur` model		
`Place`	Estimate	Standard Error	z ratio	p-value
Velar	-0.657	0.0519	-12.656	<.0001
Postalveolar	-0.753	0.0593	-12.700	<.0001
Alveolar	-0.574	0.0581	-9.890	<.0001
Bilabial	-0.622	0.1011	-6.157	<.0001

Table 2.9: Post-hoc pairwise comparison results for the interaction of `Gender` and `EojeolPosition`. Estimates show the model predicted difference between the two levels of `EojeolPosition` (Eojeol Non-initial - Eojeol Initial).

to PW-initial lenis obstruents with the same voicing profile (fully voiceless after a fully voiced vowel).

The results in §2.4 first showed that all three phonetic properties of lenis obstruents, but not for other two laryngeal categories, varied as a function of Eojeol position: lenis obstruents were more voiced, reduced, and shorter in Eojeol-medial position than in Eojeol-initial position. These prosodically dependent acoustic realizations support the view the variation in the phonetic realization of lenis obstruent should be described in terms of 'continuity lenition' (Katz, 2016).

As described in §2.1, from this perspective, obstruent lenition serves the role of signaling to the listeners that the current prosodic constituent is continuing (Harris & Urua, 2001; Kingston, 2008; Ennever et al., 2017; Katz & Fricke, 2018). For instance, by realizing the stop as an approximant within a word, the acoustic signal of the word as a prosodic constituent is interrupted less, compared to the case where the stop was produced as a stop which would necessarily introduce a closure, hence a disruption in the intensity contour. Under this view, it is also reasonable to account for the results of fully voiceless lenis obstruents after a fully devoiced vowel (Jun, 1994). When the preceding vowel is devoiced, realizing the following obstruent as voiceless disrupts signal less, and therefore being voiceless here fits the description of the continuity lenition better, as opposed to being voiced. Replicating the findings in Jun (1994), some of the fully voiceless lenis obstruents in the Eojeol medial position were preceded by fully devoiced vowels (§2.4.2). On the contrary, there were indeed vanishingly rare tokens in the corpus that were fully voiced after a fully devoiced vowel.

As reported previously (Jun, 1994), however, not every Eojeol-medially fully voiceless token was preceded by a devoiced vowel. The voicing of lenis obstruent was optional within the Eojeol (Jun, 1994; Han, 2000), as there were fully voiceless tokens after fully voiced vowels (§2.4.2). If the prosodically conditioned realization of lenis obstruent is a 'continuity lenition', we would further expect these seemingly exceptional Eojeol medial lenis obstruents that are fully voiceless after fully voiced vowels to be also lenited, compared to the tokens found in the Eojeol initial position. I found that they were indeed more reduced and shorter compared to the fully voiceless lenis obstruents that were also after fully voiced vowels but in the Eojeol initial position (§2.5).

The optionality of voicing therefore does not suggest the prosodic effect is optional, as lenis obstruents are always more lenited in the Eojeol medial position, compared to the Eojeol initial position. In light of this non-optional lenition of lenis obstruent in the prosodic medial position, we might assume that the segmental realization as a whole reliably signals the prosodic boundary to the listener while a single process (voicing) may be optional and unreliable. The results presented here call for future investigation into other so-called prosodically conditioned 'optional' segmental processes.

Given the findings that the lenis obstruents are in general 'lenited' in the prosodic constituent medial position, in one or more of the phonetic dimensions, it may be reasonable to somehow combine these phonetic measurements to describe how they vary as a function of the prosodic position. Dalcher (2007), for instance, performed the Principal Component Analysis (PCA), which is a common dimensionality reduction technique to describe the lenition patterns in Florentine Italian. More recently, Lee (2023) argued that the segmental realization of lenis obstruents, first measured in terms of voicing, intensity, and duration, and then combined using the PCA, must be considered together with the tonal transitions to reliably detect the prosodic boundary types proposed in the K-ToBI model, highlighting the importance of the segmental realization in marking the prosodic boundaries. These approaches suggest that listeners might not track one specific dimension of the segmental realization (e.g., voicing) but multiple phonetic dimensions can cause a percept of continuation of the current prosodic constituent, much like the fact that the percept of pitch is much more than just an F0 value.

These results are also in line with experimental works that has shown listeners are sensitive to such prosodically conditioned segmental realization (Cho et al., 2007; Yoo, 2020). Listeners are faster at lexical identification when the segmental realization indicates a larger, as opposed to a smaller, prosodic boundary (Cho et al., 2007) and they can use the segmental realization to disambiguate between sentences that differ only in prosodic phrasing (Yoo, 2020). In particular, Yoo (2020) demonstrated that Korean listeners pay attention to the acoustic realization of lenis obstruents in sentence disambiguation when the audio stimuli are artificially monotonized.

The results of this study are limited in that only fully voiceless lenis obstruents from different Eojeol positions were compared. If the fully voiced ones are included in the data, there might be additional variation that was not revealed in the current study. I leave this for future studies. One particular aspect of the data that needs further investigation is the fully voiced PW-initial lenis obstruents. They are predicted to be AP-medial PW-initial, according to the generalization that lenis obstruents are AP-medially voiced, but some of them might be exceptions to the generalization, i.e., fully voiced in the AP-initial position. In Chapter 3, I explore a subset of these cases.

2.7 Chapter summary

In conclusion, this study investigated the acoustic realization of lenis obstruents in a Seoul Korean spontaneous speech corpus. Previous work on this topic largely focused on voicing or duration, modulated by the speech rate and the prosodic position. Though it has been reported that lenis obstruents also become reduced, which is measurable as slower changes in intensity, such realization has not been investigated in depth. The major contribution of this study is that it investigated all three of these aspects in the acoustic realization of lenis obstruents in a spontaneous speech corpus, which was both large in the number of investigated tokens and arguably close to natural speech that speakers produced, compared to the data collected in production experiments in most previous work. I argue that the prosodic conditioning of lenis obstruent exemplifies 'continuity lenition', serving the role of signaling a continuation of the prosodic constituent (Kingston, 2008; Katz & Pitzanti, 2019).

I replicated previous findings that the voicing of lenis obstruents is optionally applied in the PW-medial position, and cannot always be predicted from the preceding vowel devoicing (Jun, 1994). However, I also demonstrated that these exceptionally voiceless PW-medial lenis obstruent tokens were nonetheless reduced and shorter, compared to the PW-initial tokens, thus indicating that the prosodic conditioning of the segmental realization might not be optional. This non-optionality suggests that the segmental realization may prove to be a reliable cue for prosodic phrasing, and promotes a more holistic approach for future lenition studies, as a single phonetic aspect (i.e. voicing) does not provide the larger picture of prosodically conditioned lenition.

2.8 Supplementary materials

Two sets of supplementary material named 'Choosing the right method and parameters to quantify the degree of reduction' and 'Random effects' are available at `https://osf.io/ukh6d/?view_only=d9bda726aebe4512830c6996f2ae4cae`.

The same repository hosts Praat, Python, and R scripts for processing the corpus annotation file and audio files, generating the figures used in this study. The scripts are explained in the document titled 'Doing phonetic research on the Seoul Corpus'.

2.9 Appendix: Model results

	CVE	
Predictor	Variance	Standard Deviation
Eojeol		
(Intercept)	0.051	0.226
Speaker		
EojeolPosition		
Initial	0.158	0.397
Non-initial	0.067	0.259
Dur		
	0.005	0.074
Place		
Velar	0.015	0.124
Postalveolar	0.011	0.106
Alveolar	0.022	0.150
Bilabial	0.131	0.362
	Dur	
Predictor	Variance	Standard Deviation
Eojeol		
(Intercept)	0.055	0.234
Speaker		
EojeolPosition		
Initial	0.043	0.208
Non-initial	0.061	0.247
CVE		
	0.008	0.088

Table 2.10: Random effects for CVE and Dur models.

	Dependent variable:	
	CVE	Dur
EojeolPosition (+Initial)	0.403*** (0.055)	0.652*** (0.047)
Place (k vs. textteshlig t p)	0.008 (0.033)	−0.111*** (0.027)
Place (textteshlig vs. t p)	0.333*** (0.049)	0.294*** (0.036)
Place (t vs. p)	0.370*** (0.071)	−0.300*** (0.052)
SpeechRate	−0.057*** (0.011)	−0.126*** (0.011)
Gender (+Female)	0.501*** (0.104)	−0.217*** (0.070)
Dur	0.255*** (0.018)	
CVE		0.264*** (0.020)
Initial:Place (k vs. textteshligtp)	0.298*** (0.052)	0.007 (0.054)
Initial:Place (textteshlig vs. tp) (0.069)	−0.326***	0.154** (0.071)
Initial:Place (t vs. p)	−0.379*** (0.102)	−0.048 (0.105)
Initial:SpeechRate	0.075*** (0.022)	−0.098*** (0.023)
Initial:Gender	−0.295*** (0.100)	−0.064 (0.083)
Initial:Dur	−0.087*** (0.026)	
Initial:CVE		0.153*** (0.025)
Constant	−0.023 (0.057)	−0.0001 (0.037)
Observations	6,262	6,262
Log Likelihood	−7,265.539	−7,578.178
Akaike Inf. Crit.	14,591.080	15,196.360
Bayesian Inf. Crit.	14,793.340	15,331.200
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 2.11: Full model results of `CVE` and `Dur` models. Standard Errors are in parentheses.

CHAPTER 3

PROSODIC ENCODING IN SPONTANEOUS SPEECH CORPUS: LENIS AND ASPIRATED OBSTRUENTS OF SEOUL KOREAN

3.1 Introduction

This chapter continues to investigate phonetic correlates of prosodic constituent junctures, i.e., the phonetic instantiation of Break Indices labels, focusing specifically on PW-initial syllables that start either with a lenis or aspirated stop/affricate (henceforth Lenis PW and Aspirated PW, respectively) with the same spontaneous speech corpus (Yun et al., 2015) used in the previous chapter.

Studying prosody with a spontaneous speech corpus poses several challenges. It is most likely that the corpus is not prosodically transcribed, which is often considered to be the essential preprocessing step before conducting a study of prosody. It also contains unpredictable speech materials that vary tremendously, possibly containing lots of disfluencies. This study presents a novel approach for studying prosodic junctures in a spontaneous speech corpus, which may be generalized to any large corpora. Instead of labeling the prosodic junctures first and investigating the labels, this study parameterizes all syllables with their acoustic characteristics, which are known to vary as a function of prosodic position, and find the prosodic juncture categories in an unsupervised way (i.e., K-means clustering analysis). The identified juncture categories are then investigated further to see if the tokens in each juncture categories are acoustically realized in the way they are expected, given their position in the syntactic structure.

The main finding of this study is that the informativity of tonal and segmental cues varies depending on the laryngeal category. In distinguishing prosodic juncture categories, segmental cues are more informative for Lenis PWs, but tonal cues are more informative for Aspirated PWs. The PWs that have the acoustic characteristics of being phrase-medial are investigated in detail, and I show that their realization seem to be affected by their syntactic position.

The rest of this chapter is structured as follows. The concept of Break Indices in K-ToBI and the idea of investigating the informativity of prosodic cue in distinguishing prosodic juncture categories are reviewed in the rest of the current section. §3.2 reviews relevant previous works and discusses the research questions. §3.3 presents the methods. §3.4 and §3.5 discuss the main results. §3.6 provides further discussion, and §3.7 concludes the cshapter with a summary.

3.1.1 Break Indices

In the ToBI framework, prosodic constituent boundaries are associated with Break Indices (BI) (Silverman et al., 1992). As reviewed in Chapter 1, there are three levels of BIs associated with three levels of prosodic constituents in the K-ToBI model for Seoul Korean: PW, AP, and IP (Jun, 1993, 2000), as summarized in Table 1.2. Jun (2006) has later proposed there is a prosodic constituent category above the AP and below the IP, called the 'intermediate phrase (ip)'. Jun (2006) argues that ip boundaries have a BI level that is larger than 2 but smaller than 3. In addition to BI levels of 1, 2, and 3, there is a BI level of 0, which corresponds to syllable onsets within a PW (i.e., PW-medial syllable), and to "cases of 'incomplete nouns', monosyllabic nouns which are, though separated by spaces [in orthography], not used by themselves but need a modifier" (Jun, 2000). These incomplete nouns are discussed further in §3.2.3.

BI level	Prosodic position
3	IP-initial
2	AP-initial
1	PW-initial
0	PW non-initial / incomplete nouns

Table 3.1: Break Indices in K-ToBI. PW non-initial includes both PW medial and PW final syllables.

The K-ToBI labeling convention manual (Jun, 2000) explains that constituent boundaries are 'defined in the prosodic model by tonal markings', and these boundaries usually coincide with Break Indices, which are defined using 'the labeler's subjective sense of disjuncture'. The labeler's sense of disjuncture may depend on a number of factors, including tonal (e.g., Steffman et al., 2022) and segmental cues (e.g., Mitterer et al., 2021). However, in K-ToBI, the tone labels in fact are indicative of prosodic boundaries, which would make the BIs completely redundant. For example, we would know that there is an AP boundary between a pair of syllables if the second of the two syllables is marked with the initial tone 'L' and is preceded by a syllable marked with the final tone 'Ha', without even needing to look at the BI tier. Furthermore, as reviewed in Chapter 1, it has been argued that the medial tones ('+H', 'L+') are phonetically distinct from the edge tones ('Ha', 'L') (e.g., Jun, 1998; Lee & Lee, 2013).

The K-ToBI manual (Jun, 2000) mentions that the BI and the tones could mismatch sometimes which should be flagged on the BI tier with the diacritic of 'm' (Jun, 2000), meaning if a 2-like break does not coincide with an AP-like tone, this BI of 2 should be marked as '2m'. Also, the BI tier can have a diacritic of '-' which flags the labeler's uncertainty of the prosodic break strength. A BI of '2-' therefore is implying that the labeler knows that it is a 2-like break because of the tones, but perceives a weaker prosodic break. The fact that the uncertainties are marked on the

BI tier, rather than the tonal tier, implies that the tonal pattern is prioritized over the BI.

The trend to use a set of tonal labels that mark both the tonal target and also the prosodic constituency is prevalent in ToBI systems of other languages: 'the BI part of ToBI has also become less emphasized in the ToBI system' (Jun, 2022, p. 169).

3.1.2 Distribution of prosodic juncture categories over phonetic correlates

One way to make BI labels more objectively defined is to investigate how prosodic junctures are phonetically marked in a large data set. Once the phonetic correlates for each BI label are identified from the data set, a classificaion model may be trained to predict for novel speech data what BI labels a listener may assign. Such a classification model for speech perception has been implemented as a Bayesian inference model in prior studies (Clayards et al., 2008; Norris & McQueen, 2008; Kurumada & Roettger, 2022). These works have shown that listeners are able to adapt to an unfamiliar talker by making an inference of what the input signal could mean, essentially by comparing the signal with the expectation they built from previous experience, and thereby updating their expectations gradually, which can be predicted by a Bayesian model (Kleinschmidt & Jaeger, 2015; Xie et al., 2021; Kurumada & Roettger, 2022). These Bayesian models are closely related to distributional learning models, which use the distribution of categories along phonetic dimensions to discover phonetic categories (e.g., Maye et al. (2002); Feldman et al. (2009)). The Bayesian model of speech perception constantly updates beliefs based on the new input, in the same way that a model of learning updates the parameters so that the resulting categories better account for the data. Just as listeners would classify a particular speech sound as one speech category vs. another (e.g., /p/ vs. /b/), they would also classify a prosodic juncture between two syllables as one type vs. another (e.g., they belong to the same AP, or separate APs).

As Xie et al. (2021) remarks, however, it is impossible to model how listeners perceive a novel input using a Bayesian framework, without actually knowing what the typical distribution of the categories looks like, which would be the expectation of the listeners have built, from their previous experiences. However, there have not been many studies that investigate what the distribution of phonetic cues to prosody is (Xie et al., 2021). This is primarily due to the practical problem that labeling large speech corpora with prosodic transcriptions is costly and labor-intensive.

In this study, I present an alternative approach to study BI junctures in spontaneous speech corpora. Instead of labeling the speech with BI labels first, and then comparing the labels, I investigate the distribution of syllables on the phonetic correlates for the prosodic boundaries, and identify the 'prosodic juncture categories' using K-means clustering analysis (Lee, 2023). Each syllable in the data is parameterized with the same set of phonetic variables, and from the acoustic space formed by the phonetic variables, clusters are identified. The resulting clusters are the syllables that have similar phonetic profiles, and therefore expected to have the same level of BI. Specific syllables in the clusters can then be further examined to see if they are acoustically realized in the way that they are expected, given the syntactic relationship they have with surrounding PWs. This is analogous to a typical vowel identification study (e.g., Klein et al., 1970). A number of vowel intervals may be plotted on a phonetic space of formant values, from which clusters are identified using a classification algorithm. The resulting clusters are then further examined to see if they correspond to the expected vowel categories.

3.2 Background

3.2.1 Detection of AP boundaries in speech corpora

Previous work has attempted to train computational models to automatically identify prosodic constituents in speech corpora of Korean. To name a few, Jung et al. (2007b) scored over 90% precision and recall for IP boundaries but 72% Precision 64% recall for AP boundaries. Kim et al. (2021) used a deep learning model equipped with rule-based features, which include part of speech taggers and common collocations, and scored an accuracy of 93% for IP boundary detection but 70% for AP boundary detection. One common feature of these studies is that they used syntactic features as well as acoustic ones. For example, the fact that the nouns and the post-nominal case markers usually form their own AP was hard-coded into these models to detect AP boundaries. However, as mentioned above, the idea that the prosodic constituents in Seoul Korean are marked with segmental and supra-segmental features is central to the existing literature, and this study focuses on investigating the distribution of the prosodic juncture categories under this assumption. Once the prosodic juncture categories are identified, some of the categories are further studied to see if they are acoustically realized in the way they are expected. In particular, I will focus on PW-initial syllables that are realized with the acoustic characteristic of being AP-medial. To preview, the results show that their acoustic realization can largely be predicted by their syntactic position.

It is worth noting that the studies have achieved much higher detection scores for IP boundaries than AP boundaries. Given that the models could make use of morphological and syntactic features, it seems obvious that the IP boundaries are found more easily, as there are distinct sentence-final morphemes in Korean, such as the question particle or the honorific ending particle. However, the fact that the models achieved less accuracy on the AP boundaries suggests that there is something fundamentally difficult in detecting AP boundaries. If, for example, an AP always has the sequence of 'L +H L+ Ha', as the prototypical AP does, then such a F0 contour can be identified in a similar way that we can identify the unit cycle from a sinusoidal curve. However, there are several ways that an AP can surface differently, which makes them hard to identify as a coherent category.

First, an AP can consist of one or multiple PWds, which means it can theoretically be of any length in terms of number of syllables, though it has been reported that most APs consist of one PWd and usually not longer than 5 syllables and around 3 syllables long on average (Kim, 2004).

Second, there are fourteen sequences of tones that an AP can be made of, which are reported to be attested according to Jun (2000). When an AP does not start with an Aspirated or Fortis segment, the possible tonal sequences are: 'L Ha', 'L La', 'L +H Ha', 'L +H La', 'L L+ Ha', 'L +H L+ Ha', 'L +H L+ La'. When an AP starts with an Aspirated or Fortis segment, the initial tone of these sequences will be an 'H'.

Third, the number of syllables and the number of tones might not always match, since some medial syllables might not get any tones associated with them though there are tendencies and restrictions. Usually, APs don't get more tones than their number of syllables, except a mono-syllabic AP can have both initial and final tones. Also, while the middle tones '+H' and 'L+' are optional, the initial and the final tones are obligatory. Kim (2004) investigated how frequently each of these patterns occurs in two corpora. She reported out of 1541 APs that were supposed to be L-initial, 1465 were in fact L-initial (95.1% of the tokens), meaning there were some tokens that did not had H-inducing Aspirated or Fortis as the onset, and yet had a seemingly high F0. Out of those 1465 L-initial APs, 1201 had one of the three patterns: L Ha, L +H Ha, and L +H L+ Ha (82%). For these three popular AP patterns, Kim found that there was a tendency that APs were mostly shorter than 4 syllables. Only APs that were longer than 3 syllables had the canonical 'L +H L+ Ha' pattern. The most frequent form of AP is 3-syllable long with 'L Ha' pattern, which amounted to 31% of the entire data.

Fourth, the tonal pattern associated with the number of syllables is further affected by other factors such as the rate of speech (Cho & Flemming, 2015) and the casualness of speech (Yoo & Jun, 2016). Cho & Flemming (2015) reported that when speaking at a faster rate, speakers were more likely to compress a four syllable AP with 'L +H L+ Ha' as 'L +H Ha' and even 'LHa'. Similarly, Yoo & Jun (2016) transcribed 647 APs from 8 different types of speech style (TV drama, free conversation, article reading, story reading, interview, news, lecture and documentary). They found that 'L+' was found more often in more formal speech ('lecture' and 'documentary') as opposed to more spontaneous speech such ('TV drama' and 'free conversation').

Fifth, while typically each AP only dominates a single PW (Kim, 2004), Jun (1993) found that when the first PW of an AP was under focus, the subsequent PW was merged into the same AP. In spontaneous speech, where such focus conditions are not controlled, there might be numerous such cases where multiple PWs form one AP, due to this focused-induced dephrasing.

The remarkable versatility of tonal melodies of AP not only poses challenges to the models but also causes variation among the human transcribers in their judgment of the presence or absence of AP boundaries, as well as the exact tonal sequences. Jun et al. (2000) collected 20 sentences from 5 sources and asked 21 transcribers who differed in their level of expertise in transcribing K-ToBI, ranging from experts in K-ToBI to beginners in prosodic transcription. They reported that the agreement rate for the AP boundary type was 77.5% for all labelers, while IP boundary type was 90.9%. While experts and non-experts did not differ in their agreement rate for IP boundaries, the experts had a higher agreement rate for AP boundaries (81.9%) than all labelers (77.5%). In addition, the agreement rate for all tone labels and AP boundary location was only 52.2%, even among the expert group. The authors explain that the reason for this low agreement rate can be attributed to the fact that there are multiple possible patterns within the AP (e.g. L Ha can be confounded with L +H Ha).

These findings suggest that the AP boundaries are hard to identify with a consistent set of acoustic cues. It calls into question whether the proposed acoustic cues to AP boundaries allow for reliable differentiation between AP-initial and AP-medial syllables. I review the acoustic cues to AP boundaries in the next section.

3.2.2 Phonetic correlates of prosodic boundaries depending on the onset segment type

In this study, I focus on syllables with a lenis or aspirated stop/affricate onset. As I did in Chapter 2, I use the term 'obstruent' to refer to stops and affricates, though techincally the term includes fricatives (/s, s*, h/) as well. Lenis obstruents include /k, t, p, ʧ/ and aspirated obstruents include /k^h, t^h, p^h, $ʧ^h$/.

As mentioned, since APs are versatile in their tonal melodies, tonal cues may not always be deterministic in identifying whether there is an AP-level juncture. In addition to the tonal cues, lenis and aspirated segments exhibit variation in their acoustic realization conditioned on whether they appear at an AP-level juncture or not.

3.2.2.1 Syllables with a lenis obstruent onset

In terms of tonal cues, a lenis PW is AP-initial if it has an initial L tone. The preceding syllable is AP-final, assuming the Strict Layer Hypothesis (Selkirk, 1986), and is most likely to bear a phrase final Ha tone (Kim, 2004). Therefore, in terms of tones, a tonal fall from Ha to L may be associated with the AP-boundary. However, a tonal fall also occurs within an AP, from +H to L+. In a short AP, the syllable after +H may bear the +L tone, and phonetically have a lower F0 than the preceding syllable. In a longer AP, the syllable after +H would still have a phonetically lower

F0, as the tone falls until the AP penultimate syllable bearing L+ tone. This is schematized in Table 3.2.

Suppose that the task is to identify whether there is an AP-level prosodic juncture between two PWs. The first PW spans from Syllable 1 to 2, and the second one spans from 3 to 6. When there is an AP boundary between Syllables 2 and 3, the tonal labels suggest that there's a fall in F0 from Syllable 2 to 3 (Row (1a)). When there isn't an AP boundary between Syllables 2 and 3, the tonal labels suggest that the third syllable is unspecified for tone, which means that F0 falls until it reaches the 'L+' target in the fifth syllable (Row (2a)). As reviewed in Chapter 1, these two kinds of falls are distinct in phonetic outcome, that is the fall from Ha to L has been reported to be larger and steeper (Jun, 1998; Lee & Lee, 2013).

	Cue type	1	2	3	4	5	6
(1a. 2 APs)	Tonal	L	Ha	L	+H	L+	Ha
(1b. 2 APs)	Segmental			Strong			
(2a. 1 AP)	Tonal	L	+H			L+	Ha
(2b. 1 AP)	Segmental			Weak			

Table 3.2: Schematization of cases where two Lenis PWs form two APs (1), where the AP boundary lies between the second and the third syllables and where two Lenis PWs form a single AP (2). Expected tonal and segmental realization of syllables are presented.

In terms of segmental cues, a lenis PW is expected to be AP-initial if the initial segment is pronounced with a strong articulation (e.g., Jun, 1993; Cho & Keating, 2001). As shown in Chapter 2, the strength of a lenis onset can be measured using 'CVE' (i.e., the maximum falling velocity of intensity change from the preceding syllable), along with other phonetic measures such as proportion of voicing (e.g., Han, 2000; Davidson, 2016, 2018), or duration (Cho & Keating, 2001). In Table 3.2, as rows (1b) and (2b) show, the third syllable would have a more extreme value of CVE when there is an AP boundary (1b), compared to when there isn't (2b). As discussed in Chapter 2, acoustic realization of lenis obstruents has been argued to be optional

(Jun, 1994; Han, 2000), and therefore may not be the most reliable cue for the prosodic juncture perception (Jun, 1994). However, Chapter 2 has shown that this conclusion may be premature and misled by focusing on the phonetic measure quantifying voicing only, while lenis obstruents also vary significantly in their intensity, as a function of their prosodic position.

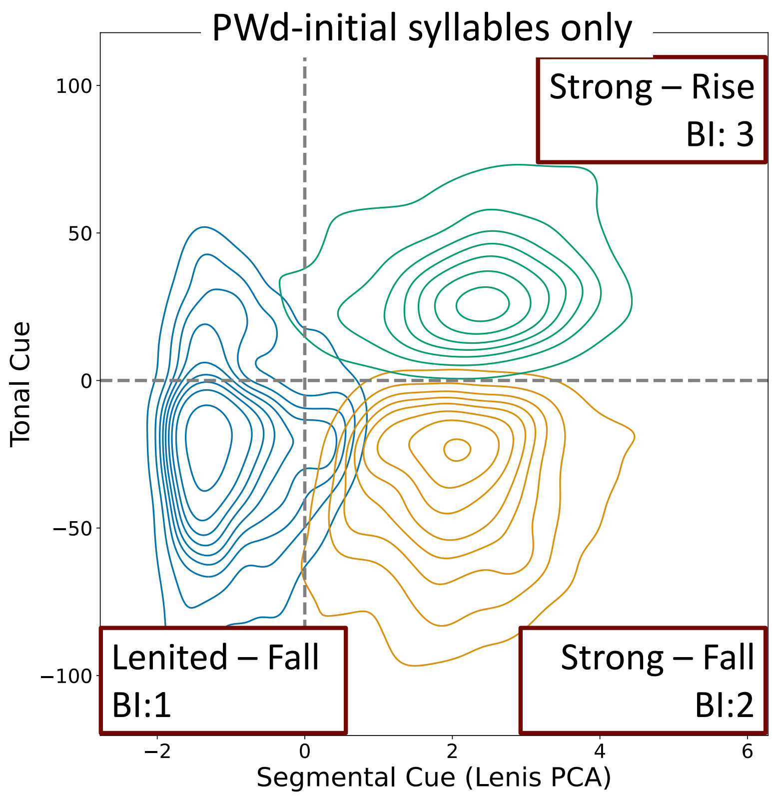


Figure 3.1: Figure 2 in Lee (2023) with extra annotation labeling the name of the clusters. See text for explanation.

Recently, Lee (2023) investigated Lenis PW initial syllables in a subset of the same spontaneous speech corpus (Yun et al., 2015) investigated in the current study. Details of how these cues were parameterized will be presented in §3.3. Each Lenis PW-initial syllable in the data was parameterized with a tonal cue quantifying the pitch change from its preceding syllable, and with a segmental cue quantifying the acoustic strength of the lenis onset. These Lenis PW-initial syllables were then plotted in a phonetic space defined over the segmental and tonal cues. From the distribution of Lenis

PW-initial syllables, three clusters were identified using a simple clustering analysis, i.e., K-means analysis. The characteristic findings of Lee (2023) are presented in Figure 3.1. Being higher on the tonal cue dimension means that a given syllable token has a higher F0 compared to the preceding syllable (i.e., tonal rise), and being lower means that it has a lower F0 compared to the preceding syllable (i.e. tonal fall). Being left on the segmental cue dimension means that the lenis onset of a given syllable is acoustically weak (i.e., lenited), and being right on this dimension means that it is acoustically strong (i.e., strong).

From the distribution of Lenis PW-initial syllables, the best number of clusters that is justified by the data turned out to be 3, which corresponds to the three levels of break indices that a PW-initial syllable be associated with: AP-medial PW-initial, AP-initial, and IP-initial. Lenis PW-initial syllables in the blue cluster in Figure 3.1 had segmental cue values that indicated that they were acoustically weak or lenited (Segmental cue below 0), and had lower F0 values compared to the preceding syllable (Tonal cue below 0). This phonetic profile would correspond to the third syllable in rows (2a) and (2b) in Table 3.2 above. Lee (2023) predicted that they would have BI values of 1 (PW boundary, AP-medial). This blue cluster spans wider vertically, and there are PW-initial syllables that have a higher F0 than the preceding syllable (i.e., 'Lenited Rise'). These tokens are overlooked in Lee (2023) as the majority of the blue cluster tokens may be characterized as 'Lenited Fall'.

Lenis PW-initial syllables in the yellow cluster in Figure 3.1 had segmental cue values that indicated that they were acoustically strong, and had lower F0 values compared to the preceding syllable. This phonetic profile would correspond to the third syllable in rows (1a) and (1b) in Table 3.2 above. Lee (2023) predicted that they would have BI values of 2 (AP boundary). If there was a difference in the size of phonetic fall between the tokens in the blue cluster and the ones in the yellow cluster, the center of the yellow cluster would be located lower than the center of the blue cluster. No such difference was visible from Figure 3.1 (Lee, 2023).

Finally, Lenis PW-initial syllables in the green cluster in Figure 3.1 had segmental cue values that indicated that they were acoustically strong (Segmental cue above 0), and had higher F0 values compared to the preceding syllable (Tonal cue above 0). Lee (2023) predicted that they would have BI values of 3 (larger than AP boundary), since a higher F0 value than a preceding syllable may be interpreted as a pitch reset (Jun, 2006), which is expected at the beginning of an ip or an IP. The strong realization of lenis onset accompanied with a pitch reset was argued to indicate a larger prosodic juncture, i.e., BI 3.

The findings in Lee (2023) suggest that tonal and segmental cues are both necessary to find expected three prosodic junctures categories from acoustically parameterized tokens of Lenis PW syllables. Crucially, contrary to previous findings that found minute tonal difference between AP-initial Lenis syllables and AP-medial Lenis syllables (Jun, 1998; Lee & Lee, 2013), there was no indication that the two kinds of junctures (blue and yellow in Figure 3.1) had different means in terms of the size of F0 fall.

However, Lee (2023) had some limitations, which are addressed in this study. First, only two speakers from the corpus (Yun et al., 2015) were investigated: one teenage female and a male in his forties since they were maximally different in terms of their age and gender. Distributions of values shown in Figure 3.1 were largely the same for the two speakers, but it needs to be confirmed that the same distribution holds for the entire corpus. Second, only lenis syllables were investigated, leaving out other laryngeal categories, e.g., aspirated obstruents. Third, more in-depth analyses of PWs in each cluster were not provided. In particular, considering that it is typical that each PW starts a new AP (Jun, 1993; Schafer & Jun, 2002; Kim, 2004), Lenis

PW-initial syllables in the blue cluster deserve a closer examination because these syllables do not seem to be AP-initial.

3.2.2.2 Syllables with an aspirated obstruent onset

Unlike initial syllables of Lenis PWs, initial syllables of Aspirated PWs are expected to be marked more distinctly by a local change in F0. To use the same schematization as in Table 3.2, the tone on the third syllable is expected to be notably high when there is an AP boundary (1a), compared to when there isn't an AP boundary (2a). Assuming that the initial H tone can be higher than the AP-final Ha, Aspirated PW-initial syllables that are AP-initial would have a higher F0 than the preceding syllable. In addition, another reason that the AP-initial syllable would also be expected to have a higher F0 than the preceding syllable is that it was previously found that the final tone of the AP preceding a H initial AP may be 'La' (Jun, 1996; Kim, 2004). On the other hand, when an Aspirated PW is AP-medial, F0 is expected to be lower than the preceding syllable (2a).

	Cue type	1	2	3	4	5	6
(1a. 2 APs)	Tonal	H	Ha	H		L+	Ha
(1b. 2 APs)	Segmental			Strong			
(2a. 1 AP)	Tonal	H	+H			L+	Ha
(2b. 1 AP)	Segmental			Weaker			

Table 3.3: Schematization of cases where two Aspirated PWs form two APs (1) and where two Aspirated PWs form a single AP (2). AP boundary appears between the second and the third syllables. Expected tonal and segmental realization of syllables are presented.

Aspirated obstruents have been found to vary in their VOT and articulation strength as a function of their prosodic position (Cho & Jun, 2000). Therefore, they may still vary in terms of segmental realization (1b and 2b), though this variation may be much smaller than the one observed for Lenis obstruents.

3.2.3 When does an AP have more than one PW?

As mentioned, the default prosodic phrasing of PWs is to have each one dominated by a single AP (Jun, 1993; Schafer & Jun, 2002; Kim, 2004). Underlying information structure such as contrastive focus can create a longer AP dominating multiple PWs (Jun, 1993; Schafer & Jun, 2002). Identifying the information structure of utterances in a speech corpus would be as labor-intensive as labeling them with prosodic transcription. Instead, a syntactic factor will be studied after finding the clusters of PW-initial syllables.

3.2.3.1 Syntactic restrictions on accentual phrasing

The influence of the syntactic structure on the prosodic structure has been noted from the beginning of the K-ToBI model Jun (1993). In particular, Jun defines the Accentual Phrase formation rules as in (1).

(1) The Accentual Phrasing rules (Example (16) on p.215 in Jun (1993))

- a. Every prosodic word may be an Accentual Phrase.
- b. A focused word must be the left-most word in an Accentual Phrase.
- c. An Accentual Phrase can include any number of prosodic words as long as:
  - (i) the last prosodic word is not the left element of a branching constituent.
  - (ii) all the prosodic words are not focused.

Jun (1993) shows that these rules explain the possible ways to prosodically phrase PWs into APs for a sentence with a nested right branching syntactic structure. The sentence in (2) is from her example (20) on p.219, which has the syntactic structure in Figure 3.2. Note that the tree is a simplified schematization of the syntactic structure of the sentence in (2) and hence lacks projections like TP.

(2) /(1)ʧʰamse-nɨn (2)ʧakɨn (3)namu-ɾɨl (4)ʧoahan-ta/

Sparrow-TOP small tree-ACC like-DECL
'A sparrow likes a small tree'

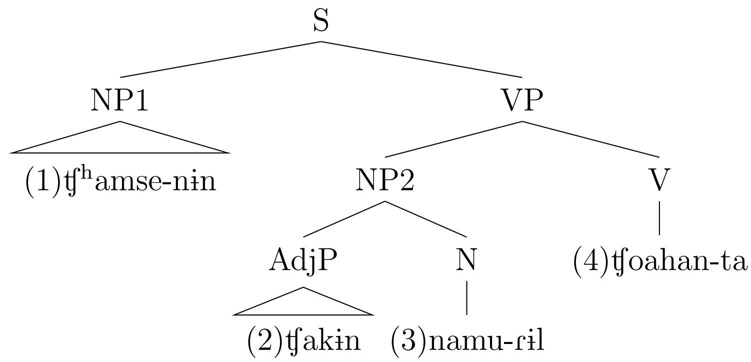


Figure 3.2: An example of a nested right branching structure (Example (20) on p.219 in Jun (1993), PWs are numbered by me.)

The rules in (1) predict that out of eight possible prosodic phrasings of the four PWs in Figure 3.2, five are possible and three are impossible. Note that only non-recursive prosodic structures are considered here, following the proposal that prosodic structures are non-recursive in Seoul Korean (Jun, 1993, 2000). The phrasings are listed in (3) where numbers indicate the index of the PWs (the first, second, third, and fourth PW) and parentheses indicate AP phrasings.

(3) (Im)possible accentual phrasings of the sentence in (2) that has a nested right branching structure as in Figure 3.2

a. Default prosody: (1)(2)(3)(4)

b. All in one (e.g., Focus on (1)): (1 2 3 4)

c. Right element tolerance:

(i) (1)(2)(3 4)

(ii) (1)(2 3)(4)

(iii) (1)(2 3 4)

Left element constraint:

(i) *(1 2)(3)(4)

(ii) *(1 2 3)(4)

(iii) *(1 2)(3 4)

The rules in (3) predict that the first (/ʧʰamse-nɨn/) and the second PW (/ʧakɨn/) cannot be phrased together in an AP, because the second PW is part of the left element (NP2) of a branching constituent (VP) in Figure 3.2. This is termed 'Left element constraint' in (3).

The only way that the first and the second PW can be phrased in the same AP is if all four PWs are phrased in the same AP. This may be the prosodic structure when the first PW is focused, which makes the following words phrased together in the same AP as the focused PW (Jun, 1993).

On the other hand, all four phrasings are allowed with respect to the second, third and fourth PWs because it is a left branching structure. Put differently, any one of them can be the last in an AP since it is the right element of a branching constituent. This is termed 'Right element tolerance' in (3).

While the rules in (1) make predictions about a set of possible and impossible ways that PWs are phrased into APs, they do not make further predictions of whether a particular phrasing is preferred or dispreferred by speakers and listeners of Seoul Korean. For instance, while the prosodic phrasing (1)(2)(3 4) is allowed by the rules, it mismatches with the syntactic structure in Figure 3.2 in that there is no syntactic constituent that only includes the third and the fourth PWs. Would this mismatch with syntactic constituency cause the prosodic phrasing (1)(2)(3 4) to be dispreferred by speakers? In general, if PWs are in the partially left branching structure, as in the VP in Figure 3.2, what would be the preferred way that they are phrased into APs? In this study, I will investigate a number of such cases involving incomplete nouns.

3.2.3.2 Incomplete nouns and Post-CP PWs

As Korean is a head-final language with the word order of 'SOV', all syntactic elements that follow the first verbal predicate in the linear order are likely to be in the right element tolerance relationship with the first verbal predicate within the sentence. For instance, consider the syntactic construction in Figure 3.3. The first VP is followed by a nominalizer /kɔ/ which originates as a head of the NP. As /kɔ/ is an example of incomplete nouns, which constitute a closed set of 'monosyllabic nouns which are, though separated by spaces [in orthography], not used by themselves but need a modifier' (Jun, 2000).

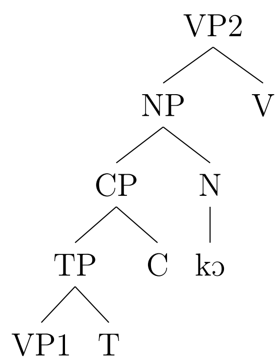


Figure 3.3: An example of a left branching structure.

Incomplete nouns are different from verbal suffixes that are assumed to originate in a head position but later merged with the verb in spell-out (Yoon, 1993; Chung, 2011). While suffixes are written together with the host in the same Eojeol, these incomplete nouns are written as an Eojeol (PW) on their own. Unlike other PWs, however, incomplete nouns tend to have a BI level of 0 and therefore /kɔ/ is expected to be phrased together with the preceding CP. This means that the CP in Figure 3.3 and the incomplete noun /kɔ/ are expected to be prosodically phrased together in the same AP.

In Korean sentences, these incomplete nouns are followed by the matrix predicates, which are typically utterance-final in default word order. I will refer to these matrix predicates as 'Post-CP PWs' in this chapter. Right element tolerance makes a further prediction about PWs after the CP and the incomplete noun. The matrix predicate (V in Figure 3.3) and the NP before it (which is a CP nominalized by the incomplete noun /kɔ/) are in 'the right element tolerance' relationship, which means that the matrix predicate may also be contained within the same AP with the NP (i.e., (...PW /kɔ/ V)$_{AP}$). However, unlike the incomplete nouns, it has not been investigated whether matrix predicates after incomplete nouns are also phrased together with PWs before them in the same AP.

In this study, some examples of incomplete nouns and Post-CP PWs will be examined in-depth, after the PW initial syllables are assigned to prosodic juncture categories in the clustering analysis. If these incomplete nouns and Post-CP PWs are prosodically phrased together with the preceding elements, their initial syllables are expected to be associated with a lower level of BI, which may indicate that they are AP-medial.

3.2.4 Research questions

There are two main research questions.

First, does the informativity of tonal and segmental cues vary by the laryngeal category of PW-initial onset? Given the results of Lee (2023), it seems that Lenis PW-initial syllables may be distinguished better with the segmental cue. On the other hand, Aspirated PW-initial syllables may be distinguished better with the tonal cue. This question is addressed in §3.4.

Second, are the incomplete nouns and Post-CP PWs acoustically realized with a lower level of BI? In the syntactic position, both the incomplete nouns and Post-CP PWs are in the right element tolerance relationship with the preceding PW(s). It is an empirical question of how they are prosodified, though from their syntactic position, it is predicted that the initial syllables of these PWs would be acoustically weak, and therefore associated with a lower level of BI. This question is addressed in §3.5.

3.3 Methods

3.3.1 Spontaneous speech corpus

The data come from syllables with a lenis or aspirated obstruent onset in the Seoul Corpus (Yun et al., 2015). Utterance initial syllables were not included because both Segmental and Tonal cues were defined as relative change from the preceding syllable, and utterance initial syllables by definition did not have a preceding syllable. Numbers of lenis and aspirated tokens per Eojeol position are presented in Table 3.4.

	Lenis	Aspirated
Initial	68,464	7,776
Medial	92,256	19,238
Total	160,720	27,014

Table 3.4: Lenis and Aspirated tokens per Eojeol position in the Seoul Corpus (Yun et al., 2015).

After parameterization, some tokens were lost due to missing value of F0 measurement. This is expected since both lenis and aspirated obstruents may be pronounced with aspiration and the subsequent vowels are heavily influenced by the aspiration sometimes to the extent that no F0 can be measured from the entire syllable interval. Numbers of tokens that were successfully parameterized are presented in Table 3.5. Disproportionately more Aspirated tokens were lost as F0 measurement was prone to failure for those due to aspiration.

3.3.2 Tonal cue ($\textsc{Max}\Delta_{\mathbf{F0}}$)

In order to parameterize the tonal changes between syllables, the following steps were taken. First, for each sound file, F0 was measured for every 10 ms in Praat (Boersma & Weenink, 2023).

	Lenis	Aspirated
Initial	61,981	6,711
Medial	85,760	15,736
Total	147,741	22,447
Original	160,720	27,014
Loss (%)	12,979 (8%)	4,567 (17%)

Table 3.5: Lenis and Aspirated tokens per Eojeol position after paremeterization.

Then, for every syllable interval, the mean was computed over the set of F0 measurements taken within that syllable interval. This mean value is referred to as 'MEAN(σ_i)'. The mean was taken for the entire syllable interval, rather than a vowel interval, because it was observed that the F0 minimum was often located in the coda if the syllable has one.

Once I computed the 'MEAN(σ_i)' for every syllable, it was converted to a fraction of the range of F0 values from the utterance that it was from, using Equation (3.1), following Lee & Lee (2013), where MIN/MAX(Utterance) means the minimum/maximum F0 of the utterance that the syllable is taken from.

Equation (3.1) therefore put the MEAN(σ_i) value within the range of 0 and 1, so that it was possible to compare between speakers of different F0 range. The normalization was done with the range of each utterance rather than with the range of a speaker throughout the dataset so as to avoid the possibility of misanalyzing the cases where one speaker would produce an entire utterance with a higher pitch than the next utterance, in which case all of the syllables in the first utterance will all be analyzed as having a high normalized F0.

$$\textsc{Mean}\text{ Fraction of Range}(\sigma_i) = \frac{\textsc{Mean}(\sigma_i) - \textsc{Min}(\text{Utterance})}{\textsc{Max}(\text{Utterance}) - \textsc{Min}(\text{Utterance})} \tag{3.1}$$

Algorithm 1 Computing MaxΔ_{F0}

$\Delta_{\text{F0}} = \textsc{Mean}(\sigma_i) - \textsc{Mean}(\sigma_{i-1})$
if $\Delta_{\text{F0}} \geq 0$ **then** ▷ This is when the σ_i has a higher F0 than σ_{i-1}, meaning a tonal rise
  $\textsc{Max}(\Delta_{\text{F0}}) = \textsc{Max}(\sigma_i) - \textsc{Min}(\sigma_{i-1})$
else ▷ And this is when the σ_i has a lower F0 than σ_{i-1}, meaning a tonal fall
  $\textsc{Max}(\Delta_{\text{F0}}) = \textsc{Min}(\sigma_i) - \textsc{Max}(\sigma_{i-1})$
end if

Every syllable mean F0 was then compared with the syllable that came before it (σ_{i-1}) by subtracting '$\textsc{Mean}(\sigma_i)$' from '$\textsc{Mean}(\sigma_{i-1})$', which is henceforth termed Δ_{F0}.

Therefore, if σ_i has a higher F0 value then σ_{i-1}, this means that the tone is rising, and the Δ_{F0} value is going to be positive. On the other hand, if σ_i has a lower F0 value than σ_{i-1}, then this Δ_{F0} value is going to be negative, and indicate that the tone is falling. Δ_{F0} was maximized to exaggerate the trend between the two syllables by the procedure described in Algorithm 1. The resulting variable, $\textsc{Max}\Delta_{\text{F0}}$, is how the tonal marking was parameterized in Lee (2023), and in this study.

3.3.3 Segmental cue (CVE)

In Lee (2023), the segmental cue of lenis obstruents was parameterized using three measurements taken over the Lenis stop closure interval that were reported to be correlated with the lenition of stops: percentage of voiced interval Davidson (2016), the difference between the maximum and minimum rate of change in intensity as in Kingston (2008), and the closure duration, which was speech-rate normalized with the method described in Yoo (2020). Following Dalcher (2007), Principal Components Analysis was used to combine them into a single variable for comparison with the tonal cue. The first component, which accounted for 77% of the variance in the data. A Lenis token that was strong (voiceless, long, with abrupt change in intensity) had a positive value of this component, whereas a Lenis token that was realized as weak had a negative value.

In contrast, this study will use simpler measure of CVE used in Chapter 2. As discussed in Chapter 2, CVE is the maximum falling velocity of intensity change from the preceding syllable. By definition, CVE is typically negative but I multiplied it by -1 to make it positive, so that a higher value of CVE would indicate a stronger articulation of the lenis obstruent onset.

3.4 Prosodic juncture categories: Lenis and Aspirated syllables

This section presents how Lenis and Aspirated PW-initial syllables are distributed in the acoustic space defined over CVE and $\textsc{Max}\Delta_{\mathrm{F0}}$.

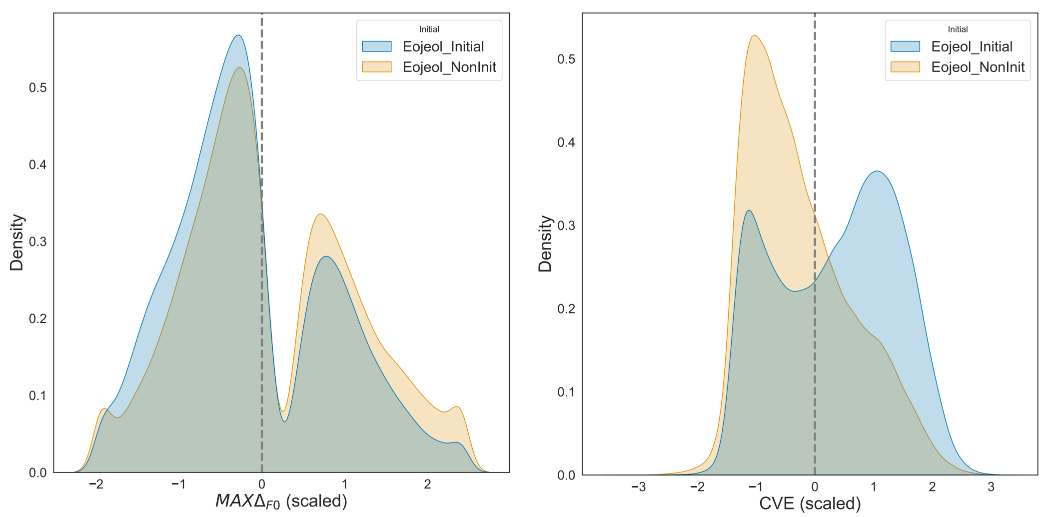
First, distributions of PW-initial and PW non-initial syllables are plotted on each dimension. Note that both CVE and $\textsc{Max}\Delta_{\mathrm{F0}}$ are meant to distinguish whether a given syllable token has a BI level higher or equal to 2, e.g., is it AP-initial or above? Assuming that a higher prosodic constituent properly contains a lower prosodic constituent ('Proper bracketing' (Selkirk et al., 1982)), all PW non-initial syllables are AP medial, but some PW-initial syllables may be AP-initial. Therefore, if a cue is informative, then the distribution of PW-initial syllables should have some areas where it does not overlap with the distribution of PW non-initial syllables. The tokens in the overlapped area are the ones that have BI lower than 2: AP-medial, and the tokens of the PW-initial syllable that are not in the overlapped area may be candidates for AP-initial syllables with BI level higher or equal to 2.

Next, PW-initial syllables are plotted on the 2-dimensional space of CVE and $\textsc{Max}\Delta_{\mathrm{F0}}$. A K-means analysis is performed to determine the best number of clusters that accounts for the variation in the data.

3.4.1 Lenis PW on Tonal and Segmental cues

To perform a K-means clustering analysis in 2-dimensional space, segmental and tonal cues are both z-score transformed using 'StandardScaler()' in the 'scikit-learn' package (Pedregosa et al., 2011). Here, I first present the distribution of lenis syllables over each dimension.

Cues are standardized within each laryngeal category: Lenis and Aspirated. Rather than comparing tokens in terms of raw values, standardizing these cues relativizes the tokens with respect to the mean. For instance, when comparing an Aspirated token with another Aspirated token, both are likely to have higher CVE values in an absolute sense, compared to CVE values of Lenis tokens. Standardizing within Aspirated tokens would make a CVE value positive if it is higher than the average values that Aspirated tokens have, and negative if it is lower than the average that Aspirated tokens have. The plots below will use the scaled dimensions, rather than the original ones. Results pattern the same with or without scaling.



(a) Lenis syllables on $\text{Max}\Delta_{\text{F0}}$ (b) Lenis syllables on CVE

Figure 3.4: Distribution of Lenis PW-initial and PW non-initial syllables

Figure 3.4a plots the density of Eojeol-Initial (PW-initial in blue) and Eojeol Non-initial (PW non-initial in yellow) syllables over $\text{Max}\Delta_{\text{F0}}$. There are two modes in the distribution of PW-initial and PW non-initial syllables: one on the left side of 0 and one on the right side of 0. The tokens that have values on the left side of 0 are the ones that have lower F0 values than the preceding syllable (i.e. tonal fall), and the tokens that have values on the right side of 0 are the ones that have higher F0 values than the preceding syllable (i.e. tonal rise).

However, the distributions of PW-initial and non-initial syllables almost completely overlap, replicating results of Lee (2023). If the size of the fall is indeed a reliable cue to distinguish an AP boundary juncture from an AP-medial fall, the left mode of the PW-initial syllable distribution (blue) should be located to the left of the left mode of the PW-medial syllable distribution (yellow). However, this does not seem to be the case.

Next, Figure 3.4b plots the density of Eojeol-Initial (PW-initial in blue) and Eojeol Non-initial (PW non-initial in yellow) syllables on CVE. Again, there are two modes in the distribution of PW-initial syllables: one on the left side of 0 and one on the right side of 0. The tokens that have values on the left side of 0 are the ones that are acoustically weak, and the tokens that have values on the right side of 0 are the ones that are acoustically strong. In contrast, there seems to be only one mode in the distribution of PW-medial syllables: one on the left side of 0 (acoustically weak). At least some of the PW-initial syllables on the right side of 0 are separated from the distribution of PW non-initial syllables, suggesting that CVE is a more informative cue than $\textsc{Max}\Delta_{\text{F0}}$.

3.4.1.1 K-means analysis of Lenis PW-initial syllables

The K-means analysis on the 2 dimensional space is performed by setting the value of K from 2 to an arbitrary number, where K is the number of clusters that the algorithm is supposed to find. Then for each K value, a silhouette analysis is performed, which is a commonly used metric to evaluate clustering analysis, using ‘silhouette_score()’ from the ‘scikit-learn’ package (Pedregosa et al., 2011). The output of the silhouette analysis ranges from 1 to -1; where 1 indicates perfect separation, 0 indicates a complete overlap, and -1 indicates complete misrepresentation of the data. The K value that gets the highest silhouette score is therefore the number of clusters that would best represent the data.

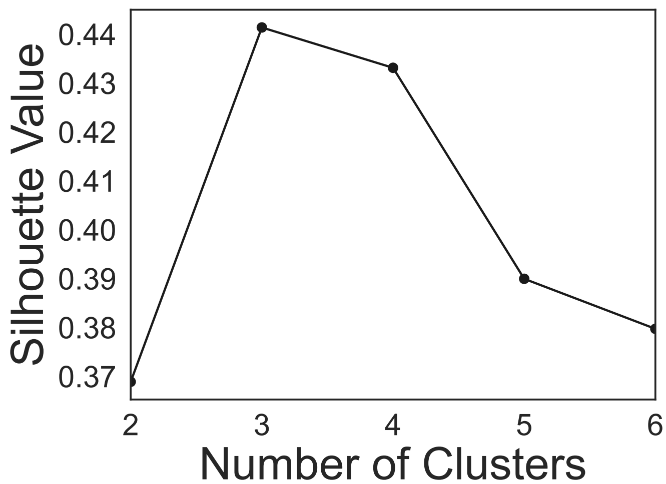


Figure 3.5: Silhouette scores for K-means analysis on Lenis PW-initial syllables, varying K from 2 to 6.

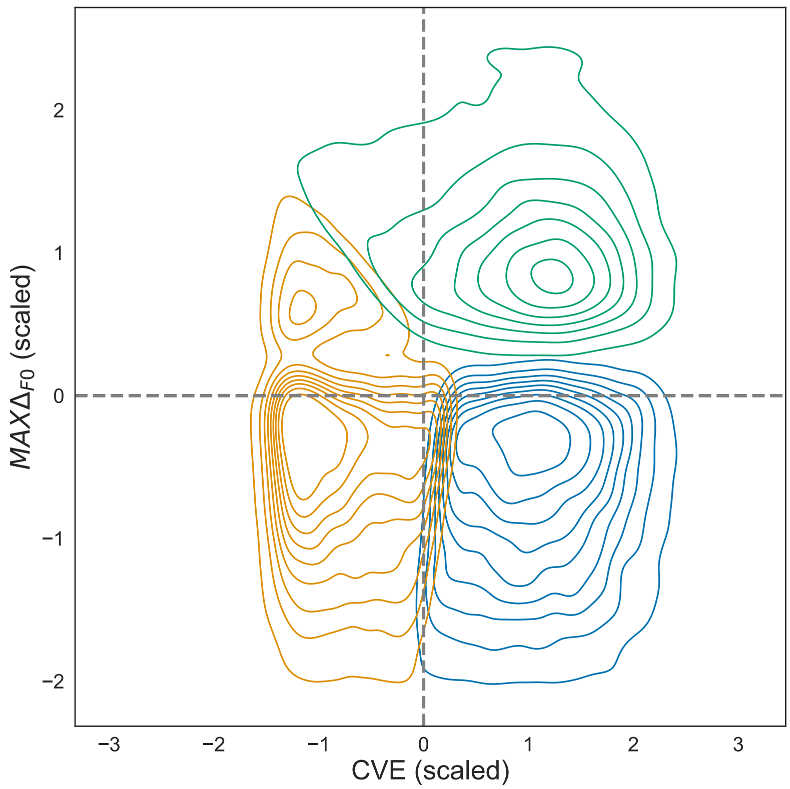
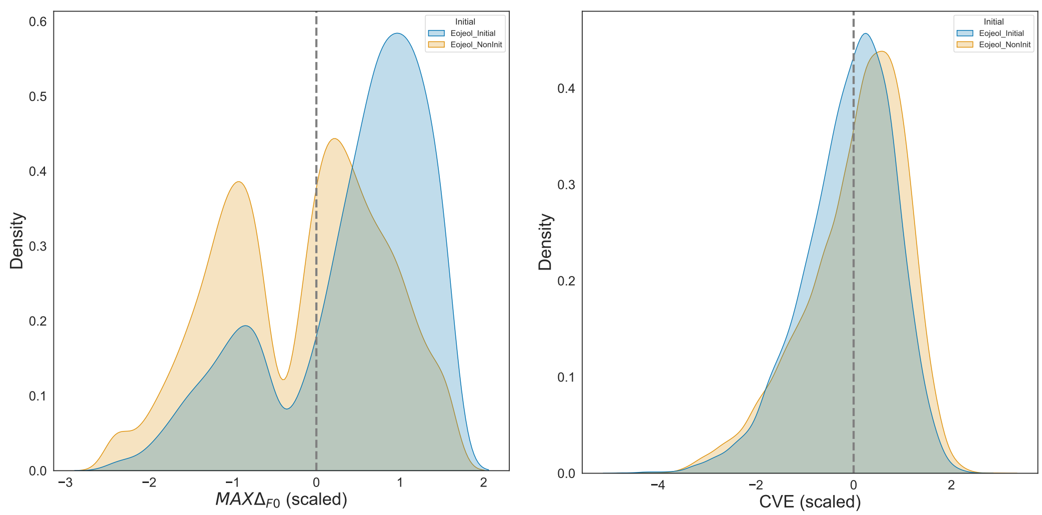


Figure 3.6: Result of K-means analysis on Lenis PW-initial syllables.

Figure 3.5 shows the result of the silhouette analysis for K values ranging from 2 to 6. The silhouette analysis indicated that the number of clusters that would best account was 3 for the PW-initial syllables, which yielded a slightly higher silhouette value than 4 clusters. These found clusters for the PW-initial syllables are plotted in Figure 3.6. The resulting three clusters mirror the clusters in Figure 3.1, thus replicating the same results in Lee (2023) with additional data. The descriptions of each cluster are the same as ones given in §3.2.2.1. The only difference would be that the color of clusters are different comparing Figure 3.6 with Figure 3.1.

3.4.2 Aspirated PW and Tonal and Segmental cues

Next, we move onto Aspirated PW-initial syllables. As for Lenis syllables, Figure 3.7a plots the density of Eojeol-Initial (PW-initial in blue) and Eojeol Non-initial (PW non-initial in yellow) syllables over $\text{MAX}\Delta_{\text{F0}}$. There are two modes in the distribution of PW-initial and PW non-initial syllables corresponding to a tonal fall (left) and a tonal rise (right). Unlike distributions of Lenis syllables, the distribution of PW-initial syllables does not completely overlap with that of PW-medial syllables. While PW-initial syllables mostly have a 'tonal rise', PW-medial syllables seem to equally have a tonal fall or a tonal rise. Moreover, the rise of PW-initial syllable seems to be larger (the mode in the blue distribution is more to the right) compared to the rise of PW-medial syllable (the mode in the yellow distribution). This suggests that local change in F0 is an informative cue for marking AP boundaries, when PW starts with an Aspirated obstruent.



(a) Aspirated syllables on MaxΔ (b) Aspirated syllables on CVE

Figure 3.7: Distribution of Aspirated PW-initial and PW non-initial syllables

On the other hand, the distributions of PW-initial and PW non-initial syllables over CVE are completely overlapping, as shown in Figure 3.7b. This means that PW-initial syllables are not guaranteed to be weaker than PW-medial syllables on average.

3.4.2.1 K-means analysis on Aspirated PW-initial syllables

The silhouette analysis in Figure 3.8 indicated that the number of clusters that would best account was 2 for the PW-initial syllables. The found clusters for PW-

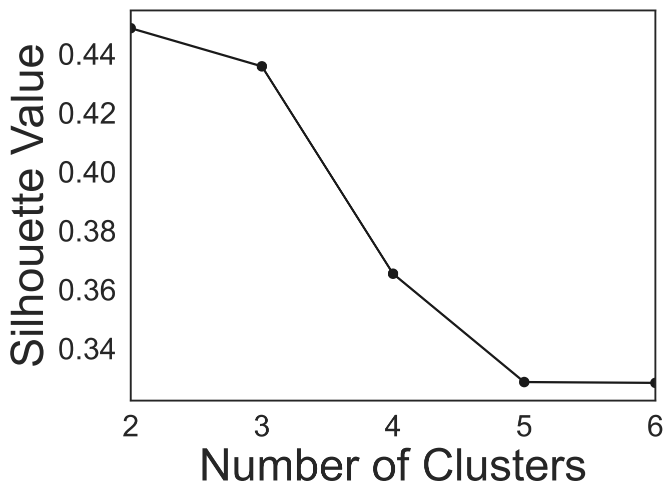


Figure 3.8: Silhouette scores for K-means analysis on Aspirated PW-initial syllables, varying K from 2 to 6.

initial syllables are plotted in Figure 3.9. Since the segmental cue does not participate in clustering subcategories of the distribution of PW-initial syllables, the resulting clusters in Figure 3.9 look the same as the two modes in one dimensional space of $\text{Max}\Delta_{\text{F0}}$. Judging from the density of two clusters, it seems that most Aspirated PW-initial syllables were realized as AP-initial (the yellow cluster), and a smaller subset of them had a different acoustic realization, which would be associated with a BI level below 2 (the blue cluster).

While the expected number of BI levels is three, the results of the K-means analysis for Aspirated PW initial syllables indicated that there were only two clusters (tonal rise or fall). This indicates that there may not be a reliable difference, between an AP-initial Aspirated syllable and an Aspirated syllable at the beginning of a prosodic constituent larger than AP, when it comes to the local change in F0 from the preceding syllable.

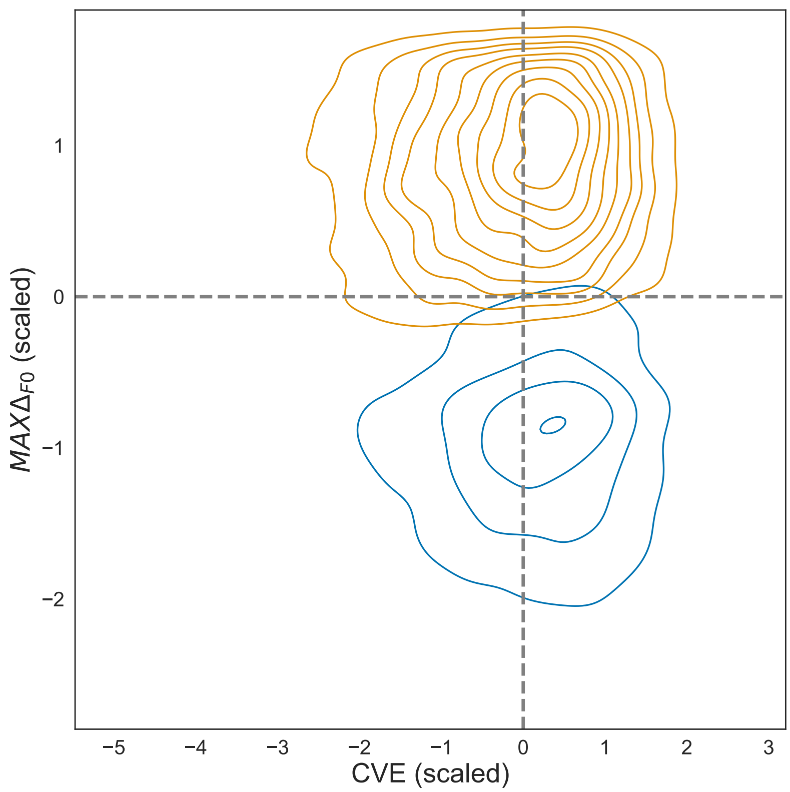


Figure 3.9: Result of K-means analysis on Aspirated PW-initial syllables.

3.4.3 Summary of results

The results presented in this section address the first research question: does the informativity of tonal and segmental cues vary by the laryngeal category of PW-initial onset?

K-means analysis for Lenis PW-initial syllables showed that segmental realization seemed to distinguish syllables that have a BI level higher or equal to 2, from those that have a BI level lower than 2 (AP-medial syllables). For those that have a BI level higher than 1, it seemed that depending on the tonal cue, there were syllables that have a BI level higher than 2 (ip or IP-initial syllables) and ones that have a BI level of 2 (AP-initial syllables). These results replicate the findings in Lee (2023), with additional data from the corpus.

On the other hand, K-means analysis for Aspirated PW-initial syllables showed that PW-initial syllables were mostly AP-initial and these were distinguished on the basis of the tonal cue. A smaller subset of Aspirated PW-initial syllables had a F0 value lower than the preceding syllable, which was argued to be associated with a BI level below 2.

The results of K-means analysis indicate that the prosodic category junctures may be phonetically encoded differently, depending on the laryngeal category of the syllable onset. When the tonal and segmental cues are parameterized as $\text{M}{\scriptsize\text{AX}}\Delta_{\text{F0}}$ and CVE, while $\text{M}{\scriptsize\text{AX}}\Delta_{\text{F0}}$ is informative in distinguishing AP-initial Aspirated PW-initial syllables, CVE is informative in distinguishing AP-initial Lenis PW-initial syllables.

3.5 AP-medial-like PWs

In the previous section, I demonstrated that in both cluster analyses, one cluster was identified as being 'AP-medial-like': the Lenited category for Lenis PW-initial syllables and the Fall category for Aspirated PW-initial syllables. These PW-initial syllables are exceptions to the trend that every PW is dominated by a single AP (Schafer & Jun, 2002; Kim, 2004).

In what follows, these AP-medial-like categories are investigated to address the second research question. The second research question concerns acoustic realizations of incomplete nouns that mark the end of CPs and PWs that follow the incomplete nouns (i.e., Post-CP PWs). As explained in §3.2.3, incomplete nouns and matrix predicates may not start a new AP though typically each PW starts a new AP (Jun, 1993; Kim, 2004). If incomplete nouns and matrix predicates after CPs do not start a new AP, they are likely to be found in the 'AP-medial-like' clusters identified by the K-means analysis.

To preview, results presented here suggest that there may be syntactic generalizations for PWs that do not seem to start new APs. It was found that incomplete nouns were prosodically phrased together with preceding PWs as reported (Jun, 2000). In addition, it was found that Post-CP PWs that follow incomplete nouns did not start new APs either. Both incomplete nouns and Post-CP PWs are in the right element tolerance relationship, which could explain why they are allowed to be prosodically phrased together in the same AP. The results however further suggest that they are not simply 'allowed', but may even be preferred to be prosodically phrased together in the same AP, given that they are more frequently phrased together with the previous PW, than they are not. The results here therefore suggest that the syntactic restriction makes a stronger prediction than previously assumed.

3.5.1 AP-medial-like Lenis PW-initial syllables

In §3.2.3, it was predicted that monosyllabic incomplete nouns would have a BI level of 0, and phrased in the same AP with the preceding PW. Out of 61,981 Lenis PW-initial syllables, 23,702 (38.2%) tokens had a value of scaled CVE below 0, i.e., more lenited than average Lenis PW-initial syllables. These would belong to the yellow cluster in Figure 3.6. This cluster can be further divided by the tonal cue to 'Weak Rise' (i.e., segmentally realized as weak and tonally a rise from the preceding syllable) and 'Weak Fall' (i.e., segmentally realized as weak and tonally a fall from the preceding syllable). The number of tokens per each subcategory and 10 most frequent PWs with their counts are presented in Table 3.6.

Two things are very noticeable from the lists of PWs in Table 3.6. First, except for 2 PWs, all other PWs start with /k/. This may be because /k/ is the most frequent consonant in Korean. In Seoul Corpus (Yun et al., 2015), out of 231,626 Eojeol tokens, 49,209 (21.2%) start with /k/. In terms of PW types, out of 37,750 unique PWs, 4,909 (13%) start with /k/ Cohen Priva (2017) mentions that a high lenition rate of /k/ may be due to its high frequency in the language (p. 581).

Weak Rise			Weak Fall		
PW	Meaning	Count	PW	Meaning	Count
kɔ	**Nominal**	621	kɔ	**Nominal**	1,851
ke	**Nominal**-NOM	287	ke	**Nominal**-NOM	738
kɨresɔ	so	173	katʃiko	with	452
katʃiko	with	113	kat^{h}ɨn	to seem-REL	451
kɨnte	but	109	kɔnɨn	**Nominal**-TOP	249
kɔn	**Nominal**-TOP	99	tʃom	a little	247
kat^{h}ɨn	to seem-REL	96	kɔn	**Nominal**-TOP	245
kɔs-to	**Nominal**-even	92	ponik*a	try	243
kɔ-nɨn	**Nominal**-TOP	81	kɔs-to	**Nominal**-even	234
kɨrɔn	that	75	kɔl	**Nominal**-ACC	217

Table 3.6: Top 10 most frequent Lenis PWs that belonged to the Weak cluster. PWs in the Weak Rise column had higher F0 compared to the preceding syllable, PWs in the Weak Fall column ad lower F0 compared to the preceding syllable.

More interestingly, the nominalizer /kɔ/ with various suffixes appear in both Weak Rise and Weak Fall clusters, as predicted in §3.2.3. Tokens of this monosyllable incomplete noun /kɔ/ are examined in detail next.

3.5.1.1 Nominalizer /kɔ/

/kɔ/ appears after a predicate in the sentence to nominalize the entire CP. An example sentence from the Seoul Corpus (Yun et al., 2015) is in (4). The entire clause [saŋhwaŋ-i tʃal matʃ-tʃi anh] ('the situation did not fit well') was relativized with /nɨn/ and nominalized with /kɔ/. This was the subject of the matrix predicate /kat^{h}/ ('be like'). The second part of the sentence ('(I) ended up coming back') is connected to the first part with the sentence final particle /sɔ/ 'because'. Drawing the entire syntactic structure of this sentence would be out of scope of this study, and not relevant for the current discussion. The important point is that /kɔ/ as a nominalizer forms a constituent with the preceding clause (a nominalized CP), and forms a left branching structure with the preceding PWs in the syntactic structure. Therefore, the onset segment /k/ of /kɔ/ is expected to have a weak realization.

(4) saŋhwaŋ-i ʧal maʧ-ʧi anh-nɨn kɔ katʰa-sɔ, toɾa

Situation-NOM well fit-COMP NEG-REL NOMINAL seem-because turn

o-ke twe-s*ɔs*-ɔ-jo

come-COMP become-PAST-DECL-HON

'Because it did not seem like the situation fit well,

(I) ended up coming back.

File name: s08f16m6, time slice: 164.30452 - 167.53552 in Seoul Corpus

Out of 2,620 tokens of /kɔ/ in the data set, 2,472 (621 + 1851) tokens (94%) had a weak realization of /k/ (CVE (scaled) < 0). To check the context that appears before /kɔ/, for each /kɔ/, the PW that occurred immediately before it was retrieved from the corpus. Table 3.7 presents the top 5 PWs which preceded /kɔ/ with their counts. All of these preceding PWs end with some variation of /ɨn/, which is a relativizer.

PW	Meaning	Count
kɨɾɔ-n	'to be like that-REL'	336
iɾɔ-n	'to be like this-REL'	143
is*-nɨn	'to exist-REL'	134
katʰ-ɨn	'to be like-REL'	97
ʧo-ɨn	'to be good-REL'	95

Table 3.7: Top 5 PWs that precede /kɔ/ in Seoul Corpus (Yun et al., 2015)

It was further checked if /kɔ/ is lenited because of its syntactic position with respect to the preceding element (i.e., right branching tolerence), rather than because of its preceding phonological context (/n/). /k/ initial PW-initial syllables were filtered to include only the ones that had a preceding PW that ended with /n/. Figure 3.10 plots CVE of other PWs that had a preceding /n/-ending PW (11,382 PWs) next to CVE of PWs that start with /kɔ/ which followed an /n/ (6,059). These PWs that start with /kɔ/ include /kɔ/, along with the ones with various suffixes, as shown in Table 3.6. Figure 3.10 shows that /kɔ/ (orange) had lower CVE values compared to other /k/-initial PWs (blue), when /n/ preceded.

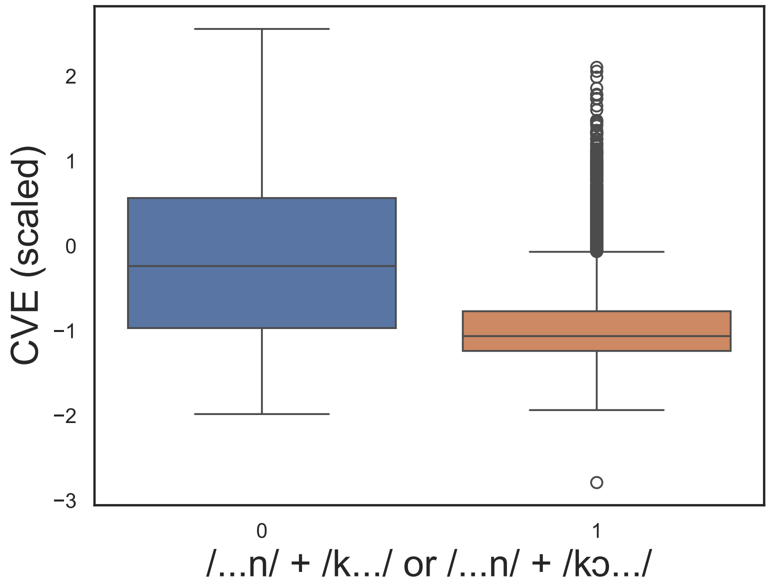


Figure 3.10: CVE comparison of /k/-initial PWs and /kɔ/ tokens that follow a preceding PW ending with an /n/.

To test if the difference between the two means were significant, Welch's T-test was performed to accomodate the fact that these two have disproportionate sample size (4,473 and 6,059). The result indicated that two means were significantly different ($t(6383.97) = 52.11$, $p < 0.0001$). The positive sign of t-statistics value suggest that the other /k/-initial PWs, which was the reference level, had a higher CVE value, and therefore more acoustically more strong when /n/ preceded. The result suggests that the incomplete noun /kɔ/ seems to be predominantly phrased together at least with the immediately preceding PW into a single AP, thus having a BI level of 0, as predicted from its position in the syntactic structure, with respect to the preceding elements.

3.5.1.2 /kɔ/ + /kat^h.../ ('it's like...')

In §3.2.3, it was also hypothesized that after /kɔ/, the matrix predicate may not start a new AP either, as it is also in the right element tolerance relationship with the preceding nominalized CP, in the syntactic structure. To test the prediction, the five most frequent (token-wise) PWs after /kɔ/ were identified. Table 3.8 presents the result. It turns out /kɔ/ is frequently followed by the verb /kat^h/ which varies in different verbal suffixes.

PW	Meaning	Count
kat^h-ajo	'seem-DECL-HON'	255
kat^h-ejo	'seem-DECL-HON'	140
kat^h-ko	'seem-COMP'	78
kat^h-ɨnte	'seem-but'	58
kat^h-asɔ	'seem-because'	36

Table 3.8: Top 5 Lenis PWs following /kɔ/

To test if /kat^h/ and /kɔ/ were in the same AP, Lenis PW-initial syllables were filtered to include only the ones that followed /kɔ/ in the sentence. Figure 3.11 plots CVE of other /k/-initial PWs that followed /kɔ/ and CVE of PWs that start with /kat^h/ which followed /kɔ/. These PWs that start with /kat^h/ include all PWs that vary in verbal suffixes. Figure 3.11 shows that in the sequence /kɔ kat^h/(orange), the initial consonant in the second PW /k/ had lower CVE values compared to other lenis segments following /kɔ/ (blue). This difference was significant in Welch's t-test ($t(97.41) = 9.48$, $p < 0.0001$).

Out of 806 instances where /kɔ/ was followed by /kat^h.../, 605 had the initial /k/ in /kat^h.../ that had a CVE (scaled) value below 0. This again means that /k/ in /kat^h.../ are typically realized as weak. In addition, there were also additional 1207 tokens where the initial /k/ that was not transcribed in the pronunciation tier (e.g., /kɔ at^h.../) as they were barely audible. To summarize, out of 2013 instances of /kɔ/ + /kat^h.../, 1862 (92%) of them were either lenited or deleted. These results suggest that the BI between /kɔ/ and /kat^h.../ should not be higher than 2, and these two PWs belong to the same prosodic constituent.

These results strongly suggest that speakers prefer to prosodically phrase the incomplete noun /kɔ/ and the matrix predicate /kat^h.../ together in the same AP.

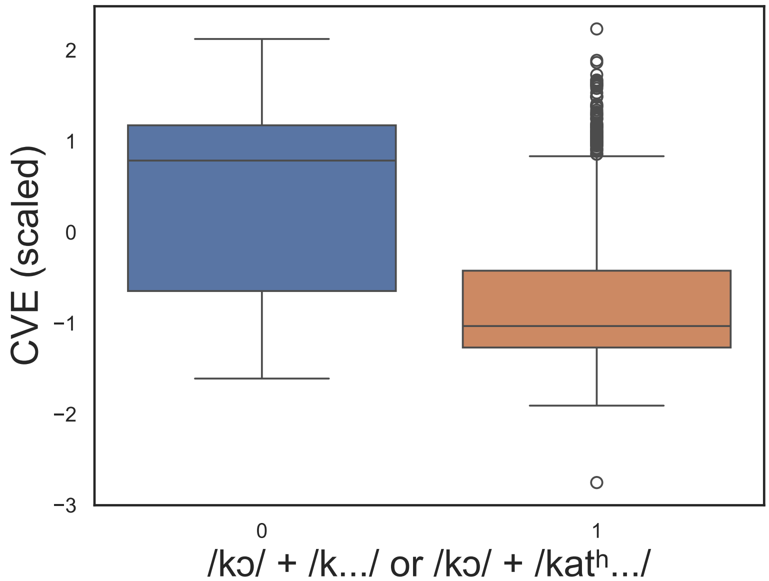


Figure 3.11: CVE comparison of other /k/-initial PWs and /kat^h.../ tokens when the preceding PW was /kɔ/.

Though the syntactic structure of these PWs may not necessitate that they are prosodically phrased together, at least in a spontaneous speech, the syntactic structure seems to have a strong influence on how these PWs are prosodically phrased.

3.5.2 AP-medial-like Aspirated PW-initial syllables

The results of K-means analysis for Aspirated PW-initial syllables indicated that most Aspirated PW-initial syllables had a higher F0 than the preceding syllable, while a small subset had a lower F0 than the preceding syllable. The latter cluster (blue in Figure 3.9) will be referred to as AP-medial-like. Out of 6,711 Aspirated PW-initial syllables, 1,672 tokens (25%) tokens had a value of scaled $\text{M}\textsc{ax}\Delta_{\text{F0}}$ value below 0.

3.5.2.1 /-l/ + /t^hente/ (Conditional ending)

One of the PWs in this AP-medial-like cluster is /t^hente/. This PW is chosen to be investigated further as it is a sister to a CP in the syntactic structure, just as the Lenis PW /kɔ/ is. Following Kwon (2013), I gloss it as 'Conditional ending'

(COND.END). An example sentence from the Seoul Corpus (Yun et al., 2015) is in (5). The first part of the sentence ends with /tʰente/ 'Conditional ending', which takes the clause ending with /l/ ('Modal'): 'Even the box price is some hundred Won'. /tʰente/ adds the conditional meaning, translated as 'would be' in (5), to this clause (Kwon, 2013). As /tʰente/ follows a CP, it is potentially prosodically phrased with the preceding PW in the same AP, as it is in the right element tolerance relationship with the preceding CP.

(5) paksɨ kakjɔk-to mjɔʧʰ pek-wɔn ha-l tʰente, iltan-ɨn
Box price-even some hundred-Won be-MOD COND.END, now-TOP
inkɔnpi-ni mwɔ-ni ha-mjɔn-ɨn ʧɔkʧa-ʧjo
cost of labor-or something-or be-COND-TOP deficit-DECL.HON
'Even the box price would be some hundred Won (Korean currency),
now if you count cost of labor or something, (the business) is in the red.
File name: s25m32f6, time slice: 189.56537 - 193.33685 in Seoul Corpus

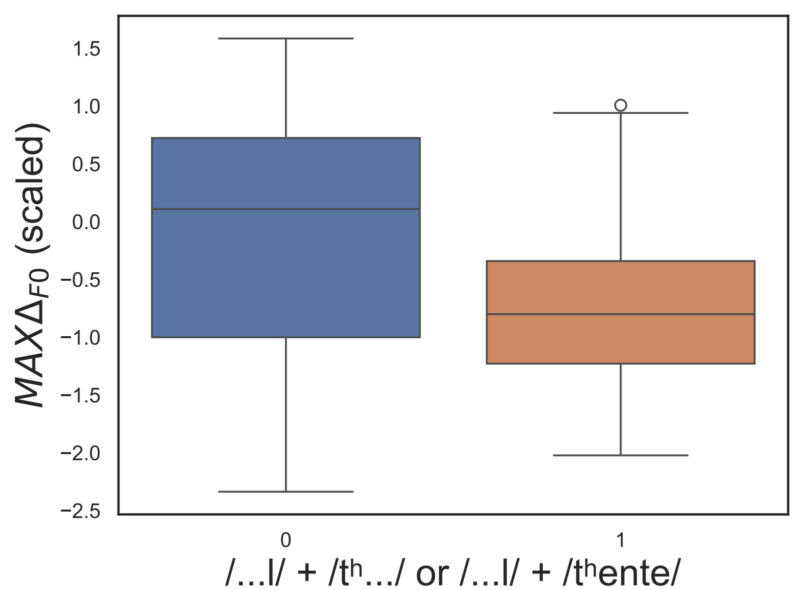


Figure 3.12: $\text{M}\textsc{ax}\Delta_{\text{F0}}$ comparison of tʰ PWs and /tʰente/ tokens that follow a preceding PW ending with an /l/.

To test if /t^hente/ was prosodically phrased together with the preceding PW ending with /l/ ('Modal') in the same AP, /t^h/-initial PWs were filtered to include only the ones that had a preceding PW ending with an /l/. Figure 3.12 plots $\textsc{Max}\Delta_{F0}$ of /t^h/-initial PWs that followed /l/ and $\textsc{Max}\Delta_{F0}$ of /t^hente/. It is clear that in the sequence /l t^hente/ (orange), the initial syllable /t^hen/ had a lower F0 compared to the preceding syllable, and this tonal fall pattern is clearly different from other PWs starting with a /t^h/ following /l/ (blue). This difference was significant in Welch's t-test ($t(92.06) = 4.01$, $p < 0.001$). The tonal profile of /t^hente/ suggests that this Post-CP PW is likely to belong to the same prosodic constituent as the preceding PW, as it is expected from its syntactic position.

3.5.3 Summary of results

The results in this section address the second research question: are the incomplete nouns and Post-CP PWs acoustically realized with a lower level of BI? Lenis and aspirated stops at the beginning of these PWs are predicted to be acoustically realized as AP-medial since they are allowed to be prosodically phrased together with the preceding PWs, according to the Accentual Phrasing rules (Jun, 1993). According to the rule in (1) if a PW is not the left element of a branching syntactic constituent, it can be prosodically phrased together in the same AP with the preceding PW. Incomplete nouns and PWs that follow CPs are not the left element of a branching syntactic constituent and they are therefore allowed to be prosodically phrased together with the preceding PW, according to the rule (c-i) in (1). However, it has not been investigated explicitly whether they are indeed consistently prosodically phrased together with the preceding PW from a spontaneous speech corpus.

Two PWs with a Lenis PW-initial syllable (/kɔ/ and /kat^h/) and one PW with an Aspirated PW-initial syllable (/t^hente/) were investigated. These were identified to be an incomplete noun (/kɔ/), or PWs that can be sister of a CP (/kat^h/ and

/t^hente/). The results showed these PWs had significantly weaker or AP-medial-like acoustic realization, compared to other PWs in the same phonological context.

These results indicate that their acoustically weak and prosodic phrase medial realization supports the syntax-prosody mapping generalization in (1). While the generalization states that these PWs that are not the left element of a branching syntactic constituent are 'allowed' to be prosodically phrased together, the results suggest that they may in fact be preferred to be prosodically phrased together with the preceding PW. This challenges the expectation that every PW starts a new AP (Schafer & Jun, 2002; Kim, 2004), and therefore suggests that the syntactic structure may have a stronger influence on prosodic phrasing than previously assumed. However, it is also important to note that these Post-CP PWs are crucially not required to be prosodically phrased together with the preceding PWs. For instance, out of 806 instances of /kɔ/ + /kat^h.../, 75% of them (605) were prosodically phrased together, but 25% of them were not.

In general, these results call for further investigations on how syntactic structure variably affects prosodic phrasing. I leave this for future work.

3.6 Discussion

Before concluding the chapter, two remaining issues about Lenis PW-initial syllables are addressed.

First, the results in the previous section suggested that the PWs after CP exhibited phonetic realization that would correspond to BI level below 2 unlike other PW-initial syllables. However, it needs to be further checked whether there is a reliable accompanying tonal evidence that suggests that there is no AP-boundary. I present an example where there seems to be a tonal evidence that there is an AP boundary before the Post-CP PW, and yet the Post-CP PW seems to be prosodically phrased together with the preceding PW, based on the segmental realization.

Second, a smaller subset of Lenis PW-initial syllables had an F0 higher than the preceding syllable. These syllables were investigated in further detail.

3.6.1 Prosodic structure of Post-CP PW

In §3.2.3.1, I have shown that Post-CP PWs (e.g., /kat^{h}/) were diagnosed to be prosodically phrased together with the preceding PW, considering that they have a weak realization of the initial lenis consonant. In the case of /kat^{h}/, in particular, its preceding PW tends to be /kɔ/, which happens to be prosodically phrased with the preceding PW as well. Consequently, the resulting AP is predicted to be quite long: [[[CP] kɔ] kat^{h}...].

I show an example where the Post-CP PW is prosodically phrased together with the preceding PW, but there is a tonal evidence that suggests that there is an AP boundary before the Post-CP PW. (6) is an example of such a construction identified from the corpus (Yun et al., 2015) The waveform and spectrogram of the production of this sentence is presented in Figure 3.13.

(6) irɔn k^{h}eisi-nin kweŋʧaŋhi ʧoh-in kɔ kat^{h}-e-jo

this case-TOP extremely be good-REL NOMINAL seem-DECL-HON
'It seems that this kind of case is extremely good.'

File name: s36m43m6, time slice: 350.51186 - 352.67951 in Seoul Corpus

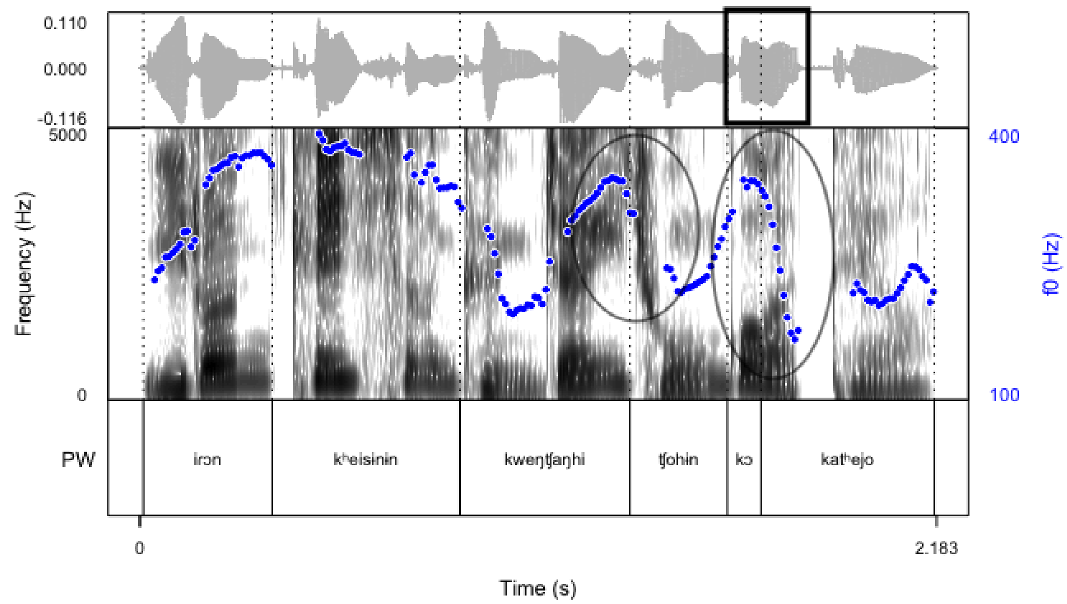


Figure 3.13: Waveform, spectrogram and TextGrid for the sentence in (6).

I will focus on the phrasing of the last three PWs ʧoh-ɨn kɔ kat^{h}-e-jo. From the waveform and spectrogram, the first two PWs seem to be phrased together, which may be annotated as having 'L ... Ha' tones for the three syllables [ʧo.ɨn.kɔ], where the second syllable /ɨn/ is not specified for tone. Note that the underlying /h/ is not pronounced in the pronunciation due to AP-medial /h/ deletion. The Ha tone on /kɔ/ seems to be as high as the Ha tone at the end of kweŋʧaŋhi ('extremely') (circled on Figure 3.13), and therefore do not seem to be the AP-medial +H tone. This H tone cannot be interpreted as a phrase-medial +H tone, since if it was, it would be expected to align with the second syllable /ɨn/, not the third syllable /kɔ/. Therefore, the fall from H on /kɔ/ to the L target at the beginning of /kat^{h}ejo/ seem like it should be labeled as an AP-boundary juncture. However, as the waveform and the spectrogram show (boxed on Figure 3.13), there is no sign that /kat^{h}ejo/ and at least the preceding PW /kɔ/ belong to different 'chunks', considering that the initial segment of /kat^{h}ejo/ is so lenited that there is hardly any consonantal friction.

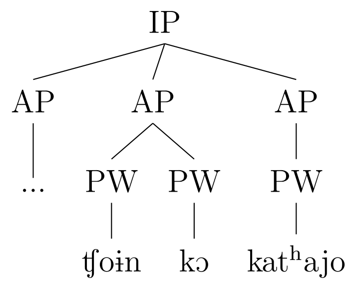
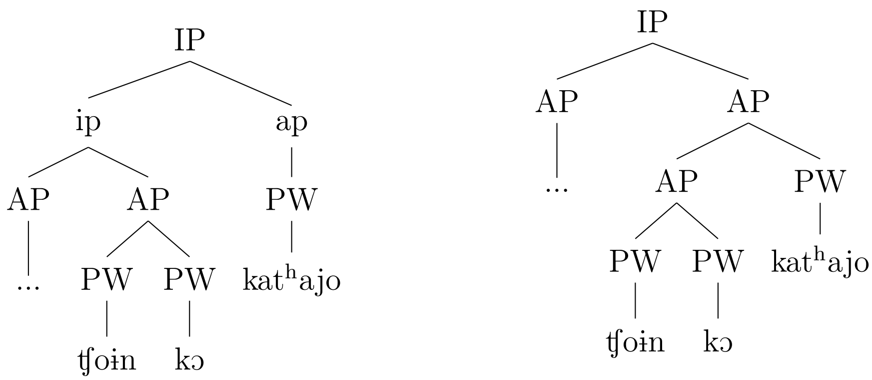


Figure 3.14: Prosodic structures for the utterance in Figure 3.13 that obey the Strict Layer Hypothesis



(a) *Exhaustivity (Jun, 2011)

(b) *Non-recursivity (Baek & Yun, 2018; Lee, 2022)

Figure 3.15: Alternative prosodic structures for the utterance in Figure 3.13 that violate the Strict Layer Hypothesis.

In a prosodic structure that fully obeys the Strict Layer Hypothesis (Selkirk, 1986), it is always true that immediately following an end of an AP, there is a beginning of the next AP. A schematic representation of the prosodic structure of (6) is in Figure 3.14. In this structure, the Post-CP PW /katʰejo/ is phrased as an AP, as it follows an AP-final Ha tone, though it does not have an acoustic realization of an AP-initial syllable. This syllable therefore may be labeled as with a BI level of '2-' ('2-like break because of the tones, but a weaker prosodic break' (Jun, 2000)). If this utterance (Figure 3.13) is analyzed to have the structure in Figure 3.14, it becomes a clear exception to the generalization that AP-initial lenis obstruent is not lenited.

One way to structurally represent the pattern in Figure 3.13, without allowing for an exception that an AP-initial lenis obstruent may be lenited, is to assume that the

Strict Layer Hypothesis (Selkirk, 1986) is not always obeyed. I consider two proposals for the intonational model of Seoul Korean that allow violation of the Strict Layer Hypothesis: a deformed AP proposal (Jun, 2011) which violates Exhaustivity, and a recursive AP proposal (Baek & Yun, 2018; Lee, 2022) which violates Non-recursivity.

The deformed ap structure was first proposed in Jun (2011) to account for pitch range reduction both before and after the focused AP. Jun (2011) showed that in addition to the post-focal dephrasing which was previously reported in Seoul Korean (e.g., Jun, 1993; Schafer & Jun, 2002), there seemed to be also post-focal pitch range reduction, when an IP-medial PW was under focus. Jun formalized this with the deformed AP (ap), and assumed that this is directly dominated by the IP, skipping ip, since it does not show the typical ip-final phrase tone ('H-'). In this proposal, the post-focal PW is dominated by the ip, as a sister to the prominent AP, while the pre-focal PW is directly dominated by the highest IP, violating the Exhaustivity constraint (Selkirk, 1996).

While the deformed ap structure was originally proposed to account for dephrasing due to focus prosody, the same idea may be adopted to account for the pattern in Figure 3.13. This is represented in Figure 3.15a. In this structure, the initial syllable of the Post-CP PW is ap-initial, and being initial of a deformed AP justifies its weaker realization. It also follows an ip-final syllable /kɔ/ which has a phrase-final H- tone. The large tonal fall in Figure 3.13 may therefore be justified (ip-final H- to ap-initial L). It is directly dominated by the IP node, skipping ip and violating Exhaustivity, and this reflects the intuition that this PW does not start a new prosodic constituent larger than a PW.

Alternatively, instead of increasing the number of prosodic categories, another widely adopted alternative in the literature (e.g., Ito & Mester, 2012) is to allow the recursion of the existing category, violating the Non-recursivity constraint (Selkirk, 1996). In Korean, Baek & Yun (2018) and Lee (2022) proposed the recursion of AP to derive the prosodic structure of relative clauses more transparently from the syntactic structure.

Assuming that the Post-CP PW occupies a high position in the syntactic structure, and is in a right element tolerance relationship with the preceding clause, a prosodic structure like that in Figure 3.15b may be posited. In this structure, the initial syllable of /kat^{h}ejo/ is AP-medial, and its weak realization is justified. However, to make sense of the large fall from /kɔ/ to /kat^{h}ejo/ shown in Figure 3.13, there needs to be additional assumptions made about how this recursive structure is tonally implemented. This has not been studied in previous works, and is not attempted in the current study either.

In short, while Post-CP PW /kat^{h}.../ exhibited acoustic characteristics of a low BI, an actual instance of the PW exhibited inconsistent tonal evidence that it may not initiate a new AP. This is a problem for a prosodic structure that fully obeys the Strict Layer Hypothesis, in the sense that it breaks the generalization that AP-initial lenis obstruents are acoustically strong. To rescue the generalization, alternative prosodic structure that violates the assumptions of the Strict Layer Hypothesis may be considered. I sketched two proposals here, one that introduces additional level of prosodic constituent labels (ap and ip), and one that makes use of prosodic recursion.

3.6.2 Weak Rise Lenis PW: Post /an/ (NEG) lenition

Finally, I will explore a small subset of PW-initial weak Lenis tokens with a higher F0 than the preceding syllable (i.e., Weak Rise tokens). Typically, Lenis PW-initial syllables have a lower F0 than the preceding syllable if that PW is AP-medial (e.g., the onset of the second PW in $[[\text{L} + \text{H}]_{PW}[\ldots\ \text{L+ Ha}]_{PW}]_{AP}$). Lenis PW-initial syllables with a weak realization can only have a higher F0 (i.e., Weak Rise) in two cases. First, if the Lenis PW is monosyllabic and it is at the end of an AP, it bears a phrase final Ha tone, and thus has a higher F0 compared to the preceding syllable. Second, if the Lenis PW is preceded by a monosyllabic PW and this Lenis PW is the second PW in the AP, the initial syllable of this Lenis PW bears a phrase medial +H tone, and thus has a higher F0 compared to the preceding syllable.

Out of 23,702 total Weak Lenis PW-initial syllables, 6,483 were considered to be Weak Rise Lenis PW-initial syllables. 638 out of these 6,483 tokens were preceded by a monosyllabic PW /an/ (NEG). An example sentence of /an/ is given in (7).

(7) kɨɾɔnik*a an pis*ata-ko seŋkakhe-sɔ

So NEG be expensive-COMP think-because
'So since I think it is not expensive...'

File name: s26f32f6, time slice: 91.31182 - 92.656395 in Seoul Corpus

/an/ precedes any predicate in the sentence to negate it. In previous works, /an/ has been described as a verbal prefix by some authors (e.g., Kim, 1999), and the negated verb as a lexical item is inserted in the syntactic structure. Sells (2015) argued against this claim by showing that /an/ can be ambiguous in its scope when it modifies a verbal complex (verb + causative). Sells (2015) illustrated this point with the following example (originally from Whitman (2005, 886)) in (8). The words are transcribed into IPA by me.

(8) ɔmɔni-k*esɔ t*al-ɨl an mɔk-ke ha-ʃi-ɔs*-ta

mother-HON.NOM daughter-ACC NEG eat-COMP do-HON-PAST-DECl
a. 'Mother did not make the child eat.'

b. 'Mother made the child not eat.'

The sentence in (8) is ambiguous, depending on whether /an/ modifies the entire complex ([an [mɔk-ke ha...]]) which corresponds to the first meaning in (a) or just the complement verb ([[an mɔk-ke] ha]), which corresponds to the second meaning in (b) (Sells, 2015). Sells (2015) argued that if /an/ is a prefix at the lexical level, it is hard to explain why in the first meaning, it seems to negate the second verb (/ha/ ‘do (causative)’) though it does not immediately precede it (p.217).

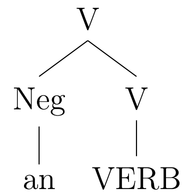


Figure 3.16: Complex head consisting of /an/ and the verb, proposed in Sells (2015)

Instead, Sells (2015) claimed that /an/ and the following predicate form a complex head consisting of two heads, as in Figure 3.16. The scopic ambiguity of /an/ and a complex verb then would be represented as in Figure 3.17.

(a) ‘did not make (the child) eat’ (b) ‘made (the child) not eat’

Figure 3.17: Structural representation of scopal ambiguity of /an/ and verbal complex in (8)

Though Sells (2015) did not discuss the prosodic phrasing of the verbal complex, it is expected that the two meanings in (8) are prosodically phrased differently, as they have different syntactic structures. In the left structure, /an/ and the first verb /mɔk-ke/ are in the left element constraint relationship, meaning /an/ and the first verb are less likely to be phrased together. On the other hand, in the right structure, /an/ and the first verb /mɔk-ke/ are allowed to be prosodically phrased together. Unfortunately, it seems that the Seoul Corpus does not have any instances where /an/ exhibits this particular scopic ambiguity and therefore testing this prediction awaits further studies. In Chapter 4, I will present a similar case of scopal ambiguity involving a polysemous word /kɨman/ ('stop'/'start'), and show how the syntactic structure affects the prosodic phrasing of PWs.

3.7 Chapter summary

To conclude the chapter, this study presents a novel way of investigating prosody in spontaneous speech corpus. The syllables in the spontaneous speech corpus were parameterized in terms of tonal and segmental cues. The tonal cue was operationalized as relative pitch difference with respect to the preceding syllable ($\text{MAX}\Delta_{\text{F0}}$), and the segmental cue was operationalized as the velocity of intensity change, compared to the preceding syllable (CVE). From PW-initial syllables parameterized with these cues, prosodic juncture categories that may be interpreted as different Break Indicies values are identified using a K-means analysis. The results indicated that the type of phonetic cue that was informative in identifying the prosodic juncture categories varied by the laryngeal category of the syllable onset. While the segmental cue was informative for Lenis, the tonal cue was informative for Aspirated. These results are unexpected given the emphasis that the prosodic boundaries are tonally delimited in Seoul Korean (Jun, 1993, 2000) and that segmental variation of lenis obstruents is unreliable due to its optionality (Jun, 1994).

PW-initial syllables that had acoustic characterization of being phrase-medial (BI < 2) were investigated further. It was found that a majority of these tokens were incomplete nouns and PWs that are sister to CPs, which are expected to be phrased together with the preceding PWs, given their syntactic position (Jun, 1993).

Further examination of a post-CP Lenis PW /kat$^{\text{h}}$.../ has revealed that a Lenis PW may follow an AP-final syllable and yet not start a new AP, which follows from the tonal and segmental profile of some of the initial syllables of /kat$^{\text{h}}$.../. Instead of arguing that these are AP-initial syllables that are exceptionally voiced, I have sketched alternative prosodic structures that do not need to make exceptions to the generalization.

Finally, Lenis PWs that follow a monosyllabic PW /an/ ('NEG') were discussed. The weak realization of lenis following /an/ supported the claim that /an/ forms a complex verbal head with the following verb (Sells, 2015).

The results of this study assume that when a PW-initial syllable exhibit the same acoustic profile as a PW-medial syllable, it is not perceived as starting a new prosodic constituent above the level of PW (AP). However, it needs to be empirically tested how listeners process the prosodic structure of cases where the tonal cue suggests a prosodic break, while the segmental cue does not, as in Figure 3.13. In order to test this empirically, a carefully designed perception experiment is in order. One way to test how listeners process the prosodic structure is to test how they behave in a prosodic disambiguation experiment, involving a pair of syntactically ambiguous sentences that differ in their prosodic phrasing. The next chapter presents such a case in Seoul Korean, which involves a polysemous word /kɨman/ ('stop'/'start').

CHAPTER 4

A CASE STUDY ON SYNTAX, PROSODY, AND SYNTAX-PROSODY MAPPING OF *KEUMAN* + VERB IN SEOUL KOREAN

4.1 Introduction

This chapter presents a case study on a less-studied pair of syntactically ambiguous sentences in Korean, which can be disambiguated by prosodic phrasing. It involves a polysemous adverb *keuman* /kɨman/, which can mean either ‘stop Verb-ing’ or ‘start Verb-ing’, when it is immediately followed by a verb (henceforth the /kɨman/ + Verb construction) (Noh, 1983; Gim, 2004; Kim & Jeong, 2019; Tan, 2023). The rest of this dissertation presents a production (see Chapter 5) and perception experiment (see Chapter 6) using this pair of sentences to investigate the role of segmental realization of lenis obstruents in prosody-syntax mapping in Seoul Korean. Before presenting these experiments, this chapter lays the groundwork for the rest of the dissertation by providing a preliminary analysis of the syntactic and prosodic structures of the /kɨman/ + Verb construction, as well as how these prosodic structures are mapped from the syntactic structures.

To investigate how prosodic structure is computed from the acoustic signal, and how such computation affects the parsing of syntactic structure, this dissertation uses a method referred to as ‘prosodic disambiguation experiments’ (see §1.3.3 for a review of the method).

As reviewed in Chapter 1, there are several experimental ways to investigate how listeners compute prosodic phrasing of an acoustic signal. One popular and arguably the most direct way is to ask listeners to label the acoustic signal using a prosodic transcription system, e.g., ToBI. Prosodic transcription labels, especially the tonal labels, often carry information about the prosodic structure, which includes the location and the size of prosodic boundaries (e.g., in K-ToBI (Jun, 2000), 'Ha' marks both the high F0 target and that there is an AP boundary). However, this method has the shortcomings discussed in §1.3.1. To briefly introduce one of the shortcomings without repeating the whole discussion, prosodic transcription necessarily relies on transcribers' subjective judgments of prosodic boundaries which may not always be totally independent of the syntactic structure of the sentence.

Instead, prosodic disambiguation experiments are based on the reasonable premise that listeners' ultimate goal is to understand the meaning of the sentence, and computation of prosodic structure is a means to achieve this goal. In a prosodic disambiguation experiment, listeners are presented with sequences of words that are ambiguous between two meanings, but can be disambiguated by their difference in prosodic structure. Assuming each meaning corresponds to a distinct syntactic structure, this means the same sequence of words have two distinct syntactic structures, which then map onto two distinct prosodic structures. Using this paradigm, experimenters can manipulate acoustic cues (e.g., F0 contour or realization of word-initial segments) in experimental stimuli. Depending on listeners' responses, we can learn about acoustic cues listeners rely on when they compute the prosodic structure of the sentence to disambiguate the meanings.

In Chapter 1, I have reviewed an example of a prosodically disambiguated syntactic ambiguity in Korean, i.e., the WH/Yes-No question ambiguity (Jun & Oh, 1996; Yun & Lee, 2022; Farinella & Lee, 2024). To briefly recapitulate, the sentence in (1) (taken from (3) on p. 42 in Jun & Oh (1996)) can be interpreted as either a WH question or a polar question, depending on the prosodic phrasing. Jun & Oh (1996) argued the indefinite pronoun (/ɔntʃe/ 'any time' in (1a)) and the verb phrase (/ɔtʃirɔwɔ-jo/) form separate APs in a Yes-No question, while the WH word (/ɔntʃe/

‘when’ in (1b)) and the verb phrase (/ɔʧirɔwɔ-jo/) form a single AP (Jun & Oh, 1996) in a WH question.

(1) /aʧumɔni-nɨn ɔnʧe ɔʧirɔwɔ-jo?/

madam-TOP when/any time be dizzy-HON

a. Yes-No question: Is there any time that you feel dizzy, madam?

b. WH question: When do you feel dizzy, madam?

(3) on p. 42 in Jun & Oh (1996)

The two sentence meanings (WH or Yes-no question) are distinguished by two prosodic factors: whether there is a prosodic boundary between the WH-word and the verb phrase and the type of boundary tone on the sentence-final syllable. When there is a prosodic boundary between the WH-word and the verb phrase, and the sentence-final syllable has an H% boundary tone, the sentence is interpreted as a Yes-No question. When there is no prosodic boundary between the WH-word and the verb phrase, and the sentence-final syllable has an LH% boundary tone, the sentence is interpreted as a WH question. As reviewed in Chapter 1, previous studies have shown that listeners can use acoustic cues such as the F0 transition between the WH-word and the verb phrase (Jun & Oh, 1996; Yun & Lee, 2022), and the segmental realization of the verb-initial lenis stop (Farinella & Lee, 2024) to disambiguate the two meanings.

However, these studies have the following three limitations. First, while it has been shown that WH-question and Yes-No question meanings can be disambiguated by prosody, why this is so is not fully worked out. Specifically, it is not clear why the verb phrase does not start a new AP in a WH question but does in a Yes-No question. Second, both Yun & Lee (2022) and Farinella & Lee (2024) have shown that listeners are more biased towards interpreting the sequence of WH-word and verb phrase as a WH question, when the acoustic signal has been manipulated so that the prosodic phrasing is made more ambiguous. In other words, listeners showed variable responses for the same acoustic signal rather than choosing one of the two meanings 100% of the time for a given acoustic signal. However, it is left unexplained why this is so and what derives this variation. Finally, from the experiment design perspective, it is hard to make a minimal set of experimental conditions for WH/Yes-No question ambiguity as there are two independent factors that can affect listeners' responses: prosodic phrasing and boundary tone. For the purpose of this dissertation, it is desirable to have a case where the two meanings are disambiguated by a single prosodic factor, i.e., prosodic phrasing. This third reason is the main reason why I do not pursue the WH/Yes-No question ambiguity in this dissertation, even though it is a well-studied case of prosodically disambiguated syntactic ambiguity in Korean. Instead, in the rest of this dissertation, I investigate a less studied pair of syntactically ambiguous sentences in Korean which can be disambiguated by prosodic phrasing only, and in so doing, address the first two limitations.

This pair of sentences involves a polysemous adverb /kɨman/. When /kɨman/ is immediately followed by a verb, the sentence can either mean 'stop Verb-ing' or 'start Verb-ing', depending on the prosodic phrasing (Noh, 1983; Gim, 2004; Tan, 2023). For example, the sentence in (2) is grammatical in both contexts in (2a) and (2b). I will refer to /kɨman/ spoken in these two contexts as kɨman$_{stop}$ and kɨman$_{start}$, respectively.

(2) /kɨman pul-mjɔn ʧoh-kes*-ta/
up to that degree/amount blow-COND be good-FUT-DECL

a. Grammatical in a context where a child has been blowing something, e.g., a balloon, to the extent that the balloon is about to pop (i.e., 'stop blowing')

b. Grammatical in a context where a child has not been blowing balloons, though the child has been asked to do so, and the speaker wants to say that it is about time to blow them (i.e., 'start blowing')

While the ambiguity of /kɨman/ has been documented in the Korean linguistics literature (Noh, 1983; Gim, 2004; Kim & Jeong, 2019; Tan, 2023), the focus of these works has been providing descriptive accounts of the syntactic and semantic properties of this adverb. Gim (2004) is the only study that mentions that the two meanings can be disambiguated by prosody but does not provide any further analysis of how the prosodic structures differ, and whether native listeners can use the acoustic difference to disambiguate the two meanings. The key observation in Gim (2004) is that when /kɨman/ is used with the 'stop' meaning, it is pronounced together with the following verb without a prosodic juncture, while when it is used with the 'start' meaning, there is a prosodic juncture between /kɨman/ and the following verb.

It is important to note early in this chapter that working out details of syntax and semantic properties of /kɨman/ is outside the scope of this dissertation. Instead, this chapter has three goals. In pursuing these three goals, I address the gaps left in the previous literature on WH/Yes-No question ambiguity in Korean: (i) why there may be a variation in prosodic phrasing of a given syntactic structure and (ii) why two meanings of the same sequence of words exhibit different prosodic phrasings.

First, based on previous studies of /kɨman/, I propose a preliminary analysis of the syntactic structures of the two /kɨman/ sentences. I will argue that the /kɨman/ + Verb construction exhibits syntactic and semantic similarities to two well-studied cases of syntactic ambiguity: the verbal/sentential LE particle in Mandarin Chinese (Soh, 2009; Wang, 2018), and the clausal/manner adverb in languages including English (Ernst, 2002; Keizer, 2018; Kojadinović, 2022), Japanese (Kubota, 2015), and Hungarian (Egedi, 2009).

Second, based on Gim (2004)'s informal observations, I make a proposal for the prosodic structures of the /kɨman/ + Verb construction. Building on Gim (2004)'s observations, I will also offer my own judgments about the variation in the prosodic phrasing /kɨman/ + Verb sentences, which are not discussed in the previous literature.

Third, I sketch an analysis of the variable mapping relations between the syntactic and prosodic structures of the /kɨman/ + Verb construction, using a constraint based grammar (Optimality Theory (OT) (Prince & Smolensky, 1993/2004) and Maximum Entropy (MaxEnt) grammar (Goldwater & Johnson, 2003)). In particular, I argue that the difference in the prosodic phrasing of the two meanings of /kɨman/ arises from the fact that different elements in the sentence are under focus in the two meanings, which would also explain the variation in prosodic phrasing.

This chapter proceeds as follows. §4.2 reviews the previous work on the semantic, syntactic, and prosodic properties of the /kɨman/ + Verb construction (Noh, 1983; Gim, 2004; Kim & Jeong, 2019; Tan, 2023). Building on these previous descriptions, §4.3 presents a preliminary proposal for the syntactic structures and §4.4 presents the prosodic structures of the /kɨman/ + Verb construction. §4.5 discusses the mapping relations between the proposed syntactic and prosodic structures. §4.6 provides a general discussion regarding limitations of preliminary analyses in this chapter, and discusses possible alternative prosodic structures based on more recent proposals for the intonational phonology of Seoul Korean involving prosodic recursion (Baek & Yun, 2018; Lee, 2022). A chapter summary is provided in §4.7.

4.2 Background

Previous studies have argued that /kɨman/ can be used in multiple discourse contexts. There are largely two meanings of /kɨman/, which I loosely translate as 'stop' and 'start' (Noh, 1983; Gim, 2004; Kim & Jeong, 2019; Tan, 2023). The meanings of /kɨman/ are subtle and cannot be perfectly translated with 'stop' and 'start', as the review in this section will show. I nonetheless use these two words as shorthand to refer to the two meanings of /kɨman/ in this dissertation.

This section proceeds as follows. §4.2.1 reviews previous studies that investigated the meaning of /kɨman/ diachronically (Kim & Jeong, 2019) and synchronically (Noh, 1983; Gim, 2004; Tan, 2023). §4.2.2 discusses a syntactic restriction on /kɨman/ (Noh, 1983; Gim, 2004), which is supported by the corpus investigation conducted in Gim (2004). §4.2.3 introduces a claim made in Gim (2004) that the meaning of the /kɨman/ + Verb construction can be prosodically disambiguated. §4.4.2 discusses the variation in the prosodic phrasing of /kɨman/ + Verb constructions.

4.2.1 Meaning of /kɨ.man/

The word /kɨman/ is comprised of two morphemes: /kɨ/ 'that (demonstrative)' and /man/ 'up to (degree/amount)' though it is likely to be lexicalized as a monomorphemic word due to its high token frequency (Gim, 2004; Tan, 2023).

Kim & Jeong (2019) investigated the development of the polysemy of /kɨman/. The authors cited historical texts that showed in the 15th and 16th centuries, this word was combined with the verb /hata/ ('to be') and was used to describe that something was 'about that size/degree', and to modify the following verb to mean that doing the verb was done 'to an appropriate degree'. Then starting in the 17th century, there were more uses of /kɨman/ as an adverb that expressed that 'the action had happened to the right degree', and 'therefore it needed to stop'. Finally, starting from the late 19th century, the meaning of /kɨman/ was expanded to mean 'discontinuing the action that is taking place but not explicitly spoken in the utterance', and also to mean that a new action took place, implying it had not taken place before (Kim & Jeong, 2019). In short, Kim & Jeong (2019) showed the original meaning of /kɨman/ developed from a morphologically transparent 'to that degree' to its present day meaning that is more abstract and subject to discourse pragmatics, like 'stop' and 'start', giving rise to the ambiguity, when a verb immediately follows /kɨman/.

Other authors have reported that /kɨman/ is polysemous in modern Korean, and one of the meanings can be translated as 'stop' (Noh, 1983; Gim, 2004; Tan, 2023). As for the other non-'stop' meaning, authors have disagreed slightly about the variety of contexts in which the word is used. Gim (2004) identified two contexts where the non-stop meaning of /kɨman/ is used: one where the new action is simply starting and one where the new action has taken place unintentionally. Tan (2023) classified the contexts differently: one where the new action is consciously executed, one where it is unconsciously executed, and one where the new action is inevitable. While these distinctions are interesting from a discourse-pragmatics perspective, I will simply group these contexts into one, as they all describe the execution of an action denoted by the verb, which I loosely translate as 'start'. While Gim (2004) and Tan (2023) described the various semantic/pragmatic flavors of the 'start' meaning of /kɨman/, they did not provide a structural analysis of how the two meanings 'start' and 'stop' arise. It is therefore not clear whether /kɨman/ has a single denotation, and the difference in the meaning is due to its position in the syntactic structure; or if there are two distinct lexical items which share the same pronunciation.

On the other hand, Noh (1983) argued that the core meaning of /kɨman/ is 'discontinue', and that the context decides what action it modifies.[1] The sentences in (3) are from Noh (1983). Glosses and translations are mine.

(3) (a-c) from Example (8) on p. 136 and (d) from (23) on p. 149 in Noh (1983)

a. /toɾ-ija kɨman koŋpuhe-ɾa/
   tol-VOC kɨman study-IMP
b. Grammatical if Tol (name of a person) has been studying
c. Grammatical if Tol has been doing something else, e.g., playing (with toys)
d. /toɾ-ija kɨman nol-ko koŋpuhe-ɾa/
   tol-VOC kɨman play-and study-IMP

[1]Noh (1983) discussed the meaning of /kɨman/ and /iman/. /iman/ is a very similar word, used in slightly different contexts, as it is comprised of another demonstrative /i/ ('this') and the same /man/ ('up to degree/amount') in /kɨman/. I limit the discussion to /kɨman/ here.

'Tol, stop playing and study' (a more explicit version of the sentence in (a) spoken in the context (c))

The sentence (3a) spoken in the context in (3b) means 'study only up to that point, and no more', which I loosely translate as 'stop studying'. The sentence (3a) spoken in the context in (3c) means 'do the current ongoing action (e.g., playing with toys) only up to that point and no more, and study', which I loosely translate as 'start studying'.

Noh wrote that the sentence in (3a) spoken in the context in (3c) can be paraphrased as in (3d), where the verb 'to play' is explicitly mentioned. Noh argued that the meaning of /kɨman/ is still 'discontinue', except that it modifies the verb 'to play'. In his analysis, the ambiguity of /kɨman/ essentially arises from the fact that the verb that it modifies can be elided from the sentence, and therefore it can be interpreted as modifying the verb that follows it in the surface form ('to study'), or the verb that is not spoken ('to play'). While Noh's effort to provide a unified analysis of the two meanings of /kɨman/ is commendable, the details of how the meaning is composed semantically is left unspecified. The analysis also suffers from the major problem that /kɨman/ does not always imply the discontinuation of the action that is not spoken. This point is also addressed in Kim & Jeong (2019), who argued that the recently developed (late 19th century) meaning of /kɨman/ includes the meaning of 'a change in state', without implying that the ongoing action has stopped (p. 262). I will illustrate this with the following example in (4b).

(4) a. /minsu-nɨn kɨman uɾ-ɔs*-ta/

Minsu-TOP kɨman cry-PAST-DECL
Loosely: 'Minsu started crying'

b. /minsu-nɨn kɨman nore-rɨl tɨt*a-ka uɾ-ɔs*-ta/

Minsu-TOP kɨman song-ACC listen-while cry-PAST-DECL

'Minsu started crying while listening to songs (e.g., because he became sad)'

(4a) and (4b) are meant to be similar to the Verb-elided version in (3a) in the context in (3c) and the explicit unambiguous paraphrased sentence in (3d), respectively. If Noh (1983) is right, the sentence in (4a) implies that some ongoing action has been discontinued, and Minsu started crying. However, from its explicit paraphrase in (4b), we know that the salient ongoing action from the context is 'listening to songs', and the sentence does not imply that Minsu stopped listening to songs as he started to cry. Instead, it is pragmatically more likely that Minsu kept listening to songs as he started crying. This example therefore clearly illustrates that the 'start' meaning of /kɨman/ does not presuppose 'discontinuation' of the 'salient action in the context', contrary to Noh (1983).

Noh (1983) in fact also used examples like (4a) where /kɨman/ does not imply the discontinuation of any ongoing action. For those cases, Noh argued that /kɨman/ introduces a presupposition of a state where the following verb has not taken place and expresses the discontinuation of that state. Using the example in (4a), it works as follows. In (4a), kɨman$_{start}$ modifies the state where Minsu cried. In doing so, it first presupposes the negation of this state, i.e., a state where Minsu did not cry. Then it expresses the discontinuation of this state, meaning he stopped *not* crying, i.e., he started crying.

It is not clear from Noh (1983)'s proposal whether in some cases, /kɨman/ implies the discontinuation of some unspoken action, and in others, it conveys the presupposition that the event did not take place before. I instead argue that kɨman$_{start}$ always makes a presupposition of the negation of the state denoted by the following verb, and discontinuation of that state. Coming back to (3a) in the context (3c), /kɨman/ also expresses the presupposition that Tol has not been studying, and that the speaker wishes a discontination of this state of 'Tol's not studying'. We only get the inference that Tol is going to discontinue the ongoing action (i.e. ‘playing with toys’), since it is not pragmatically likely to both play with toys and study simultaneously.

To summarize, it seems clear that authors in Korean linguistics have reached the consensus that /kɨman/ has more than one meaning. While Noh (1983) argued that there is a single denotation for /kɨman/, he left the mechanisms in which the word is used in different contexts unspecified. I follow Noh (1983) and argue that the core meaning of /kɨman/ is ‘discontinue’, and the two meanings arise depending on what it modifies. kɨman_{stop} simply modifies the following action, and therefore conveys the meaning of ‘Stop V-ing’. On the other hand, kɨman_{start} seems to modify the ‘negation of the state where the verb took place’, and it conveys the discontinuation of that state, i.e., ‘Stop not V-ing’, which may be loosely expressed as ‘Start V-ing’.

4.2.2 Linear adjacency constraint on Adverb (including /kɨman/) + Verb constructions

While previous studies on /kɨman/ did not provide a structural analysis of how the ambiguity arises, Gim (2004) made a valuable observation that there is a linear adjacency constraint that is only imposed on kɨman_{stop}, which seems to be a general constraint on adverbs, like /kɨman/ that can be used with two meanings. This linear adjacency constraint is given in (5). This is illustrated in (6a) and (6b). Note that the discourse particle /ʧom/, whose distribution is free in Korean, can intervene between kɨman_{stop} and the following verb. This is argued to be the only exception to the linear adjacency constraint (Gim, 2004).

(5) Linear adjacency constraint

When /kɨman/ is used with the ‘stop’ meaning, it must be immediately followed by a verb, without any other words intervening between them.

When it is used with the ‘start’ meaning, there can be other words intervening between /kɨman/ and the verb.

(6) /Example (17b, p. 58) in Gim (2004)(numberings, glosses and translations are mine)

a. /ton kɨman ʧom pilljɔ-ʧwɔ/

money(-ACC) kɨman ʧom borrow-give
Ambiguous: 'Please lend (me) some moeny'

or 'Please stop lending money'

b. /kɨman ton ʧom pilljɔ-ʧwɔ/

kɨman money(-ACC) ʧom borrow-give
Not ambiguous: 'Please lend (me) some money'

(*'Please stop lending money')

When /ton/ ('money') intervenes bewteen /kɨman/ and the verb 'to lend', as in (6b), the sentence cannot mean 'stop' lending.

/kɨman/ V	Stop meaning	Start meaning	Indistinguishable	Total
Yes	191	243	35	469
No	15	629	0	646
Total	206	874	35	1,115

Table 4.1: Distribution of /kɨman/ in corpora studied in Gim (2004)

Gim (2004) further provided evidence for this linear adjacency constraint from Korean text corpora. He studied the 21st Century Sejong planned tagged corpus, which contains 8,719,237 Eojeols. Eojeols were tagged with syntactic categories which allowed Kim to search 1,115 occurrences of /kɨman/ with the adverb tag in this corpus. The cases where /kɨman/ was used with the transparent meaning of 'to be adequate amount/degree' was not included in this number of occurrences. Out of 1,115 occurrences, it was used to mean 'stop' 206 times (18.48%), and it was used to mean something other than 'stop' (i.e., 'start' using the binary distinction used in the current dissertation) 874 times (78.39%), which was considerably higher than 206. For the remaining 35 occurrences, it was not possible to tell by the author which was the intended meaning based on the context in the corpus.

When it meant 'stop', the immediately following word was a verb 92.72% of the times (191/206), whereas when it meant 'start', it was a verb only 27.80% of the times (243/874). Furthermore, for all 35 cases where the meaning was not identifiable, the immediately following word was a verb. These results provide support for the linear adjacency constraint that requires the verb to be immediately adjacent to /kɨman/, when it means 'stop'.

The linear adjacency constraint suggests that the position of /kɨman/ in the syntactic structure may be different, depending on the meaning that it is used. The fact that there can be intervening words between kɨman$_{start}$ and the verb suggests that kɨman$_{start}$ may be adjoined to the structure at a higher position than the Verb Phrase. This is in line with the idea that the meaning of kɨman$_{start}$: the negation of the proposition that the verb took place. From a structural perspective, when it occupies a higher position, it can take a wider scope than just the verb, i.e., the entire proposition. In §4.3, I show that this linear adjacency constraint generalizes to other adverb + verb constructions in Korean, and possibly adverbs of other languages. I also present a preliminary proposal for the syntactic structures of the /kɨman/ + Verb construction.

4.2.3 Prosodic disambiguation of 'kɨman + Verb' construction

Among the previous studies on /kɨman/, Gim (2004) seems to be the only one that mentions the sentence in (2) differs in prosody depending on the meaning. The following quote from Gim (2004) is a summary of the author's observations on the prosodic characteristics of the two meanings of /kɨman/, which provide valuable insights for both prosodic and syntactic structures for the '/kɨman/ + Verb' construction. Note that the original source is written in Korean and I translated it into English, trying to keep the terminology as close to the original as possible. I leave both the original Korean and my English translation below.

> '그만'의 경우 '금지'의 용법으로 사용되면, '그만'의 억양(강세)이 높아진다. 전체 명제에서 명제의 진리값을 변경하는 중요한 요소이기 때문에 초점이 놓일 수밖에 없다. 또한 후행하는 동사와 거의 연접시켜(붙여서) 발화된다. '그만 + 동사' 구성이 하나의 서술 단위를 형성한다.
>
> ...
>
> 반면에 [그 외의 용법으로 사용되면], '그만'은 후행하는 동사와 연접되지 않고 발화된다 ... '그만'이 후행하는 사건(진술) 전체를 수식하기 때문에 동사에 종속적으로 연접할 필요가 없다. 또한 '금지'의 용법과 비교했을 때, 후행 요소 (주로 동사)에 비해 억양(강세)을 약하게 받는다. 정보의 초점 역할을 하지 못하기 때문이다. (Gim, 2004, p. 62)

> When /kɨman/ is used with the 'stop meaning', the intonation (accent) on /kɨman/ becomes higher. It is natural that it receives a focus, since it is an important factor that changes the truth condition of the entire proposition. Moreover, it is pronounced together (without a juncture) with the following verb. The '/kɨman/ + Verb' construction forms a predicative unit.
>
> ...
>
> [On the other hand, when it is used with the 'start meaning'], /kɨman/ is not pronounced as a connected unit with the following verb. It does not need to be pronounced together with the verb because it modifies the entire event (proposition) that follows it. Moreover, compared to when it is used with the 'stop meaning', it also receives a weaker intonation (accent) compared to the following element (usually a verb), as it does not serve as the focus of information. (Gim, 2004, p. 62)

There are two key points in Gim (2004)'s description. First, the prosodic phrasing of /kɨman/ and the verb that follows it is different between the two meanings. When it means 'stop', /kɨman/ and the verb are pronounced together without a prosodic juncture, while when it means 'start', there is a prosodic juncture between /kɨman/ and the verb.

Second, it suggests that the information structure of the two /kɨman/ + Verb constructions is different. When it means 'stop', /kɨman/ is under focus, while when it means 'start', it is not, and the verb is under focus. I use 'focus' here to refer to an element that is compared with a set of alternatives (Rooth, 1992; Katz & Selkirk, 2011).

While Gim (2004) makes a clear claim that /kɨman/ receives focus when it means 'stop', it is less clear whether it is the verb that receives focus when it means 'start'. I argue that this assumption is not ungrounded given that kɨman$_{start}$ seems to occupy a high position in the syntactic structure. Moreover, I argue that this assumption is essential to account for the variation of prosodic phrasing.

I discuss these points further in §4.4 and propose an autosegmental-metrical structure of /kɨman/ + Verb construction with two meanings, based on the K-ToBI model (Jun, 1993). I show further in §4.5 that the proposed prosodic structures which capture the observations of Gim (2004) are expected given the syntactic and informational structures of the /kɨman/ + Verb construction.

4.3 Syntactic analysis

Building on previous work on the adjacency restriction of /kɨman/ and the following verb (Gim, 2004), as well as work on the polysemy of /kɨman/ (Noh, 1983), I propose that kɨman$_{stop}$ is adjoined at a lower position, within the verbal projection, while kɨman$_{start}$ is adjoined at a higher position than the verbal projection.

I argue that the syntactic structure and the polysemy of /kɨman/ is similar to two well studied cases of syntactic ambiguity, i.e., the verbal/sentential LE in Mandarin Chinese and the clausal/manner adverbs of other languages, in §4.3.2 and §4.3.3.

I argue three points in discussing these cases. First, though it may seem unintuitive that a single word /kɨman/ can have seemingly opposite meanings of 'stop' and 'start', it is not undocumented in other languages, i.e., verbal/sentential LE in Mandarin Chinese. Second, the fact that /kɨman/ can occupy a high or low position in the syntactic structure is not unique to this word, but a general tendency for a subset of adverbs in Korean as well as in other languages.

4.3.1 Syntactic structures of '/kɨman/ + Verb'

In §4.2.2, I have reviewed the linear adjacency constraint (see (5)) that /kɨman/ with the 'stop' meaning needs to be adjacent to the verb and /kɨman/ with the 'start' meaning does not have this syntactic restriction. The example in (6) is repeated in (7) without the discourse particle /ʧom/ for simplicity. One way to model this difference is by adjoining kɨman$_{stop}$ within the verbal projection, as a sister to the verb, and adjoining kɨman$_{start}$ at a higher position than the verbal projection. Note that the exact position, or the name of its projection is not specified here. Future work should investigate what this projection may be, or where exactly kɨman$_{start}$ gets adjoined. Here I only argue that this is at a higher position outside of the verbal projection. Figure 4.1 illustrates my proposed syntactic structures for the sentences in (7).

(7) /Example (17b, p. 58) in Gim (2004) (repeated from (6))

a. /ton kɨman pilljɔ-ʧwɔ/

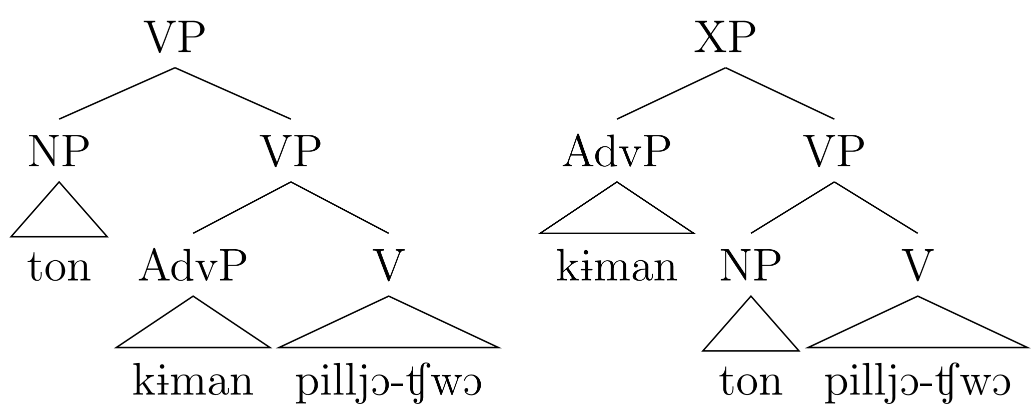
money(-ACC) kɨman borrow-give
Ambiguous: 'Please lend (me) some moeny'

or 'Please stop lending money'

b. /kɨman ton pilljɔ-ʧwɔ/

kɨman money(-ACC) borrow-give
Not ambiguous: 'Please lend (me) some money'

(*'Please stop lending money')



(a) Money + kɨman$_{stop}$ + Lend ('Stop lending money')

(b) kɨman$_{start}$ + Money + Lend ('Start lending money')

Figure 4.1: Proposed syntactic trees for the sentences in (7)

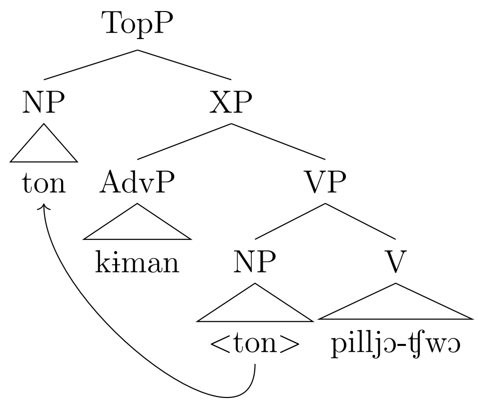


Figure 4.2: Money + kɨman$_{start}$ + Lend
('Start lending money')

In (7a), the word order 'Noun + /kɨman/ + Verb' is ambiguous between the two meanings of /kɨman/. The structure in Figure 4.1a is for kɨman$_{stop}$. To get the same word order while maintaining the higher position of kɨman$_{start}$ we need to move /ton/ to the initial position. While the default word order is 'SOV' in Korean, 'OSV' can appear either by scrambling or object topicalization (Lee & Ramsey, 2000; Frenck-Mestre et al., 2022). I propose the word order 'Noun + kɨman$_{start}$ + Verb' is derived via Object fronting, either by scrambling or by topicalization. This is illustrated in Figure 4.2 where the noun phrase /ton/ ('money-ACC') has moved to the initial position of the sentence via toplicalization. As a result, the surface word order in Figure 4.2 is the same as in Figure 4.1a, and the difference in meaning is derived from the difference in the position that /kɨman/ occupies in the syntactic structure.

To determine the exact syntactic structures for the /kɨman/ + Verb construction would require longer sentences with more words that would show the exact height that /kɨman/ occupies. In this dissertation, I only propose that kɨman$_{start}$ is attached to the structure at a higher position, while kɨman$_{stop}$ is immediately attached to the verb as its sister.

4.3.2 Verbal/sentential *LE* in Mandarin Chinese

In §4.2.1, I've argued the denotation of /kɨman/ is 'discontinue', following Noh (1983). Additionally, I have argued that the meaning of 'start' is derived from this denotation, as /kɨman/ can refer to the negation of the state denoted by the verb when it is adjoined to the structure at a higher position.

While the fact that a word that means 'discontinue' can also mean 'start' seems counterintuitive, something similar has been reported in Mandarin Chinese with the particle LE, which has both perfective and inchoative meanings, depending on the position in the sentence as shown in (8), borrowed from Example (11) on p. 628 in Soh (2009).

(8) a. Ta ma-le ta de haizi

S/he scold-LE s/he POSS child
'S/he scolded her/his child' (verbal le: perfective)

b. Ta ma ta de haizi le

S/he scold s/he POSS child le
'S/he is scolding his child (which s/he was not doing before/contrary to what one may expect).'

When the particle LE appears right after the verb (/ma/) as in (8a), it expresses the meaning that the action denoted by the verb is terminated (i.e. the perfective meaning), while the sentence final LE in (8b) expresses a 'change of state' (i.e. the inchoative meaning) or the action is 'contrary to expectation' (Soh, 2009). The former is refered to as the verbal LE, and the latter is referred to as the sentential LE, due to their syntactic distribution (e.g., Soh, 2009; Wang, 2018).

Similar to the /kɨman/ + Verb construction, the meaning of LE can be ambiguous between the perfective and inchoative readings in sentences like (9), borrowed from Example (8b) on p. 4 in Wang (2018). The notes in parentheses are added by me. Wang argues that the perfective LE is still the same verbal LE though it does not immediately follow the verb. Going through his analysis of the syntactic structures of two LEs in this particular example is out of scope for this study. Here, I only note that the particle LE can be ambiguous between two meanings in the same sequence of words, just as /kɨman/.

(9) Ta he jiu le
He drink wine LE
a. He drank wine (Verbal LE: perfective)
b. He has begun drinking wine (used to not drink wine) (Sentential LE: inchoative)

While the particle LE of Mandarin Chinese has been extensively studied and providing a review of this literature is out of scope of this dissertation (see Wang (2018) for a recent review). Instead, I draw a parallel between the analysis of sentential LE in Wang (2018) and the 'start' meaning of /kɨman/.

Wang (2018) proposes that the sentential LE is a focus marker, which is adjoined to the structure at a higher position, and 'expresses the assertion of the situation it takes by setting it in contrast with potential alternative situations' (p. 249). To paraphrase Wang (2018)'s description of the meaning of the sentential LE, it expresses the assertion of the situation it takes (i.e., 'drinking wine') by setting it in contrast with potential alternative situations (i.e., 'not drinking wine'). This contrast creates the meaning that the person 'has begun' drinking wine, and 'used to not' drink wine.

It is important to note that researchers of Mandarin Chinese have not fully reached consensus about how to analyze the sentential LE semantically and syntactically (Soh, 2009; Wang, 2018), and there are different views about whether there is a single lexical item that has different syntactic positions or there are two lexical items that happen to occupy different syntactic positions (see Wang (2018) for a review). Likewise, while it is possible that /kɨman/ has a single denotation of 'discontinue', as I adopt from Noh (1983), it is also possible that there are simply two lexical items with the same pronunciation. I leave for future studies to investigate the details of syntax and semantics of /kɨman/ in relations to the verbal/sentential LE in Mandarin Chinese.

Instead, the main thing to highlight is that the meaning of the sentential LE is similar to the meaning of kɨman$_{start}$ in that both seem to create a contrast with potential alternative situations, which may be described as the negation of the situation it takes. For this reason, I argue that when /kɨman/ occupies a high position, it introduces focus on the later element, which is typically the verb. Syntactically, both the sentential LE and kɨman$_{start}$ occupy a higher position in the syntactic structure, which is often associated with discourse-pragmatic functions (e.g., Cinque, 1999).

4.3.3 Clausal/Manner Adverb ambiguity

In addition to the verbal/sentential LE in Mandarin Chinese, I briefly review another case of ambiguity, which exhibits some similarities to /kɨman/, which is clausal/manner adverb ambiguity. I will show that /kɨman/ is one of many adverbs that have both clausal and manner readings, and such adverbs are attested both in Korean and in other languages. These adverbs show a similar linear adjacency constraint to /kɨman/'s, and similar prosodic phrasing patterns that /kɨman/ exhibits.

Crosslinguistically, adverbs like 'cleverly', 'stupidly', and 'clearly' cause syntactic ambiguity as they can either modify a verb phrase or an entire clause (e.g., Jackendoff, 1972; Ernst, 2002; Egedi, 2009; Kubota, 2015). Consider (10), which is borrowed from Example (1) on p. 1020 in Kubota (2015).

(10) a. John stupidly danced. (Ambiguous)
  (i) 'The manner in which John danced was stupid.'
  (ii) 'It was stupid of John to have danced.'
 b. Stupidly, John danced. (Unambiguous)
  (i) 'It was stupid of John to have danced.'

c. John danced stupidly. (Unambiguous)

(i) 'The manner in which John danced was stupid.'

In the sentence-final position, 'stupidly' only modifies the verb phrase 'danced' and it can only refer to the manner of dancing. This is referred to as the 'manner reading'. In the sentence-initial position, on the other hand, 'stupidly' is interpreted to modify the entire proposition of John's dancing. This is referred to as the 'clausal' or 'sentential' reading. In the preverbal position in (10a), both manner and clausal readings are available.

Similar adverbs in Korean also show the same clausal and manner ambiguity. As an example, I illustrate this with /hwaksilhi/ 'clearly' in (11). Note that not every Korean adverb is ambiguous between clausal and manner readings, just as not every adverb is amgbiguous in English (e.g., 'tightly' can only be a manner adverb in English (Ernst, 2002)).

(11) /hwaksil-hi po-as-ta/
clear-ly see-PAST-DECL

a. Clausal reading: 'Clearly, (I) saw (it).'

b. Manner reading: 'I saw (it) clearly.'

The two readings are subject to the same adjacency restriction reviewed earlier for /kɨman/ in that there must not be any word intervening between the adverb and the verb, if the adverb is to be interpreted with the manner reading. This is illustrated in (12).

(12) a. (i) /na-nɨn **hwaksil-hi** motu-ka ʧoaha-nɨn kɨrim-ɨl
I-TOP clear-ly everyone-NOM like-REL painting-ACC
po-as-ta/
see-PAST-DECL
(*manner, clausal) 'Clearly, I saw the painting that everyone likes.

(ii) /hatʃiman tʃal mot po-asɔ a.swip-ta/

But well NEG see-because regrettable-DECL
But it's a shame that I didn't see it well.'

b. (i) /na-nɨn motu-ka tʃoaha-nɨn kɨrim-ɨl **hwaksil-hi**

I-TOP clear-ly everyone-NOM like-REL painting-ACC

po-as-ta.

see-PAST-DECL
(manner/clausal, disambiguated prosodically) 'I clearly saw the painting that everyone loves'.

(ii) hatʃiman tʃal mot po-asɔ a.swip-ta/

But well NEG see-because regrettable-DECL
#/? 'But it's a shame that I didn't see it well'.

Both (12a) and (12b) have two sentences. The first one has the adverb /hwaksil-hi/ ('clearly') and the second one is a continuation from the first sentence: 'But it's a shame that I didn't see it well'. In (12a), when /hwaksil-hi/ is separated from the verb phrase, only the clausal reading is available, and the sentence means that it is without any doubt that I saw the painting. As 'clearly' does not modify the manner of my seeing, the second sentence can be a good continuation, meaning that though there is no doubt that I saw it, the manner of seeing was not good, perhaps because there were too many people trying to see it at the same time, and I did not have enough time to see it in detail.

On the contrary, in (12b), /hwaksilhi/ is adjacent to the verb phrase. Here, the continuation in (ii) becomes infelicitous if we interpret the sentence in (i) with the manner reading, because it is contradictory to 'see something clearly but not well'. The only felicitous reading would be to interpret the adverb with the clausal reading ('there is no doubt that I saw it'). In line with the /kɨman/ + Verb construction, I predict that the prosody will again play a role and the continuation in (ii) becomes less acceptable if /hwaksilhi/ and the verb is prosodically phrased together, though I leave for future work to test this prediction.

Example (11) illustrates three points. First, some adverbs in Korean can be ambiguous preverbally between clausal and manner readings. Second, when they are separated from the verb in the linear order, only the clausal reading becomes available. Third, when they are adjacent to the verb, the manner reading is available, and the clausal reading seems to be available only if there is a prosodic break between the adverb and the verb. All three of these characteristics are what we already saw for /kɨman/. This suggests that the polysemy of /kɨman/ may be a general characteristics of adverbs that have both clausal and manner readings.

The clausal/manner adverb ambiguity is extensively studied in the literature and there are several proposals available for how the ambiguity arises in the syntactic/semantic structure (e.g., Ernst, 2002; Kubota, 2015; Keizer, 2018). Providing a thorough review of this literature is out of scope of this dissertation (see Ernst (2020) for a recent review of this literature). Here, I will only discuss an aspect of one of these proposals that concerns the notion of 'comparison class' (Ernst, 2002; Kubota, 2015), as it is relevant for understanding the 'start' meaning of /kɨman/.

To derive the clausal/manner meaning of adverbs, Ernst (2002) and Kubota (2015) argue that the degree adverbs (stupidly, cleverly, rudely etc.), which describe a degree of an event on some scale (e.g., stupidity for stupidly), are evaluated with different 'comparison classes'. For instance, consider the example in (13) borrowed from Example (2.9) on p. 43 in Ernst (2002). The notes in parentheses are added to the original sentences to clarify how the adverb is interpreted in each sentence.

(13) a. Rudely, she left (clausal, *manner)

b. She left rudely (*clausal, manner)

Ernst (2002) argues that in (13a), 'she is judged rude because of the event of her leaving, as opposed to other things she could have done, most especially not leaving' (p. 57). On the other hand, in (13b), 'she is judged rude on the basis of something about her leaving', i.e. her manner of leaving. For instance, she 'left without saying good-bye' or 'by slamming the door' (p. 57). Interestingly, Ernst (2002) mentions that the 'other things she could have done' can be described as the negation of leving (i.e., 'not leaving').

Kubota (2015) adopts the idea of comparison class to analyze Japanese data similar to (10) and (13). Kubota uses two mechanisms to derive the difference in meaning. First, the adverb is adjoined to the sentence at a lower position when it has a manner reading, and at a higher position when it has a clausal reading. Second, when it has a clausal reading, a new vacuous event is introduced to serve as the non-restrictive comparison class. In Japanese, this has a morphological exponent /mo/. For example, a manner adverb for 'stupidly' is /orokani/ whereas the clausal verb of it is /orokani-mo/, as in (14), borrowed from Example (2-3) on p. 1021 in Kubota (2015).

(14) a. John-wa orokani odotta

John-TOP stupidly danced
'John danced stupidly' (*clausal, manner)

b. orokani John-wa odotta

stupidly John-TOP danced
'John danced stupidly' (*clausal, manner)

c. John-wa orokani-mo odotta

John-TOP stupidly danced
'John danced stupidly' (clausal, *manner)

d. orokani-mo John-wa odotta

stupidly John-TOP danced
'John danced stupidly' (clausal, *manner)

In the semantic derivation Kubota presents, when /orokani/ is evaluated, it is compared with other dancing events, and the particular dancing event mentioned in the sentence is judged to be more 'stupid' than the average degree of stupidity. On the other hand, when /orokani-mo/ is evaluated, it is compared with the events in a vacuous predicate. In this way, John's dancing event is judged to be more stupid than other things he could have done, e.g., 'John's moving' (p. 1032). While Kubota describes the set of events that acts as the comparison class as nonrestricted, it seems that it can also be described as the negation of the event it modifies, 'John's not dancing', in the same way that in (13), 'her leaving' is compared with 'her not leaving'.

Put this way, the clausal/manner meanings of adverbs are similar to start/stop meanings of /kɨman/ in two ways. First, in terms of syntactic position, both the clausal adverb and /kɨman/ with the 'start' meaning seem to be adjoined at a higher position, judging from the fact that words can intervene between the adverb and the verb. Second, for both the clausal reading of adverbs and /kɨman/ with the 'start' meaning, the evaluation of the adverb involves the negation of the event, as summarized in (15). As /kɨman/ with the 'start' meaning introduces a comparison between the verb happening and the verb not happening, I argue that the verb receives focus in the kɨman$_{start}$ + Verb construction.

(15) a. Rudely, she left.

Meaning of rudely: more rude compared to x.

Meaning of the sentence: Compared to her *not* leaving, her leaving was rude.

b. /kɨman minsu-nɨn ul-ɔs*-ta/

kɨman Minsu-TOP cry-PAST-DECL

'Minsu started crying'

Meaning of /kɨman/: discontinue x.

Meaning of the sentence: The state of Minsu's *not* crying is discontinued.

As noted earlier, working out the details of the semantic derivation of /kɨman/ is not the goal of this chapter, and I leave for future work to explore this idea. The main point to note is that /kɨman/ may not be so special in its meaning, and it exhibits similar syntactic and semantic properties to other adverbs that are ambiguous.

4.3.3.1 Role of prosody in adverb disambiguation

Another similarity between the clausal adverb and /kɨman/ with the 'start' meaning is that both are often pronounced with a prosodic boundary.

For instance, (16)–adopted from Example (4.121) on p. 139 in Potts (2005)–illustrates the role of prosodic breaks, transcribed in text with ','. Similar observations have been made in Selkirk (2005) with a minimal pair in (17) (her Examples (26) and (27) on p. 24).

(16) a. Willie luckily won the pool tournament (manner)

b. Willie, luckily, won the pool tournament (clausal)

(17) a. Mary remained true to herself.

b. Mary remained, true to herself.

In both examples, when there is no ',', which is acoustically implemented as an Intonational Phrase boundary (Selkirk, 2005), the adverbial ('luckily' or 'true to herself') modifies the verb, and when there is ',', it adds an additional meaning to the entire proposition, for instance the speaker's attitude towards the proposition (Potts, 2005).

Egedi (2009) reports a similar pattern in Hungarian. In postverbal position, adverbs like 'cleverly' are ambiguous between the clausal and manner reading.[2] The meaning of the sentence is disambiguated by the Intonational Phrase boundaries both before and after the adverb, as in (18b) (Example 38 and 40 on p. 113 in Egedi (2009)).

(18) a. Hugó nem válaszolt okosan a kérdésre
   Hugo not answered cleverly to the question
   'Hugo did not answer the question cleverly (*clausal, manner)'

  b. Hugó nem válaszolt, okosan, a kérdésre
   Hugo not answered, cleverly, to the question
   'Cleverly, Hugo not answered to the question (clausal, *manner)

The cross-linguistic tendency seems to be that adverbs that are interpreted with the clausal reading tend to be prosodified separately from the verb phrase (Keizer, 2018; Kojadinović, 2022).[3] The relations between the syntactic position and the prosodic phrasing will be discussed further for the /kɨman/ + Verb construction in §4.5.

4.3.4 Section summary

In this section, I have argued that kɨman$_{stop}$ and kɨman$_{start}$ have different positions in the syntactic structure. The former occupies a lower position and is syntactically phrased together with the verb that it modifies, while the latter occupies a higher position and seems not to be contained within the verbal projection. I have also argued that the /kɨman/ + Verb construction exhibits syntactic, semantic, and prosodic similarities with two cases of ambiguities that are extensively studied: the verbal/sentential LE in Mandarin Chinese (Soh, 2009; Wang, 2018), and the

---

[2]It is not clear from Egedi (2009) whether the syntactic structure is different for the two readings, when the adverb is postverbal.

[3]Keizer (2018) and Kojadinović (2022) argue that this is not always true, though there is a high correlation between the availability of clausal reading for adverbs and large prosodic junctures.

clausal/manner adverb in many languages including English (Ernst, 2002), Japanese (Kubota, 2015), and Hungarian (Egedi, 2009).

First, for all of these ambiguities, one of the readings is semantically interpreted closer to the verb phrase ('stop' meaning of /kɨman/, manner adverbs, and verbal LE), while the other reading is interpreted at a higher position in the structure, providing a discourse-pragmatic meaning ('start' meaning of /kɨman/, clausal adverbs, and sentential LE) (Kubota, 2015; Wang, 2018).

Second, when it occupies a higher position, the semantic interpretation of the word seems to involve a comparison or a contrast with a set of alternatives, which may be described as the negation of the predicate/situation it takes (Ernst, 2002; Kubota, 2015; Wang, 2018).

Third, when it occupies a higher position, it is also prosodically phrased differently than when it is interpreted together with the verb phrase ('start' meaning of /kɨman/ and clausal adverbs) (Potts, 2005; Selkirk, 2005; Egedi, 2009; Keizer, 2018; Kojadinović, 2022). These similarities suggest that there may exist a structural explanation for the polysemy of /kɨman/, just as such explanations have been proposed for the other ambiguities reviewed here.

4.4 Prosodic analysis

Next, we move onto the prosodic analysis of the /kɨman/ + Verb construction.

The observations in Gim (2004) mentioning the prosodic phrasing and the information structure (i.e., word under focus) of the /kɨman/ + Verb construction are summarized in Table 4.2. While Gim (2004) does not specifically mention that the verb is under focus when /kɨman/ means 'start', I have argued that it is reasonable to assume that the verb receives focus, given that semantically kɨman$_{start}$ introduces a comparison set, which may be described as the negation of the event that the verb denotes.

Meaning of /kɨman/	Prosodic phrasing	Word under focus
Stop	[/kɨman/ Verb]	/kɨman/
Start	[/kɨman/] [Verb]	Verb

Table 4.2: Summary of the observations on p. 62 in Gim (2004)

As a native speaker of Korean, I share the same intuitions as those reported in Gim (2004), and agree that the prosodic phrasing of /kɨman/ and the verb is different, depending on the meaning of /kɨman/. Based on these observations, I propose the prosodic structure of the /kɨman/ + Verb construction in §4.4.1.

In addition to these observations, I report new data on the prosodic phrasing of the /kɨman/ + Verb construction in §4.4.2. The key observation is that the prosodic phrasing between the verb and the following PW can vary, while the prosodic phrasing between /kɨman/ and the verb is consistent for each meaning of /kɨman/.

4.4.1 Proposal for prosodic structures of '/kɨman/ + Verb'

Gim (2004)'s first observation about the prosodic phrasing gives us a straightforward analysis in the K-ToBI model of intonation (Jun, 1993). We will again use the sentences in (2) as examples, which is repeated in (19).

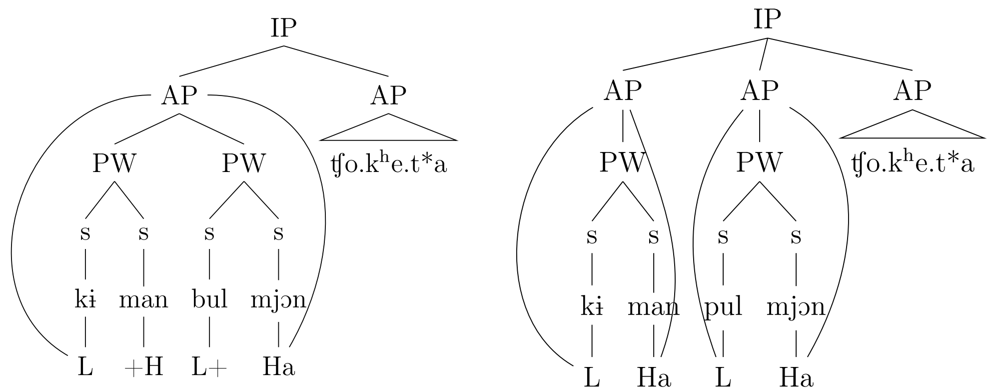
In Seoul Korean, the prosodic constituent above the prosodic word is AP. Based on the observations in Gim (2004), it is reasonable to assume that when /kɨman/ means 'stop', the following verb is phrased together with /kɨman/ in the same AP, as in Figure 4.3a. On the other hand, when /kɨman/ means 'start', the verb starts a new AP, as in Figure 4.3b.[4]

(19) /kɨman pul-mjɔn ʧoh-kes*-ta/
up to that degree/amount blow-COND be good-FUT-DECL

[4] /ʧo.k^hes*.ta/ is pronounced as [ʧo.k^he.t*a] due to neutralization of place of articulation in the coda position (/s/ → [t]) and post-obstruent tensification of /t/ (/t/ → [t*]). Finally, as the coda [s] (/t/) and the onset [t*] (/t/) has the same place of articulation, I have transcribed the sequence as simply [t*].

a. ‘Stop blowing’

b. ‘Start blowing’



(a) Stop blowing

(b) Start blowing

Figure 4.3: Prosodic trees for the sentences in (19)

	Stop V-ing (Fig. 4.3a)	Start V-ing (Fig. 4.3b)
Tones	L **+H** **L+** Ha	L **Ha** **L** Ha
Verb initial lenis voicing	[**b**]	[**p**]

Table 4.3: Comparison of acoustic implementation of the two prosodic structures in Figure 4.3

In the prosodic structure in Figure 4.3a, the verb is contained within the same AP as /kɨman/ making a long AP with four syllables (/kɨ.man.pul.mjɔn/). In this case, every syllable is assigned a tone. The first syllable /kɨ/ is assigned the AP-initial L tone because the first segment /k/ is not an aspirated or fortis obstruent, and therefore not assigned a high tone. The second syllable /man/ and /pul/ are assigned AP-medial tones +H and L+, respectively. The last syllable in this AP /mjɔn/ is assigned the AP-final Ha tone, which is a high tone that marks the end of the AP.

In the prosodic structure in Figure 4.3b, the verb starts a new AP, following the AP that contains /kɨman/. Again, the first syllable /kɨ/ is assigned the AP-initial L tone, but the second syllable /man/ is assigned the AP-final Ha tone, as it is the last syllable in the AP. The verb initial syllable /pul/ is assigned the AP-initial L tone, as it is the first syllable in the new AP, and it is not an aspirated or fortis obstruent. Finally, the verb-final syllable /mjɔn/ is assigned the AP-final Ha tone, as it is the last syllable in the AP. As a result, the two prosodic structures in Figure 4.3 are expected to have different tonal patterns, as summarized in Table 4.3.

The difference in the prosodic phrasing should also be marked via the segmental quality of the verb initial lenis stop /p/. When the verb is AP-medial, as in Figure 4.3a, it should be lenited (Jun, 1993, 1994, 1998). On the other hand, when the verb is AP-initial, as in Figure 4.3b, it should be unlenited. This is again summarized in Table 4.3. These predictions for the acoustic implementation of the /kɨman/ + Verb construction will be tested in Chapter 5.

In this chapter, we will assume that the prosodic phrasing in Figures 4.3a and 4.3b is correct, and discuss how these prosodic phrasing may be derived from the syntactic structure of the /kɨman/ + Verb construction in §4.5.

4.4.2 Variable prosodic phrasing of /kɨman/ + Verb construction

In Figure 4.3, I focused on the prosodic phrasing between /kɨman/ and the verb, and assumed that the last PW /ʧo.k^{h}e.t*a/ is prosodically phrased separately from the verb in its own AP. Building on Gim (2004)'s observations, I report that there is a variation in the prosodic phrasing of the /kɨman/ + Verb construction with respect to the verb /pul-mjɔn/ ('blow') and the following PW /ʧoh-kes*-ta/ ('be good-FUT-DECL').

Consider the sentences in (20) and (21), repeated from (2). The sentence in (20) is used to mean 'stop blowing', while the sentence in (21) is used to mean 'start blowing'.

When /kɨman/ means 'stop' as in (20), I prefer the prosodic phrasing that puts /kɨman/ and the verb together as in (20a) and (20b). On the other hand, when

/kɨman/ means 'start' as in (21), I prefer the prosodic phrasing that puts /kɨman/ and the verb separately as in (21a) and (21b). However, I find the prosodic phrasing between /pulmjɔn/ and /ʧoh-kes*-ta/ to be variable. That is, I find the prosodic phrasing in (20a) and (20b) equally acceptable, and the prosodic phrasing in (21a) and (21b) equally acceptable.

(20) /kɨman pul-mjɔn ʧoh-kes*-ta/

kɨman$_{stop}$ blow-COND be good-FUT-DECL

a. Acceptable: (kɨman pul-mjɔn)$_{AP}$(ʧoh-kes*-ta)$_{AP}$

b. Acceptable: (kɨman pul-mjɔn ʧoh-kes*-ta)$_{AP}$

c. *(kɨman)$_{AP}$(pul-mjɔn)$_{AP}$(ʧoh-kes*-ta)$_{AP}$

d. *(kɨman)$_{AP}$(pul-mjɔn ʧoh-kes*-ta)$_{AP}$

(21) /kɨman pul-mjɔn ʧoh-kes*-ta/

kɨman$_{start}$ blow-COND be good-FUT-DECL

a. Acceptable: (kɨman)$_{AP}$(pul-mjɔn)$_{AP}$(ʧoh-kes*-ta)$_{AP}$

b. Acceptable: (kɨman)$_{AP}$(pul-mjɔn ʧoh-kes*-ta)$_{AP}$

c. *(kɨman pul-mjɔn)$_{AP}$(ʧoh-kes*-ta)$_{AP}$

d. *(kɨman pul-mjɔn ʧoh-kes*-ta)$_{AP}$

Next, I will discuss the mapping relations between the proposed syntactic and prosodic structures of the /kɨman/ + Verb construction. I show that both the reason that the prosodic phrasing becomes different depending on the meaning of /kɨman/ as well as the reason that the prosodic phrasing between the verb and the following PW is variable, is a consequence of the syntactic and information structure of the /kɨman/ + Verb construction.

4.5 Modeling variable syntax-prosody mapping of /kɨman/ + Verb

In this section, I will present an analysis of how the syntactic structures of the /kɨman/ + Verb construction proposed in §4.3 can be mapped to the prosodic structures proposed in §4.4. In particular, I argue that Gim (2004)'s second claim about the difference in the information structure of the /kɨman/ + Verb construction depending on the meaning of /kɨman/ explains why there is a difference in the prosodic phrasing between /kɨman/ and the verb, and also why the prosodic phrasing between the verb and the following PW is variable.

This section proceeds as follows. In §4.5.1, I start with a brief review of the literature on focus-driven prosodic restructuring, which is well-documented in English (e.g., Ladd, 2008; Pierrehumbert, 1980) Japanese (e.g., Pierrehumbert & Beckman, 1988), Bengali (Hayes & Lahiri, 1991; Truckenbrodt, 2002), and Chichewâ (Truckenbrodt, 1999), among other languages. In particular, I introduce OT constraints used in the analysis of Truckenbrodt (1999) for Chichewâ and Bengali, which insert a prosodic boundary after the focused constituent, unlike Korean. In this proposal, violable OT constraints 'Align-XP', and 'Wrap-XP' are ranked to derive the syntactic influence on prosodic phrasing, and their interaction with the constraint 'Align-Foc L/R' accounts for the focus-driven prosodic restructuring (Truckenbrodt, 1999).

In §4.5.2, I bring the discussion back to Korean. I start by partially repeating the review from §3.2.3.1 and explain what the original proposal for the K-ToBI model (Jun, 1993) says about the syntactic influence on prosodic phrasing of PWs. In the original proposal, one of the situations that can force multiple PWs to be prosodically phrased in the same AP is when the first PW of that AP is focused. I present Kim (2015)'s analysis of focus-driven prosodic restructuring, which uses Truckenbrodt (1999)'s constraints to model the syntactic influence on prosodic phrasing in Korean. Kim (2015) has argued that while Align-Foc R can model focus-driven prosodic boundary insertion in Chicheŵa and Bengali, its mirror constraint Align-Foc L cannot model the focus-driven prosodic boundary deletion that has been empirically observed in Korean. As a solution, Kim (2015) proposes a new constraint 'Edgemost' (Right/Leftmost) to replace Align-Foc L/R, which motivates the focused PW to be right/leftmost in the AP ('phonological phrase' in Kim (2015)).

In §4.5.3, I argue that Edgemost constraint (Kim, 2015) fails to capture variation in focus-driven dephrasing. Instead, I propose that two constraints are responsible for variation in focus-driven dephrasing: a markedness constraint against having an AP right edge after a focused constituent (*FocAP-R) and an alignment constraint that prefers to have fewer AP right edges between the right edge of the focused constituent and an IP right edge (Align(Foc, R, IP, R, AP)). In addition, I argue that Align-Foc-L is still needed to account for focus-driven dephrasing in Korean.

In §4.5.4, I give an analysis of how the proposed prosodic phrasing of the /kɨman/ + Verb construction can be derived using these constraints.

§4.5.5 concludes the section with a summary.

4.5.1 OT accounts of focus-driven prosodic restructuring

The optimal prosodic phrasing of a syntactic structure can be modeled in Optimality Theory (OT) (Prince & Smolensky, 1993/2004) with a tableau. The syntactic structure of interest becomes the input of this tableau, and possible prosodic phrasings of PWs in the sentence become candidates. These candidates compete with each other based on a set of ranked constraints, which are used to evaluate the candidates and determine the optimal prosodic phrasing. The candidate that violates the lowest ranked constraint is chosen as the optimal prosodic phrasing of the input syntactic structure. There is a long list of literature on OT accounts of syntax-prosody mapping, and for a recent review, see Bellik et al. (2023). Here, I briefly introduce the key aspects of the analysis for focus-driven prosodic restructuring in Chicheŵa in Truckenbrodt (1999).

In general, a focused constituent requires a phonological phrase boundary at its right edge in Chicheŵa (Truckenbrodt, 1999). This pattern of inserting a prosodic phrase boundary is also seen in Bengali (Hayes & Lahiri, 1991; Truckenbrodt, 2002). Truckenbrodt (1999) models this phenomenon using the OT constraints in (22). Among them, the first constraint Align-XP-R/L is motivated in Selkirk (1986, 1996) to model the cross-linguistic tendency that edges of syntactic constituents and prosodic constituents are often aligned, which takes the format of Generalized Alignment constraints suggested in McCarthy & Prince (1993).

(22) Syntax-prosody mapping constraints (Selkirk, 1986, 1996; Truckenbrodt, 1999)

a. **Align-XP-R/L**: For each syntactic constituent XP, there is a prosodic constituent [here, a phonological phrase] such that the right/left edge of XP coincides with the right/left edge of a phonological phrase.

b. **Wrap-XP**: Each XP is contained in a phonological phrase.

c. **Align-Focus-R/L**: Each focused constituent is right/left-aligned with a phonological phrase boundary.

d. ***P-Phrase**: Assign a violation for every phonological phrase.

In Chicheŵa, the syntactic constituent [V NP PP] is prosodically phrased as (V NP PP)$_{pph}$ by default, where ()$_{pph}$ indicates the phonological phrase boundary, which is the prosodic constituent above the prosodic word in Chicheŵa (Truckenbrodt, 1999). V, NP, and PP stand for verb, noun phrase, and prepositional phrase, respectively. When the verb is under focus, the same syntactic constituent becomes prosodically phrased as (V)$_{pph}$(NP)$_{pph}$(PP)$_{pph}$ (Truckenbrodt, 1999). This pattern can be modeled by assuming that the constraints in (22) are ranked as in Table 4.4. While the original analysis considers recursive prosodic structure (e.g., ((V NP)$_{pph}$PP)$_{pph}$) as a possible candidate, I do not show those candidates for simplicity. In Chicheŵa, at least, there is no evidence that recursive structure is allowed, and this can be modeled by assuming that a constraint that disprefers recursive structure ('Nonrecursivity') is ranked highly and blocks such candidates (Truckenbrodt, 1999).

A. Default phrasing in Chicheŵa					
[V NP PP]$_{VP}$		AlignFoc-R	Wrap-XP	Align-XP-R	*P-Phrase
	(V)(NP)(PP)		VP!		***
	(V NP)(PP)		VP!		**
	(V)(NP PP)		VP!	NP!	**
→	(V NP PP)			NP	*
B. Focus-driven phrase boundary insertion in Chicheŵa					
[V$_{Foc}$ NP PP]$_{VP}$		AlignFoc-R	Wrap-XP	Align-XP-R	*P-Phrase
→	(V)(NP)(PP)		VP		***
	(V NP)(PP)	*!	VP		**
	(V)(NP PP)		VP	NP!	**
	(V NP PP)	*!		NP	*

Table 4.4: OT Tableaux showing the ranking of the constraints in (22). '()' mark phonological phrase boundaries. The candidate marked with '→' is the optimal candidate in each tableau. '!' marks the critical violation that causes a candidate to lose.

In the upper tableau (A) in Table 4.4, the input syntactic structure is [V NP PP], where none of the constituents are under focus. In this case, the optimal candidate is (V NP PP)$_{pph}$, which violates Align-XP-R by not having a phonological phrase boundary after the NP, but satisfies the higher ranked Wrap-XP constraint. Wrap-XP requires that every syntactic XP is included in a phonological phrase. Note that any other phrasing (e.g., (V NP)(PP)) can't be the optimal candidate, as they would always violate Wrap-XP for the VP, as there is no phonological phrase that contains the entire VP. Wrap-XP is not violated by the fact that other XPs (i.e., NP and PP) are included in a phonological phrase on their own. It is satisfied when the XP is wrapped by a phonological phrase, even when other XPs are also included in that phrase. Therefore, (V NP PP)$_{pph}$ satisfies Wrap-XP completely, because for each XP (NP, PP, and VP), there is a phonological phrase that contains it.

This default phrasing is overriden when the verb is under focus, as shown in the lower tableau (B) in Table 4.4. The candidate (V NP PP)$_{pph}$ which was optimal in the upper tableau is no longer optimal, as it violates the highest ranked Align-Foc-R constraint by not having a phonological phrase boundary after the focused verb. Candidates (V NP)(PP) and (V)(NP)(PP) equally violate Wrap-XP (because there is no phonological phrase that contains the VP), but only (V)(NP)(PP) satisfies Align-XP-R, and thus is the optimal candidate in this tableau.

To summarize, Wrap-XP is ranked above Align-XP-R to derive the default prosodic phrasing to include PWs in one phonological phrase. This is overriden by a higher ranked AlignFoc-R constraint, which forces a phonological phrase boundary after the focused constituent.

Truckenbrodt (1999) has argued that the constraint 'AlignFoc-L' is responsible for the opposite pattern of inserting a phrase boundary to the left of a focused constituent in Japanese (Pierrehumbert & Beckman, 1988)[5] and in Korean (Jun, 1993), though the relevant prosodic constituent would be 'Accentual Phrase' rather than 'phonological phrase' in Korean. However, Kim (2015) has shown that while AlignFoc-L captures the fact that a focused syntactic constituent initiates a new AP, it cannot model why there is dephrasing, which overrides the general tendency to phrase every PW in its own AP. Instead, Kim (2015) proposes a constraint 'Edgemost-L/R' (Prince & Smolensky, 1993/2004; Selkirk, 2011) is responsible for prosodic dephrasing in Korean. We now turn to the discussion of focus-driven prosodic restructuring in Korean.

[5]The relevant prosodic constituent is called 'intermediate phrase' in Pierrehumbert & Beckman (1988)

4.5.2 Focus-driven prosodic restructuring in Korean

In §3.2.3.1, I have reviewed how syntactic structure influences prosodic phrasing in the original K-ToBI model (Jun, 1993). I repeat the Accentual Phrase formation rules as in (23) (repeated from (1) in §3.2.3.1). While I will not repeat the discussion on these rules, as they have been explained with an example in §3.2.3.1, I will reformulate them in (24) and highlight some important implications of these rules.

(23) The Accentual Phrasing rules (Example (16) on p. 215 in Jun (1993))

a. Every prosodic word may be an Accentual Phrase.

b. A focused word must be the left-most word in an Accentual Phrase.

c. An Accentual Phrase can include any number of prosodic words as long as:

   (i) the last prosodic word is not the left element of a branching constituent.

   (ii) all the prosodic words are not focused.

(24) Reformulation of the Accentual Phrasing rules in (23)

a. (23a) says every PW 'may' be an AP. However, as mentioned in previous chapters, it is considered to be the 'default' that every PW becomes an AP (Kim, 2004; Jun & Kim, 2004) which is a stronger statement than (23a).

b. (23b) states that if there is a focused PW, it must start a new AP. This is by far the most straightforward and strongest statement among the rules in (23).

c. (23c-i) states that an AP can include multiple PWs if the last PW is not the left element of a branching constituent. This rule does not imply that if the last prosodic word is the right element of a branching constituent, it is prosodically phrased together with the preceding PW.

However, I showed in Chapter 3 that a PW following a CP is more likely to be prosodically phrased together with the last PW of that CP (see §3.5). Therefore, it suggests that the default tendency in (23a) may be overridden by the syntactic structure.

d. (23c-ii) does not state what happens when one of the PW is focused. Jun (1993) describes that all PWs after the focused PW within the same Intonational Phrase are prosodically phrased together with the focused PW (p. 214-215). When this happens, the PWs that follow the focused PW are said to be 'dephrased' (p. 186).

As the implications in (24) suggest, there seem to be two motivations for a PW to be prosodically phrased together with the preceding PW in the same AP, overriding the default tendency to start its own AP (23a). These motivations are stated in (25). In this section, I focus specifically on the first one, focus-driven dephrasing, and argue that this is responsible for the prosodic difference between the two /kɨman/ + Verb sentences. I briefly discuss the second motivation, syntactic dephrasing, in §4.6.2.

(25) Two motivations to dephrase a PW

a. Focus-driven dephrasing: A PW that follows a focused PW is prosodically phrased together with the focused PW in the same AP.

b. Syntactic dephrasing: A PW that belongs together with the preceding PW in a syntactic constituent can be prosodically phrased together with the preceding PW in the same AP. In particular, a PW that follows a CP tends to be prosodically phrased together with the last PW of that CP in the same AP.

The rules in (24) can be formalized by ranking the relevant OT constraints. First, (24a) states that every PW tends to be prosodically phrased in an AP. While a head-initial language may have a case where the right-edge of a word does not align with the right-edge of an XP (e.g., [V NP PP]$_{VP}$), a head-final language like Korean tends to have every word right-aligned with an XP boundary (e.g., [NP PP V]$_{VP}$). Assuming there is no recursive prosodic structure, as proposed in the original K-ToBI model (Jun, 1993), this may be modeled by ranking Align-XP-R above Wrap-XP. This derives the default prosodic phrasing that every PW is prosodically phrased in an AP, which contrasts with the default phrasing of Chicheŵa (Truckenbrodt, 1999). This is shown in Table 4.5, which presents the default phrasing of the sentence in (26). Note that Kim (2015) uses Lenis voicing to diagnose the location of phrase boundaries. When a PW-initial Lenis is AP-medial, it is realized as voiced. Following Kim (2015), I assume Wrap-XP does not evaluate whether there is an AP that wraps the entire sentence (S).

(26) mijɔŋ-ika pomi-eke kapaŋ-ɨl pilljɔʧu-ɔs*-ta

Miyoung-NOM Pomi-DAT Bag-ACC loan-PAST-DECL
'Miyoung loaned (her) bag to Pomi' (Example (35) in Kim (2015))

Default phrasing in Korean [[mijɔŋ-ika]$_{NP}$[[pomi-eke]$_{NP}$[kapaŋ-ɨl]$_{NP}$[pilljɔtʃuɔtta]$_{V}$]$_{VP}$]$_{S}$	Align XP-R	Wrap XP
→ a. (mijɔŋ-iga)(pomi-ege)(kabaŋ-ɨl)(pilljɔtʃuɔtta)		*
b. (mijɔŋ-iga)(pomi-ege)(kabaŋ-ɨl billjɔtʃuɔtta)	*	*
c. (mijɔŋ-iga)(pomi-ege gabaŋ-ɨl billjɔtʃuɔtta)	*!*	*
d. (mijɔŋ-iga bomi-ege gabaŋ-ɨl billjɔtʃuɔtta)	*!**	

Table 4.5: OT Tableau for the sentence in (26). '()' indicates AP boundaries. '→' marks the optimal candidate.

Second, the rule in (24b), which requires a focused PW to start a new AP, can be modeled by ranking Align-Foc-L above Wrap-XP. However, Kim (2015) argues that AlignFoc-L is not enough to derive the prosodic dephrasing. Kim (2015) explains that AlignFoc-L only ensures that there is an AP boundary at the left edge of the focused PW. It does not necessitate that PWs following the focused PW are prosodically phrased together with the focused PW ((24d)). This is shown in Table 4.6. The input sentence is the same as in (26), but now [Pomi-eke] is under focus. For this input sentence, the expected prosodic phrasing is such that there is an AP-boundary on the left edge of [Pomi-eke], but the following PWs are prosodically phrased together with this focused PW as in candidate (d) in Table 4.6. However, the phrasing in candidate (a), marked with '💣' cannot be made less optimal than the expected optimal phrasing in (d) (marked with '😟') because (a) satisfies Align-XP-R, while (d) does not. In fact, AlignFoc-L only serves to rule out (e), which is already not optimal as it violates the highly ranked Align-XP-R.

AlignFoc-L fails to capture focus-driven dephrasing in Korean [[M-ika]$_{NP}$[[P-eke]$_{Foc}$[kapaŋ-ɨl]$_{NP}$[pilljɔtʃuɔtta]$_{V}$]$_{VP}$]$_{S}$	Align Foc-L	Align XP-R	Wrap XP
💣 a. (mijɔŋ-iga)(pomi-ege)(kabaŋ-ɨl)(pilljɔtʃuɔtta)			VP
b. (mijɔŋ-iga)(pomi-ege)(kabaŋ-ɨl billjɔtʃuɔtta)		*	VP
c. (mijɔŋ-iga)(pomi-ege gabaŋ-ɨl)(pilljɔtʃuɔtta)		*!	VP
😟 d. (mijɔŋ-iga)(pomi-ege gabaŋ-ɨl billjɔtʃuɔtta)		*!*	VP
e. (mijɔŋ-iga bomi-ege gabaŋ-ɨl billjɔtʃuɔtta)	*!	***	

Table 4.6: OT Tableau for the sentence in (26), when [Pomi-eke] is under focus. '()' indicates AP boundaries. '😟' marks the expected optimal candidate which is not optimal given the ranking of the constraints in the tableau. '💣' marks the most acceptable candidate, though empirically unacceptable, given the ranking of constraints in the tableau.

To account for the post-focus dephrasing in (24d), Kim (2015) proposes a new constraint 'Edgemost' (Right/Leftmost). In this analysis, Rightmost replaces Align-Foc L, and motivates the AP containing the focused constituent to be rightmost in the IP. The Edgemost constraint is proposed in previous studies (e.g., Prince & Smolensky, 1993/2004; Selkirk, 2011) but it had not been discussed to motivate post-focus dephrasing. Kim (2015)'s definition of Edgemost is given in (27).

(27) Definition of Edgemost (Example (43) (Kim, 2015))

Edgemost: Focus prominence is Rightmost/Leftmost within a relevant prosodic domain.

In this definition, a prosodic constituent containing the syntactic constituent under focus is assumed to bear 'focus prominence'. Kim (2015) assumes that the prosodic constituent that bears focus prominence is an AP in Korean. The AP that bears focus prominence is then required by the constraint 'Rightmost' (the Rightmost variant of the Edgemost constraint) to be rightmost within the prosodic constituent above the AP, which is assumed to be IP in Kim (2015). Table 4.7 illustrates the effect of Rightmost. In both (d) and (e), the AP containing the focused constituent is also the rightmost AP in the IP, thus satisfying Rightmost. Candidate (d) is more optimal than (e) because it violates Align-XP-R one less time than (e). Note that in this tableau, Align-Foc-L would not change the winning candidate, whether ranked above or below Rightmost.

Rightmost (Kim, 2015) captures focus-driven dephrasing in Korean $[[\text{M-ika}]_{NP}[[\text{P-eke}]_{Foc}[\text{K-ɨl}]_{NP}[\text{pilljɔtʃuɔtta}]_{V}]_{VP}]_{S}$	Right most	Align XP-R	Wrap XP
a. [(mijɔŋ-iga)(pomi-ege)(kabaŋ-ɨl)(pilljɔtʃuɔtta)]	*!		VP
b. [(mijɔŋ-iga)(pomi-ege)(kabaŋ-ɨl billjɔtʃuɔtta)]	*!	*	VP
c. [(mijɔŋ-iga)(pomi-ege gabaŋ-ɨl)(pilljɔtʃuɔtta)]	*!	*	VP
→ d. [(mijɔŋ-iga)(pomi-ege gabaŋ-ɨl billjɔtʃuɔtta)]		**	VP
e. [(mijɔŋ-iga bomi-ege gabaŋ-ɨl billjɔtʃuɔtta)]		***!	

Table 4.7: OT Tableau for the sentence in (26), when [Pomi-eke] is under focus. '()' indicates AP boundaries and '[]' around each candidate indicates IP boundaries

Kim (2015) argues that using Rightmost to account for focus-driven dephrasing in Korean is supported by the fact that when an AP with a tonal melody 'LHLH' is focused, there is a larger initial pitch rise (initial LH), compared to a regular (non-focused) AP (Oh, 2008). As a result, the prosodic constituent containing the focused constituent (e.g., a focused AP) is acoustically distinct from other prosodic constituents of the same level within a larger prosodic constituent (e.g., other APs within the same IP). This acoustic distinction (e.g., an exaggerated pitch rise of a focused AP) that a prosodic constituent has compared to other prosodic constituents of the same level within a larger prosodic constituent, is what Kim (2015) and studies cited below refer to as 'prominence'.

It is a common claim discussed in the literature that a focused constituent is maximally prominent (acoustically) in a sentence (Truckenbrodt, 1995; Büring, 2009). It has also been a motivation for Jun (2006) to introduce a prosodic constituent above the AP, the intermediate phrase (ip), though Kim (2015) assumes a simpler prosodic hierarchy without the intermediate phrase.

However, the claim that the prosodic constituent containing the focused constituent is the most (acoustically) prominent within the same prosodic domain has been challenged empirically (Lee, 2017).

Lee & Xu (2010) has shown that prosodic (acoustic) means to mark (contrastive) focus is weak in Seoul Korean, compared to English. A constituent under contrastive focus was marked with a wider range of pitch in both Seoul Korean and English, but the difference in maximum pitch between the focused and unfocused conditions was two times bigger in English than Seoul Korean (2.52 semitones in American English; 1.18 semitones in Seoul Korean). Similarly, Downing (2008) showed that in three Bantu languages (Chicheŵa, Durban Zulu, and Chitumbuka), there is a mismatch between the most prominent prosodic constituent and the position of focus in these languages. Downing showed that regardless of the position of focus, the last prosodic word was consistently the most prominent prosodic constituent. In addition to these production studies, Lee (2017) has shown that Korean listeners struggle to find which word the speaker has intended to focus. Korean listeners were only able to identify the correct location of focus only 37% of the time, which was surprisingly low though it was higher than a random chance (10% in their experiment). These studies challenge the claim that a prosodic constituent containing the syntactic constituent under focus is maximally prominent in a sentence, at least in Seoul Korean (Lee, 2017).

Moreover, even if we accept the claim that a focused AP is the most prominent AP within an IP, the definition of Rightmost in (27) makes a strange prediction for the acoustic realization of focus. The definition of Rightmost only requires that an AP containing the focused constituent be the rightmost in an IP. This makes the prosodic phrasing in (e) that does not have the left edge at the beginning of the focused constituent [/Pomi-eke/] meet the requirement of Rightmost. However, Kim (2015) assumes that a focused AP is acoustically distinct from other APs, in that it has a large initial pitch rise. This means that the prosodic phrasing in (e) would have an expanded initial rise, but this rise would misalign with the actual constituent under focus, which is the second PW ([/Pomi-eke/]), not the first PW, in this AP.

Finally, as I show in the next section, Rightmost fails to capture variation in focus-driven prosodic phrasing.

4.5.3 Modeling variation in prosodic phrasing with weighted constraints

This section presents a proof-of-concept analysis for variation in prosodic phrasing in Korean. I use a Maximum Entropy (MaxEnt) grammar (Goldwater & Johnson, 2003) to model this variation. In a MaxEnt grammar, constraints are weighted as opposed to being strictly ranked (Smolensky & Legendre, 2006; Goldwater & Johnson, 2003). Constraint weights appear right below names of constraints in the tableaux in this section. Each candidate is assigned a weighted sum of violations of constraints. These weighted sums are called the Harmony ($\mathcal{H}$) and each candidate in a tableau is assigned to a probability (p), which is proportional to the exponential of its Harmony. I assume that these probability values represent how acceptable each candidate is as a prosodic phrasing, for a given input. If two candidates have the same Harmony, they will be assigned the same probability. Unlike strict ranking, a candidate that violates a constraint with a higher weight may end up with the same Harmony value (and therefore, the same probability) as a candidate that violates multiple constraints with lower weights. As a result, weighted constraints have been used to model variation in phonology (Pater, 2009). Note however, this approach may not be the only possible way to model variation in prosodic phrasing. For instance, Truckenbrodt (2002) has proposed using Output-Output correspondence to model variation in prosodic phrasing in Bengali. I leave for future work to investigate whether such an approach may be used to model variation in prosodic phrasing in Korean.

The goal of the MaxEnt analyses here is to capture two kinds of variation in prosodic phrasing in Korean: (i) variation in prosodic phrasing due to syntactic structure, and (ii) variation in prosodic phrasing after a focused constituent.

In Kim (2015)'s analysis, Align-XP-R is ranked above Wrap-XP to derive the default phrasing that every PW is prosodically phrased in an AP, as shown in Table 4.5. However, in the original proposal in Jun (1993), prosodic phrasing of PWs under broad focus is more variable. The rule in (24c) only states which prosodic phrasing is not allowed, i.e., an AP may not end with a PW that is the left element of a branching constituent. I show that this restriction can be modeled with the interaction of Align-XP-R and Wrap-XP.

While in a typical MaxEnt analysis, constraint weights are learned from the data, I will not discuss learning of weights in this chapter. Instead, I have chosen constraint weights carefully such that they reflect the relative ranking of the constraints in Kim (2015)'s analysis. In Kim (2015), Align-XP-R is ranked above Wrap-XP to motivate the default phrasing in (a). I have chosen a weight of 4 for Align-XP-R, and a weight of 2 for Wrap-XP, to reflect the strict ranking in Kim (2015). In addition, I assume there is a *AP constraint, with a weight of 1, which is a version of *P-Phrase in Truckenbrodt (1999).

I first present an analysis for the sentence in (26) without focus. The analysis is presented in Table 4.8. The MaxEnt tableau in Table 4.8 has more candidates than the OT tableau in Table 4.7 because it considers all possible ways that PWs may be phrased into APs, without breaking the Strict Layer assumption (Selkirk et al., 1982; Selkirk, 1986). Another difference from Kim (2015)'s analysis in Table 4.7 is the input syntactic structure. Rather than assuming there is a VP containing two NPs and the verb in a ternary branching structure, I assumed a right-branching structure with recursive VPs, as demonstrated in Figure 4.4. The effect of assuming a different syntactic structure is that Wrap-XP will incur a different number of violations, depending on which PWs are phrased into APs. The candidate in (b) [(mijɔŋ-iga)(pomi-ege gabaŋ-ɨl)(pilljɔtʃuɔtta)] incurs two violations of Wrap-XP for each VP in the input, while the candidate in (d) [(mijɔŋ-iga)(pomi-ege)(kabaŋ-ɨl billjɔtʃuɔtta)] incurs only one violation of Wrap-XP for the VP2 (the higher VP). Violation counts are indicated with negative integers in Table 4.8.

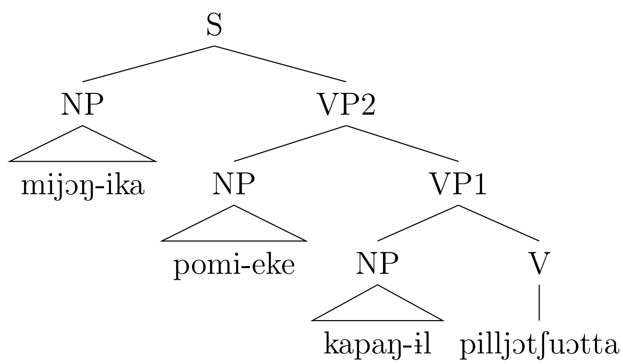


Figure 4.4: Syntactic structure for the sentence in (26) 'Miyoung loaned a bag (/kapaŋ-ɨl/) to Pomi'.

According to the rule in (24c) ('an AP may not end with a PW that is the left element of a branching constituent'), the candidates in (b), (e), and (f) are not allowed as the last PW in one of the APs of these candidates is the left element of a branching constituent (NP1 /gabaŋ-ɨl/ in (b) and (f); NP2 /bomi-eke/ in (e)). The possible phrasings according to this rule are in (a), (c), (d), and (g). The weights of constraints in Table 4.8 assign the largest probability to the default candidate in (a) (0.62), as desired. This candidate violates Wrap-XP twice but it completely satisfies AlignXP-R, and therefore gets the highest probability (which I assume to be acceptability of

A MaxEnt analysis for variable phrasing in Korean [[mijɔŋ-ika]$_{NP3}$[[pomi-eke]$_{NP2}$[[kapaŋ-ɨl]$_{NP1}$[pilljɔtʃuɔtta]$_V$]$_{VP1}$]$_{VP2}$]$_S$		Align XP R	Wrap XP	*AP		
		4	2	1	$\mathcal{H}$	p
→	a. [(m)(p)(k)(p)]		-2	-4	-8	0.62
	b. [(m)(p g)(p)]	-1	-2	-3	-11	0.03
→	c. [(m)(p g b)]	-2		-2	-10	0.08
→	d. [(m)(p)(k b)]	-1	-1	-3	-9	0.23
	e. [(m b)(k)(p)]	-1	-2	-3	-11	0.03
	f. [(m b g)(p)]	-2	-2	-2	-14	0.00
(?)	g. [(m b g b)]	-3		-1	-13	0.00

Table 4.8: MaxEnt Tableau for the sentence in (26). '()' indicates AP boundaries and '[]' around each candidate indicates IP boundaries. Candidates only show the first letter of each PW. When a PW-initial lenis stop is AP-medial, it is voiced. '→' marks the expected acceptable candidate(s).

prosodic phrasing given the input). Next, the candidate in (d) gets a probability of 0.23, as it violates AlignXP-R and Wrap-XP once each. This is considered to be better accepted by the grammar than the candidate in (c) which violates AlignXP-R twice but not Wrap-XP, as AlignXP-R has a higher weight than Wrap-XP. Though *AP is not the main component of this analysis, it also plays a minor role as it is what makes the candidate in (c) more optimal than (b). Without this constraint, both candidates would get equal Harmony values (b: 4 * (-1) + 2 * (-2) = -8 and c: 4 * (-2) = -8).

The MaxEnt analysis in Table 4.8 correctly assigns high probabilities to three phrasings (a), (c), and (d), which are predicted to be acceptable by the rule in (24c) (Jun, 1993). Among them, it also assigns the highest probability to the default prosodic phrasing in (a) (Jun, 1993; Kim, 2004; Jun & Kim, 2004). The analysis also predicts specifically that the phrasing in (d) is more acceptable than the phrasing in (c), while the rule in (24c) only predicts that both are acceptable. I leave this issue for future research. In addition, the analysis predicts that the phrasing in (g) is not acceptable, contra the prediction in Jun (1993) that the phrasing in (g) is acceptable (hence the '?' in Table 4.8). While I suspect that the phrasing in (g) is only allowed when the sentence is spoken in fast speech, I leave this issue for future research as well.

Next, I move onto the sentence in (26), with [pomi-eke] under focus.

Regarding focus-driven prosodic dephrasing, Cho (1990) has proposed that only the PW that is immediately following the focused PW is prosodically phrased together with the focused PW in the same AP. Jun (1993) has challenged this generalization by showing that her consultants consistently produced speech where more than one PW following the focused PW was prosodically phrased together with the focused PW (e.g., Example (4) on p. 187 in (Jun, 1993)). I propose that both generalizations may be correct if we accept the possibility that prosodic phrasing can vary given an input syntactic and information structure. The rule in (24d) may then be reformulated as one of the two following statements as in (28). These two reformulations are formalized as violable constraints as in (29).

(28) Reformulation of focus-driven dephrasing rule in (24d)

a. It is not desirable that the PW immediately following the focused constituent starts a new AP.

b. It is desirable to phrase as many PWs as possible together with the focused PW in the same AP, within an Intonational Phrase.

(29) Constraints motivated from (28)

a. *Foc, R, AP, R (*FocAP-R): Assign a violation for every focused constituent whose right edge aligns with an AP right edge.

b. Align(Foc, R, IP, R, AP) (AlignFocIP-R): Align the right edge of the focused constituent with the right edge of the IP that contains the focused constituent. Assign a violation for every intervening AP.[6]

The first reformulation in (28) is formalized as a markedness constraint and the second reformulation in (28) is formalized as a generalized alignment constraint (McCarthy & Prince, 1993) in (29).

Note that the first reformulation is more general as it does not distinguish between a phrasing where all PWs following the focused PW are prosodically phrased together (e.g., $(\mathrm{PW}_{Foc}$ PW PW $\mathrm{PW})_{AP}$) and a phrasing where only the immediately following PW is prosodically phrased together with the focused PW (e.g., $(\mathrm{PW}_{Foc}\ \mathrm{PW})_{AP}$ $(\mathrm{PW})_{AP}$ $(\mathrm{PW})_{AP}$). Both of these phrasings will satisfy *FocAP-R. On the other hand, the second reformulation will prefer the former phrasing over the latter, as there are fewer number of AP right edges intervening between the focused constituent and the right edge of the IP, when there is just one AP inside the IP. I argue that both constraints are needed to capture the expected variation in prosodic phrasing after a focused constituent. In addition, I argue that AlignFoc-L cannot be left out of the constraint set, contra Kim (2015).

Kim (2015) has proposed that Rightmost is ranked above Align-XP-R to derive the focus-driven dephrasing observed in Korean. Similarly, I have chosen a weight of 5 for Align-Foc-L, *FocAP-R, and AlignFocIP-R, which is higher than the weight of 4 for Align-XP-R. I will show how Align-Foc-L and *FocAP-R are mainly responsible for deriving the prosodic dephrasing. I will also demonstrate how AlignFocIP-R is

[6] I followed McCarthy & Prince (1993) and Hyde (2012) and in using the notation 'Align(ACat1, Edge1, ACat2, Edge2, SCat)'. This particular definition appears in Hyde (2012): 'the Edge1 of every ACat1 coincides with the Edge2 of some ACat2. Assess a violation mark for every SCat that intervenes between edges that fail to coincide' (p. 791).

responsible for making the candidate in (c) the only acceptable candidate, which Jun (1993) has argued to be the case.

We use the same sentence in (26) as in Table 4.8, but now with the focused PW [pomi-eke]. The MaxEnt tableau in Table 4.9 presents the analysis. In this tableau, the predicted two possible phrasings are in (b) and (c). If we follow Cho (1990)'s generalization that only the immediately following PW is dephrased, the acceptable phrasing is (b). In contrast, if we follow Jun (1993)'s generalization that all PWs following the focused PW within the IP are dephrased, the acceptable phrasing is (c). If we allow for both these phrasings to be acceptable, which is what I argue here, we expect both (b) and (c) to get some probabilities in the MaxEnt tableau.

A MaxEnt analysis for focus-driven dephrasing in Korean [[mijɔŋ-ika]$_{NP}$[[**pomi-eke**]$_{NPFoc}$[[kapaŋ-ɨl]$_{NP}$[pilljɔtʃuɔtta]$_{V}$]$_{VP1}$]$_{VP2}$]$_{S}$								
	Align Foc L 5	*Foc AP R 5	Align FocIP R 5	Align XP R 4	Wrap XP 2	*AP 1	$\mathcal{H}$	p
a. [(m)(**p**)(k)(p)]		-1	-2		-2	-4	-23	0.00
b. [(m)(**p** g)(p)]			-1	-1	-2	-3	-16	0.00
→ c. [(m)(**p** g b)]				-2		-2	-10	1.00
d. [(m)(**p**)(k b)]		-1	-1	-1	-1	-3	-19	0.00
e. [(m **b**)(k)(p)]	-1	-1	-2	-1	-2	-3	-31	0.00
f. [(m **b** g)(p)]	-1		-1	-2	-2	-2	-24	0.00
g. [(m **b** g b)]	-1			-3		-1	-18	0.00

Table 4.9: MaxEnt Tableau for the sentence in (26), when [pomi-eke] is under focus. '()' indicates AP boundaries and '[]' around each candidate indicates IP boundaries. Candidates only show the first letter of each PW. When a PW-initial lenis stop is AP-medial, it is voiced. '→' marks the expected acceptable candidate(s).

We first consider Jun (1993)'s generalization in Table 4.9. AlignFoc-L assigns a violation for (e), (f), and (g) as these candidates are missing an AP boundary at the left edge of the focused constituent [pomi-eke]. *FocAP-R assigns a violation for (a), (d), and (e) as these candidates have an AP boundary at the right edge of the focused constituent [pomi-eke]. The only two candidates that do not incur any violations from these constraints are (b) and (c), which are the two expected acceptable candidates.

As this is the first tableau where AlignFocIP-R is introduced, it is worth explaining how it assigns violations to the candidates in this tableau. AlignFocIP-R assigns two violations for (e), as there are two APs ((kapaŋ-ɨl)$_{AP}$, (pilljɔtʃuɔtta)$_{AP}$) between the right edge of the focused constituent [pomi-eke] and the right edge of the IP (the end of the sentence). It assigns one violation for (f), as there is one AP ((pilljɔtʃuɔtta)$_{AP}$) between the right edge of the focused constituent and the right edge of the IP. I assume that the right boundary of the AP containing the focused constituent ((mijɔŋ-iga bomi-ege gapaŋ-ɨl)$_{AP}$) does not incur violation for AlignFocIP-R, as the right edge of the focused element is contained within this AP, and therefore this AP as a whole may not be said to intervene between the right edge of the focused constituent and the right edge of the IP. However, the analysis is essentially the same if we assume that AlignFocIP-R counts every 'AP right edge', rather than every AP, intervening between the right edge of the focused constituent and the right edge of the IP. It would simply just increase the number of violations for AlignFocIP-R by one. In Table 4.9, AlignFocIP-R has a weight of 5, which makes the Harmony of (b) deviate from that of (c). As a result, the only acceptable candidate in Table 4.9 is (c), as Jun (1993)'s focus-driven dephrasing rule predicts ('all PWs after the focused PW becomes dephrased').

Now consider the MaxEnt tableau in Table 4.10, where AlignFocIP-R has a weight of 1. Decreasing the weight on AlignFocIP-R has the effect of increasing the probability on (b). As a result, (b) gets a probability of 0.12. I argue that individual speakers may have different weight on AlignFocIP-R, which may explain why some speakers consistently produce the phrasing in (c) where all PWs are dephrased after the focused PW (Jun, 1993), while some speakers produce the phrasing in (b) where only the immediately following PW is dephrased after the focused PW (Cho, 1990).

A lower weight on AlignFocIP-R increases probability of (b) [[mijɔŋ-ika]$_{NP}$[[**pomi-eke**]$_{NPFoc}$[[kapaŋ-ɨl]$_{NP}$[pilljɔtʃuɔtta]$_{V}$]$_{VP1}$]$_{VP2}$]$_{S}$								
	Align Foc L 5	*Foc AP R 5	Align FocIP R 1	Align XP R 4	Wrap XP 2	*AP 1	$\mathcal{H}$	p
a. [(m)(**p**)(k)(p)]		-1	-2		-2	-4	-15	0.01
→ b. [(m)(**p** g)(p)]			-1	-1	-2	-3	-12	0.12
→ c. [(m)(**p** g b)]				-2		-2	-10	0.87
d. [(m)(**p**)(k b)]		-1	-1	-1	-1	-3	-15	0.01
e. [(m **b**)(k)(p)]	-1	-1	-2	-1	-2	-3	-23	0.00
f. [(m **b** g)(p)]	-1		-1	-2	-2	-2	-20	0.00
g. [(m **b** g b)]	-1			-3		-1	-18	0.00

Table 4.10: MaxEnt Tableau for the sentence in (26), when [pomi-eke] is under focus. '()' indicates AP boundaries and '[]' around each candidate indicates IP boundaries. Candidates only show the first letter of each PW. When a PW-initial lenis stop is AP-medial, it is voiced. '→' marks the expected acceptable candidate(s).

Interestingly, the current analysis also predicts that there is an upper boundary on the probability of (b), assuming we do not assign negative weights on constraints (Pater, 2009). When we decrease the weight of AlignFocIP-R further, the probability of (b) will continue to increase, but it cannot go above 0.26, which is the probability of (b) when AlignFocIP-R has a weight of 0. I leave empirically testing this prediction for future research.

If AlignFocIP-R may have a weight of 0, is it also possible for *Foc-R to have a weight of 0? If *Foc-R has a weight of 0 and AlignFocIP-R has a high weight as in Table 4.9, it also results in (c) being the only acceptable candidate. However, when AlignFocIP-R has a lower weight, while holding *Foc-R at 0, the probability of the acceptable candidate (b) [(m)(p g)(p)] cannot ever be higher than a nonacceptable candidate (d) [(m)(p)(k b)] because (d) incurs fewer number of Wrap-XP violations than (b) does. As a result, I argue that *Foc-R is necessary in order to capture both generalizations in Cho (1990) and Jun (1993). As I have shown in Table 4.9 and

Table 4.10, AlignFocIP-R is then responsible for how acceptable (b) [(m)(p g)(p)] becomes.

In these two tableaux, another thing to note is that AlignFoc-L seemingly plays little role in determining the acceptable candidate(s). Consider the three candidates that violate this constraint, (e), (f) and (g), in Table 4.10. Candidates (e) and (f) already have a high enough Harmony if we imagine there is a zero weight on AlignFoc-L (-18 for (e) and -15 for (f)). Candidate (g) would have a harmony of -13 if there is a zero weight on AlignFoc-L and it would make it closer to (b) which has a harmony of -12. As a result, it would have a non-zero probability on (g) (0.04 to be exact). Still, the probability on (g) would not go beyond the probability on the expected acceptable candidates (b) and (c).

However, when focus-driven dephrasing misaligns with the syntactic constituent that satisfies Wrap-XP, AlignFoc-L plays a crucial role in ruling out the candidates that satisfy Wrap-XP. For an example, consider the sentence in (30), which has a (partial) left branching syntactic structure (see Figure 4.5).[7]

(30) PRO musɔun pɔmkoɾe-e kapaŋ-ɨl po-as*-ta
PRO scary orca-GEN bag-ACC see-PAST-Decl.
PRO saw scary orca's bag.

[7]Since the subject of this sentence (PRO) is not spoken, I treat this as a left branching structure.

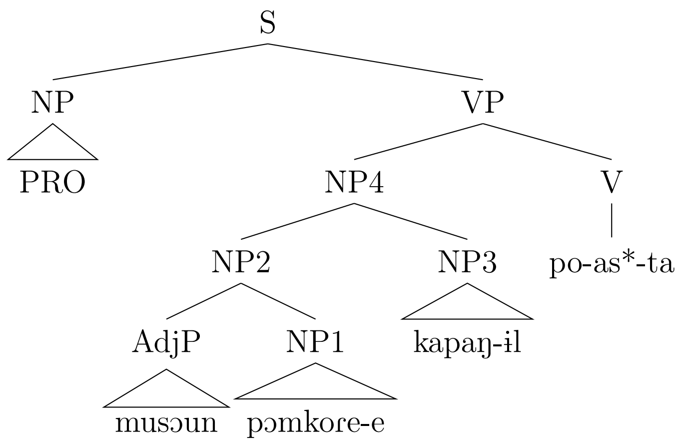


Figure 4.5: Syntactic structures for the sentence in (30)

AlignFoc-L needs to have a high weight for a left branching structure
[[[[[musɔun]$_{AdjP}$[**pɔmkoɾe-e**]$_{NPFoc}$][kapaŋ-ɨl]$_{NP}$][[po-as*-ta]$_{V}$]$_{VP}$]]$_{S}$

		Align Foc L 0	*Foc AP R 5	Align FocIP R 5	Align XP R 4	Wrap XP 2	*AP 1	$\mathcal{H}$	p
	a. [(m)(**p**)(k)(p)]		-1	-2		-2	-4	-23	0.00
→	b. [(m)(**p** g)(p)]			-1	-1	-2	-3	-16	0.03
→	c. [(m)(**p** g b)]				-2	-2	-2	-14	0.24
	d. [(m)(**p**)(k b)]		-1	-1	-1	-2	-3	-21	0.00
	e. [(m **b**)(k)(p)]	-1	-1	-2	-1	-1	-3	-24	0.00
	f. [(m **b** g)(p)]	-1		-1	-2		-2	-15	0.09
*	g. [(m **b** g b)]	-1			-3		-1	-13	0.64

Table 4.11: MaxEnt Tableau for the sentence in (30), when [/pɔmkoɾe-e/] is under focus. '()' indicates AP boundaries and '[]' around each candidate indicates IP boundaries. Candidates only show the first letter of each PW. When a PW-initial lenis stop is AP-medial, it is voiced. '*' marks the most acceptable candidate, though empirically unacceptable, given the weights of constraints in the tableau. '→' marks the expected acceptable candidate(s).

The MaxEnt Tableau for this sentence is in Table 4.11. Note that everything except for the violations of Wrap-XP should be the same between the two MaxEnt tableaux in Table 4.10 and Table 4.11. As Table 4.11 shows, if AlignFoc-L has a zero weight, the candidate (g) gets the highest probability because it satisfies both *FocAP-R and AlignFocIP-R, just as the expected acceptable candidate (c) does, while satisfying Wrap-XP completely. In Kim (2015)'s analysis, the syntactic structure of the input sentence is never a problem since constraints are strictly ranked as opposed to being weighted, and the analysis does not intend to model variation. In Kim (2015), both (g) and (c) would satisfy Rightmost, but (g) violates AlignXP-R more than the optimal candidate (c) does, and therefore gets ruled out. Once we assume constraints are weighted, we need AlignFoc-L back to rule out (g), i.e., a nonacceptable candidate that satisfies Wrap-XP. The correct analysis with a high weight on Align-Foc-L (5) is presented in Table 4.12.

Variable phrasing for a left branching structure as in (30) [[[[[musɔun]$_{AdjP}$[**pɔmkoɾe-e**]$_{NPFoc}$][kapaŋ-ɨl]$_{NP}$][[po-as*-ta]$_{V}$]$_{VP}$]]$_{S}$								
	Align Foc L 5	*Foc AP R 5	Align FocIP R 5	Align XP R 4	Wrap XP 2	*AP 1	$\mathcal{H}$	p
a. [(m)(**p**)(k)(p)]		-1	-2		-2	-4	-23	0.00
→ b. [(m)(**p** g)(p)]			-1	-1	-2	-3	-16	0.12
→ c. [(m)(**p** g b)]				-2	-2	-2	-14	0.86
d. [(m)(**p**)(k b)]		-1	-1	-1	-2	-3	-21	0.00
e. [(m **b**)(k)(p)]	-1	-1	-2	-1	-1	-3	-29	0.00
f. [(m **b** g)(p)]	-1		-1	-2		-2	-20	0.00
g. [(m **b** g b)]	-1			-3		-1	-18	0.02

Table 4.12: MaxEnt Tableau for the sentence in (30), when [/pɔmkoɾe-e/] is under focus. '()' indicates AP boundaries and '[]' around each candidate indicates IP boundaries. Candidates only show the first letter of each PW. When a PW-initial lenis stop is AP-medial, it is voiced. '→' marks the expected acceptable candidate(s).

One prediction that the current MaxEnt analysis makes is that the relative acceptability of (b) may be different depending on the syntactic structure of the input sentence (right-branching in Figure 4.4 and left-branching in Figure 4.5). Recall that in Table 4.8, when AlignFocIP-R has a weight of 5, the probability of (b) becomes vanishingly small. However, in Table 4.12, when AlignFocIP-R has a weight of 5, the probability of (b) is still 0.12. To make the probability on (b) become 0, AlignFocIP-R needs to have a weight larger than 8. This makes an interesting hypothesis that can be tested empirically. If we assume individuals differ in the weight of AlignFocIP-R, we predict that there may be a speaker who has a weight of 5 for AlignFocIP-R. We expect this person to never accept the phrasing in (b) for a right-branching structure ((b) would get a probability of 0 in this person's grammar) but finds the phrasing in (b) acceptable for a left-branching structure. Again, I leave empirically testing this prediction for future research.

To summarize the section thus far, I have presented a MaxEnt analysis for the variable phrasing in Korean. This analysis extends the analysis of Kim (2015) in two aspects. First, besides modeling the default phrasing that ever PW becomes an AP, it also assigns probabilities to other phrasings that are allowed by the rules in (24) (Jun, 1993). Second, in modeling the focus-driven dephrasing, it captures how a prosodic phrasing that only dephrases the immediately following PW after the focused PW may be acceptable, in addition to the phrasing that phrase all PWs after the focused PW. With this set up, I now return to the analysis of the prosodic phrasing of the /k1man/ + Verb construction.

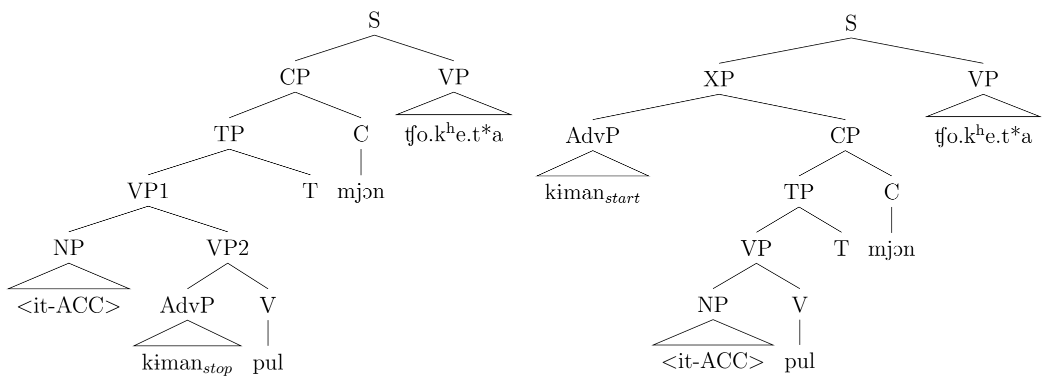
4.5.4 Analysis of prosodic phrasing of the /k1man/ + Verb construction

I start by presenting syntactic structures of the sentences in (19) (the 'stop/start blowing' sentences), for which I've proposed the prosodic structures in §4.4. These syntactic structures will be the input for the tableaux. The prosodic parse candidates are introduced next, followed by the MaxEnt analysis.

4.5.4.1 Input: Syntactic structures

I repeat the example sentences in (31) and present the syntactic structures of the two meanings of this sentence in Figure 4.6. The syntactic structures are meant to be a schematization of the difference in the syntactic structures of the two sentences, and are subject to further syntactic investigations. For instance, I do not propose that kɨman$_{start}$ is necessarily adjoined at a higher position than CP, but that it is at least not in the same verbal projection with the verb. I use a transitive verb (‘to blow’) in the examples but intransitive verbs would have similar syntactic structures, except that they would not have their object NP in the verbal projection.

(31) /kɨman pul-mjɔn ʧoh-kes*-ta/
up to that degree/amount blow-COND be good-FUT-DECL

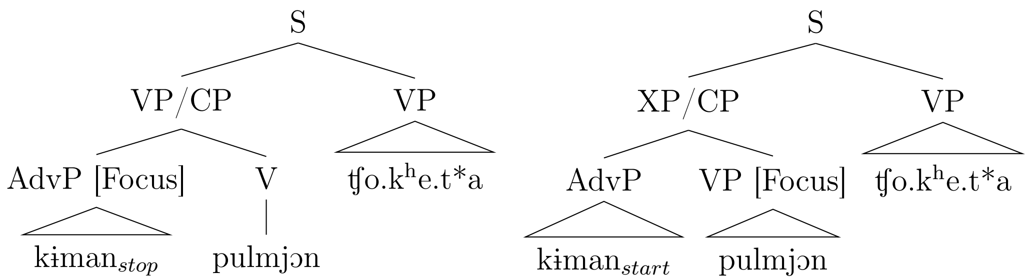


(a) ‘It would be good if (you) could stop blowing (it).’

(b) ‘It would be good if (you) could start blowing (it).’

Figure 4.6: Syntactic trees for the two meanings of the sentence in (31)

In the syntactic structure, I follow previous studies in that the verbal suffixes (such as the conditional suffix /mjɔn/) are projected as heads in the syntactic structure (e.g., Yoon, 1993; Chung, 2011). The consensus in the Korean literature seems to be that these suffixes are not just phrased together with their host (e.g., verb) in a larger prosodic constituent (AP), but that they form a single PW with their host (e.g., Yoon, 1993; Chung, 2011). This is in line with the K-ToBI model that assume verbal suffixes and case-marking suffixes are assumed to form a single PW with their host in Jun (2000). I follow these studies and assume that the verbal suffix is syntactically generated as a functional head of a higher phrase, but it is prosodified together with a host in the lower position. I also flatten the syntactic structure in Figure 4.6 assuming an empty projection like TP does not affect the prosodic phrasing, following Bennett et al. (2016). The resulting simplified syntactic structures are in Figure 4.7.



(a) 'It would be good if (you) could stop blowing (it).'

(b) 'It would be good if (you) could start blowing (it).'

Figure 4.7: Simplification of the syntactic trees in Figure 4.6

I claim that the constituent containing kɨman$_{stop}$ and the constituent containing the verb following kɨman$_{start}$ are under focus. The syntactic structure of these two sentences in Figure 4.7 do not differ in terms of where XP boundaries are located. The only difference is that the verb after kɨman$_{stop}$ is not in a VP on its own, while the verb after kɨman$_{start}$ is. However, in both syntactic structures there is an XP boundary after the verb. Consequently, there is no motivation to posit a difference in prosodic phrasing based on the proposed syntactic structure alone, if we do not assume different constituents are under focus in these two sentences.

Note that this is under the assumption that none of the candidates in the tableau have recursive prosodic structures, following the proposal that Korean does not allow recursive prosodic structures (Jun, 1993; Kim, 2015). I discuss an alternative analysis that allows recursive prosodic structures in subsection 4.6.2.

4.5.4.2 Candidates

In the input structures in Figure 4.7, there are three PWs to be prosodically phrased in AP(s). Consequently, there are four candidates in the tableaux in (32). These are all possible phrasings of the three PWs into APs, without breaking the Strict Layer assumptions (Selkirk, 1986).

(32) a. (kɨman)(pulmjɔn)(ʧo.k^{h}e.t*a)

b. (kɨman pulmjɔn)(ʧo.k^he.t*a)

c. (kɨman)(pulmjɔn ʧo.k^he.t*a)

d. (kɨman pulmjɔn ʧo.k^he.t*a)

As a reminder, I repeat the proposed difference in prosodic phrasing in (33) and (34) (= (20) and (21)).

(33) Stop Verb-ing in Figure 4.7a

a. *[kɨman][pul-mjɔn][ʧoh-kes*-ta]

b. *[kɨman][pul-mjɔn ʧoh-kes*-ta]

c. Acceptable: [kɨman pul-mjɔn][ʧoh-kes*-ta]

d. Acceptable: [kɨman pul-mjɔn ʧoh-kes*-ta]

(34) Start Verb-ing in Figure 4.7b

a. Acceptable: [kɨman][pul-mjɔn][ʧoh-kes*-ta]

b. Acceptable: [kɨman][pul-mjɔn ʧoh-kes*-ta]

c. *[kɨman pul-mjɔn][ʧoh-kes*-ta]

d. *[kɨman pul-mjɔn ʧoh-kes*-ta]

4.5.4.3 MaxEnt tableaux

I now present the MaxEnt tableaux for the two sentences in Figure 4.7. First, I assume an arbitrarily high weight on AlignFocIP-R (10) which makes only one candidate in each tableau acceptable. Table 4.13 presents the analysis.

The upper tableau is for the Stop Verb-ing sentence in Figure 4.7a and the lower tableau is for the Start Verb-ing sentence in Figure 4.7b. In the upper tableau, /kɨman/ is under focus, causing all following PWs to be dephrased (candidate (d)). In the lower tableau, the verb after /kɨman/ is under focus, and causes the following PW to be dephrased (candidate (b)). This difference in the location of focus leads to the important difference that the verb after /kɨman/ starts a new AP in the Start

	Align Foc L	*Foc AP R	Align FocIP R	Align XP R	Wrap XP	*AP		
A. Prosodic phrasing for Stop Verb-ing sentence in Figure 4.7a [[[[**kɨman**$_{Stop}$]$_{Foc}$pul-mjɔn]$_{VP}$]$_{CP}$[tʃok^{h}et*a]$_{VP}$]$_{S}$								
	5	5	5	4	2	1	$\mathcal{H}$	p
a. [(**k**)(p)(tʃ)]		-1	-2		-1	-3	-30	0.00
b. [(**k**)(p dʒ)]		-1	-1	-1	-1	-2	-23	0.00
c. [(**k** b)(tʃ)]			-1	-1		-2	-16	0.00
→ d. [(**k** b dʒ)]				-2		-1	-9	1.00
B. Prosodic phrasing for Start Verb-ing sentence in Figure 4.7b [[[kɨman$_{Start}$]$_{AdvP}$[**pul-mjɔn**]$_{Foc}$]$_{CP}$[tʃok^{h}et*a]$_{VP}$]$_{S}$								
a. [(k)(**p**)(tʃ)]		-1	-1		-1	-3	-20	0.00
→ b. [(k)(**p** dʒ)]				-1	-1	-2	-8	1.00
c. [(k **b**)(tʃ)]	-1	-1	-1	-1		-2	-26	0.00
d. [(k **b** dʒ)]	-1			-2		-1	-14	0.00

Table 4.13: MaxEnt Tableau for the sentences in Figure 4.7

Verb-ing sentence, while it does not in the Stop Verb-ing sentence. If we do not assume this difference in the input structures, we would predict exactly the same prosodic phrasing for the two sentences because in these two tableaux, the violations for AlignXP-R, Wrap-XP, and *AP are identical.

Now, consider the case where AlignFocIP-R has a weight of 0, which is presented in Table 4.14. Again, the upper tableau is for the Stop Verb-ing sentence in Figure 4.7a and the lower tableau is for the Start Verb-ing sentence in Figure 4.7b.

As shown in §4.5.3, this should lead to increasing the probability on the alternative candidate, which is one where only the immediately following PW after the focused PW is dephrased. The alternative candidate should be (c) ([(kɨman bulmjɔn)(tʃok^{h}et*a)]) for the Stop Verb-ing sentence. In the upper tableau, as expected, the alternative candidate in (c) ([(kɨman bulmjɔn)(tʃok^{h}et*a)]) gets a higher probability than unacceptable candidates (a) and (b). There is a small probability assigned to the default candidate in (a) but this may be removed if we assume an even higher weight on *FocAP-R. However, somewhat unexpectedly this candidate gets a

	Align Foc L	*Foc AP R	Align FocIP R	Align XP R	Wrap XP	*AP		
A. Variable phrasing for Stop Verb-ing sentence in Figure 4.7a A zero weight on AlignFocIP-R makes (c) more acceptable than (d) [[[[**kɨman**$_{Stop}$]$_{Foc}$pul-mjɔn]$_{VP}$]$_{CP}$[tʃok^{h}et*a]$_{VP}$]$_{S}$								
	5	5	0	4	2	1	$\mathcal{H}$	p
a. [(**k**)(p)(tʃ)]		-1	-2		-1	-3	-10	0.02
b. [(**k**)(p dʒ)]		-1	-1	-1	-1	-2	-13	0.00
→ c. [(**k** b)(tʃ)]			-1	-1		-2	-6	0.94
→ d. [(**k** b dʒ)]				-2		-1	-9	0.05
B. Variable phrasing for Start Verb-ing sentence in Figure 4.7b [[[kɨman$_{Start}$]$_{AdvP}$[**pul-mjɔn**]$_{Foc}$]$_{CP}$[tʃok^{h}et*a]$_{VP}$]$_{S}$								
→ a. [(k)(**p**)(tʃ)]		-1	-1		-1	-3	-10	0.12
→ b. [(k)(**p** dʒ)]				-1	-1	-2	-8	0.88
c. [(k **b**)(tʃ)]	-1	-1	-1	-1		-2	-16	0.00
d. [(k **b** dʒ)]	-1			-2		-1	-14	0.00

Table 4.14: MaxEnt Tableau for the sentences in Figure 4.7 when AlignFocIP-R has a weight of 0.

even higher probability than (d), since it satisfies both Align-XP-R and Wrap-XP better than (d) does. It is a result of a combination of factors. This Stop Verb-ing sentence has a left branching structure, has only three PWs, and the first PW of those is under focus. When these conditions are met, the candidate in (c) gets the highest probability, as opposed to (d), when we lower the weight on AlignFocIP-R to 0. When these specific conditions are met, even when the first PW (/kɨman/) is under focus, the most acceptable prosodic phrasing may not be the one that dephrases all PWs after the focused PW, but the one that only dephrases the immediately following PW after the focused PW, when the weight on AlignFocIP-R is low.

In the lower tableau, the two generalizations of Cho (1990) and Jun (1993) only predict that the candidate (b) should be the most acceptable because there are not enough PWs after the focused PW (the second PW in the sentence) to differentiate the two generalizations. In this case, the candidate in (a) gets a probability of 0.12 once we assume a zero weight on AlignFocIP-R. This is because it satisfies both

AlignFoc-L and AlignXP-R, while (c) and (d) do not. As a result, though (a) violates a high-weighted constraint *FocAP-R, it is still more acceptable than (c) and (d).

4.5.5 Section summary

In this section, I have discussed the syntax-prosody mapping of the /kɨman/ + Verb construction, and argued that the difference in the information structure (which word is under focus) derives the difference in prosodic phrasing discussed in §4.4.

I have based my analysis on the proposals made in Truckenbrodt (1999) which uses AlignXP, WrapXP, and AlignFoc, in capturing focus-driven prosodic resturing. I have reviewed Kim (2015)'s analysis which proposes to use Rightmost to account for focus-driven dephrasing in Korean. While these works have primarily focused on how the most acceptable phrasing becomes optimal in an OT tableau, I have extended their analyses and proposed a MaxEnt analysis that assigns probabilities to the prosodic phrasing candidates, which I assume to correspond to how acceptable each candidate is.

I have formalized AP phrasing rules of Jun (1993) which predicts a set of acceptable and unacceptable candidates in terms of weighted constraints where AlignXP-R, WrapXP, and *AP interact. I have shown that these constraints can correctly assign reasonable probabilities to the prosodic phrasing candidates which are predicted by Jun (1993)'s AP phrasing rules.

In addition, I have shown constraint interaction between AlignFoc-L, *FocAP-R, and AlignFocIP-R, may account for variation in focus-driven dephrasing (e.g., at least the immediately following PW after the focused PW is dephrased, but PWs following it may or may not be dephrased) in Korean.

I have shown that this set of constraints can account for the variable prosodic phrasing of the /kɨman/ + Verb construction, which is the main focus of this chapter.

I have argued that /kɨman/ is under focus in the Stop Verb-ing sentence, while the verb following /kɨman/ is under focus in the Start Verb-ing sentence.

The MaxEnt analyses have made specific predictions about the relative acceptability of prosodic phrasings given a particular syntactic structure (left-branching or right-branching). Testing these predictions empirically awaits further studies, in order to corroborate these analyses.

4.6 Discussion

This section discusses two remaining issues.

First, I briefly discuss why /kɨman/ may be under focus in the Stop Verb-ing sentence, while the verb following /kɨman/ is under focus in the Start Verb-ing sentence.

Second, I discuss how assuming recursive prosodic structures can account for the difference in prosodic phrasing of the two /kɨman/ sentences, without having to assume either /kɨman/ or the verb following /kɨman/ is under focus. However, assuming recursive prosodic structures fails to capture the variation in prosodic phrasing of the /kɨman/ + Verb construction.

4.6.1 Difference in information structure

In §4.5, I have assumed that there is a difference in information structure between the two /kɨman/ + Verb sentences, which is why the two sentences are prosodically phrased differently. I discuss issues related to this assumption here. I repeat the relevant statements in Gim (2004)'s observation, which is introduced in §4.2.3. Gim (2004)'s original claim is that /kɨman/ receives a focus in the Stop Verb-ing sentence while it does not in the Start Verb-ing sentence.

> When /kɨman/ is used with the 'stop meaning', ... it is natural that [/kɨman/] receives a focus, since it is an important factor that changes the truth condition of the entire proposition.
>
> ...

> [On the other hand, when it is used with the 'start meaning'], /kɨman/ ... it does not serve as the focus of information. (Gim, 2004, p. 62)

In previous studies and in the current chapter, the term 'focus' refers to an element in the sentence that is compared with a set of alternatives (Rooth, 1992; Katz & Selkirk, 2011). Gim (2004) instead seems to use the term 'focus' in a slightly different way, as an element that is important for the truth condition of the proposition. In a sense, the way Gim (2004) uses the term 'focus' is similar to a negation marker, as it reverses the truth condition of the proposition. In Chapter 3, I have shown that the negation marker /an/ in Korean has been proposed to be a verbal prefix by some authors (Kim, 1999), meaning the verb following /an/ does not start a new AP. I suspect there may be a unifying semantic account of why kɨman$_{stop}$ and /an/ behave similarly in terms of prosodic phrasing, but I leave this for future research. In short, though I have to assume kɨman$_{stop}$ is under focus in the Stop Verb-ing sentence, I do not have a clear semantic answer why it is under focus, other than pointing out that a semantically similar word /an/, which is a negation marker, behaves in a similar way in terms of prosodic phrasing.

On the other hand, I have argued that it is reasonable to assume that the verb following kɨman$_{start}$ is under focus in the Start Verb-ing sentence, though it has not been explicitly stated in Gim (2004). I summarize the main arguments here. In §4.3, I have compared kɨman$_{start}$ with the sentential LE in Mandarin Chinese and the clausal adverbs, all of which is assumed to be adjoined to a higher position in the syntactic structure. In particular, one analysis of the sentential LE (Wang, 2018) has proposed that the sentential LE works as a focus marker, which sets a comparison between the proposition that it modifies (P) and a set of alternatives (p. 249), which may be described as the negation of that proposition (not P). Similarly, in §4.3.3, I have argued that clausal adverbs also set a comparison between the proposition that it modifies (P) and a set of alternatives, which may be described as the negation of that proposition (not P). I have argued that the meaning of kɨman$_{start}$ is similar to that of the sentential LE and clausal adverbs in that it sets a comparison between the proposition that it modifies (P) and the negation of that proposition (not P). For instance, /kɨman$_{start}$ pulɔs*-ta (blow-PAST-DECL)/ means that the state where the action of blowing has taken place (P), specifically in contrast to the state where it has not taken place (not P). In this regard, /kɨman/$_{start}$ may also be similar to 'only' in English, which has been argued to be a focus particle, which signals that some later element in the sentence is under focus (Rooth, 1992). Diagnosing whether /kɨman/$_{start}$ functions as a focus particle in the same way as 'only' does, requires further investigation.

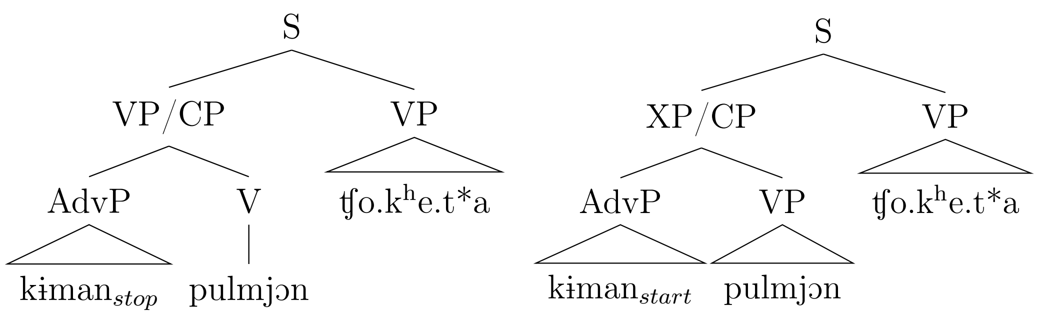
4.6.2 Recursive prosodic structures in Korean

In the analyses presented in §4.5, I have only considered prosodic phrasings that do not have recursive APs. However, in the literature, it has been proposed that recursive prosodic structures may be more suitable for capturing how syntactic structures, which are recursive in nature, map onto prosodic structures (e.g., Ito & Mester, 2012). In these proposals, the idea is that the constraint against prosodic recursion in the Strict Layer Hypothesis (Selkirk, 1984) may be violated, in the same way that other OT constraints (e.g., WrapXP) are violable. For instance, Truckenbrodt (1999) has shown that in a Bantu language, Kimatuumbi, diagnostics for prosodic phrasing suggest that certain syntactic structures may be mapped onto recursive prosodic structures.

In Korean, Baek & Yun (2018) and Lee (2022) proposed the recursion of APs to derive the prosodic structure of relative clauses more transparently from the syntactic structure. In these studies, each syntactic phrase (XP) is assumed to be matched with an AP (Selkirk, 2011).

Following these studies, I show briefly that if we assume prosodic recursion, it is possible to derive the difference in prosodic phrasing of the two /kɨman/ + Verb sentences, without having to assume either /kɨman/ or the verb following /kɨman/ is under focus.

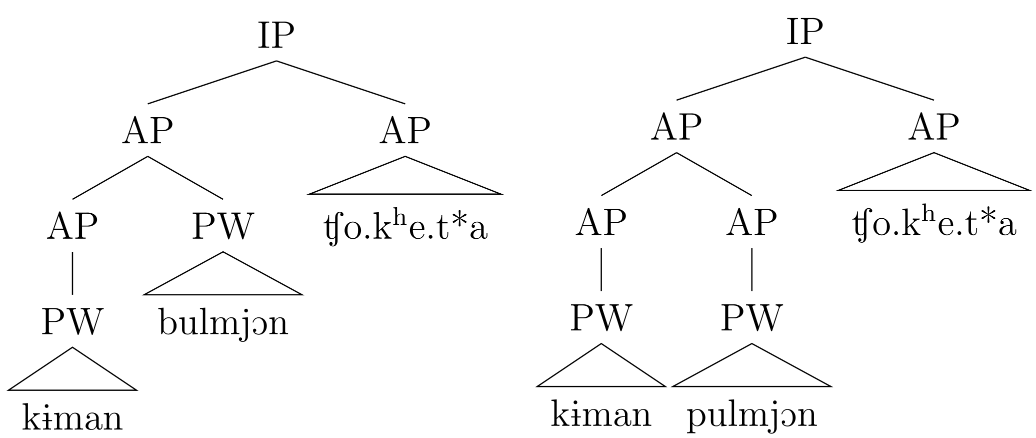
The syntactic structure for the /kɨman/ + Verb construction is in Figure 4.8, repeated from Figure 4.7. Adopting the recursive AP proposal, the prosodic structures for the /kɨman/ + Verb construction would look like the ones in Figure 4.9.



(a) 'It would be good if (you) could stop blowing (it).'

(b) 'It would be good if (you) could start blowing (it).'

Figure 4.8: Syntactic structures for the two /kɨman/ sentences, repeated from Figure 4.7



(a) Stop blowing

(b) Start blowing

Figure 4.9: The recursive AP proposal

I do not discuss how the prosodic phrasings in Figure 4.9 may be derived in an OT tableau, as a typical study that assumes prosodic recursion does (e.g., Truckenbrodt, 1999). Instead, I only outline the implications of this recursive AP proposal in Figure 4.9.

One major problem for assuming this prosodic structure is that we do not have an immediately obvious way to account for the expected variation in prosodic phrasing, when no PW is under focus (Jun, 1993). For a left-branching structure, such as the one in Figure 4.8a (the syntactic structure for the Stop Verb-ing sentence), the prosodic structure in Figure 4.9a may always be the optimal prosodic phrasing. It becomes a puzzle that why then the alternative prosodic phrasing where all PWs are phrased into one AP may also be acceptable. While it may be possible to account for this variation using recursive prosodic structures, I leave this for future studies.

The second implication of the recursive AP proposal concerns how we can diagnose the 'correct' prosodic structure given a syntactic structure. For instance, in the analysis for Kimatuumbi, Truckenbrodt (1999) uses two different phonological processes to diagnose left and right edges of a prosodic constituent. Truckenbrodt (1999) shows that 'shortening of a long stem vowel' is sensitive to right edges of phonological phrases while 'phrasal tone insertion' is sensitive to left edges. When there is a shortening on the right edge of the first PW, which is not immediately followed by a phrasal tone insertion on the left edge of the second PW, it is interpreted as evidence for a recursive prosodic structure, i.e., $((\mathrm{PW})_{pph}\ \mathrm{PW})_{pph}$ by Truckenbrodt (1999).

I will briefly explain why it may be not as straightforward to diagnose such a case in Korean. In Korean, a phrase final tone (Ha) is used as a diagnostics to identify the right edge of an AP (Choi et al., 2020; Hatcher et al., 2024, e.g.,). Acoustic realization of lenis stops serves as a diagnostics for an AP left boundary: if it is lenited, there is no AP boundary on the left edge of that lenis stop. The main problem of this diagnostics is that it crucially relies on the fact that the phrase final tone is acoustically distinct from the phrase medial '+H' tone. Depending on the assumptions of how prosodic structure is acoustically realized, we may end up with different diagnoses of the prosodic structure of the same acoustic signal.

Suppose we have a sequence of four syllables with a melody of 'LHLH'. For the sake of discussion, I assume that the first and the second H tones are similar in F0, and likewise, the first and the second L tones are similar in F0. The onset of the third syllable is a lenis stop, which is realized as lenited. There may be three interpretations of this acoustic signal, as listed in (35).

(35) Multiple interpretations of a sequence of four syllables with a melody of 'LHLH' where the onset of the third syllable is a lenited lenis stop

a. $(\mathrm{LH})_{AP}(\mathrm{LH})_{AP}$: Lenis is optionally lenited in the AP-initial position (Jun, 1993)

b. $(\mathrm{L\ {+}H\ L{+}\ Ha})_{AP}$: +H is not acoustically distinct from Ha, and L is not acoustically distinct from L+ either (Chapter 3)

c. $((\mathrm{LHa})_{AP}\mathrm{LHa})_{AP}$: Both H tones are interpreted as AP-final tones. Lenis is lenited as it does not align with an AP left edge.

The first interpretation prioritizes the tonal evidence and essentially treats lenition of lenis stop as a noise in the data (lenis may be lenited in the AP-initial position, just as it may be non-lenited in the AP-medial position). This seems to be the most common assumption in the literature which follows the original proposal of Jun (1993).

In previous chapters of this dissertation, I have put forward the second interpretation. In Chapter 2, I have presented that there are no clear separate clusters of high F0 values, which may correspond to +H and Ha, in the PW-initial position. In contrast, I have identified two separate clusters of lenited and non-lenited lenis stops in the PW-initial position. In addition, in Chapter 2, I have also argued that lenition of lenis obstruent may not be as optional as previously assumed. These results support the interpretation in (b) in (35), for a sequence of four syllables with a melody of 'LHLH' where the onset of the third syllable is a lenited lenis stop.

However, the recursive prosodic structure in (c) in (35) offers a new way to interpret the same acoustic signal, which assumes that both diagnostics (tonal and segmental) are both reliable. I do not attempt to present evidence that one interpretation should be favored over the others in this dissertation. I only point out that the recursive prosodic structure in (c) needs to be considered as a possible alternative, though it has not been extensively explored in the literature (cf. Baek & Yun (2018); Lee (2022)).

In order to test the validity of the recursive AP proposal in Korean, we need a carefully designed production experiment where we can derive the three prosodic structures in (35), by designing a sequence of syllables that may differ in syntactic structure. For instance, for (a), we need a sequence of two disyllabic NPs, spoken in a list, and in a neutral context where no word is under focus, so that we would expect every NP to be phrased on its own. For (b), we would need a four-syllable-long NP, spoken in a context where we would phrase this NP on its own as an AP. Finally, for (c), we would need a sequence of two disyllabic NPs again, but spoken in a context where the second NP is higher than the first NP in the syntactic structure (e.g., [[AdjP NP-GEN] NP-ACC], as in Figure 4.5). Acoustic data from this experiment should then be analyzed to see if we can find a reliable acoustic difference between the three different prosodic structures.

In the next chapter, I consider different prosodic structures, including the recursive AP, for the /kɨman/ + Verb construction, and evaluate them with acoustic data from a pilot production experiment. The sentences from the pilot experiment are not precisely controlled in terms of their syntactic structure, since I only recorded /kɨman/ + Verb constructions with two meanings. Still, I aim to present a proof-of-concept study which considers multiple interpretations of the prosodic structure for a given acoustic signal, rather than assuming one prosodic structure that prioritizes the tonal evidence (e.g., (a) in (35)) as the only possible interpretation.

4.7 Chapter summary

To conclude, I summarize how the three main goals of this chapter have been achieved.

First, I have proposed a preliminary syntactic analysis for the /kɨman/ + Verb construction (see Figure 4.7). I have argued that depending on the meaning, kɨman/ seems to occupy different positions in the syntactic structure. I have shown that the proposed syntactic structure for the /kɨman/ + Verb construction has similarities to two well-studied cases of syntactic ambiguity: the verbal/sentential ambiguity of the particle LE in Mandarin Chinese (Wang, 2018) and the clausal/manner adverbial ambiguity (Ernst, 2002; Egedi, 2009; Kubota, 2015). In both cases of ambiguity, when the ambiguous word (the particle LE or the adverb) occupies a syntactically lower position, it modifies the verb phrase, and when it occupies a higher position, it serves a discourse-pragmatic function (Ernst, 2002; Kubota, 2015; Wang, 2018). I have argued the ambiguity of /kɨman/ exhibits syntactic and semantic similarities to these cases.

Second, I have proposed that the two meanings of /kɨman/ have distinct prosodic structures based on the initial observations of Gim (2004). The major difference between the prosodic structures is whether /kɨman/ is prosodically phrased with the verb that follows it together in an AP (see Figure 4.3). When /kɨman/ and the verb are phrased together in the same AP, the sentence is interpreted as 'Stop V-ing' as in the context described in (2a). When they are phrased in different APs, the sentence is interpreted as 'Start V-ing' as in the context described in (2b). I have additionally reported that for each meaning of /kɨman/, the prosodic phrasing of the verb following /kɨman/ and the sentence-final PW (the matrix verb) may vary, and does not affect the overall interpretation of the sentence.

Third, I have formalized previously described syntactic restrictions on forming APs in Korean (Jun, 1993) in the form of violable constraints, and presented a MaxEnt analysis that captures both kinds of variation in prosodic phrasing: when no PW is under focus and when there is focus-driven dephrasing.

In the next chapter, I will present a pilot production experiment to confirm the proposed prosodic structures using acoustic data. As mentioned at the beginning of this chapter, this chapter also lays the necessary groundwork for Chapter 6, which presents a prosodic disambiguation experiment, using the /kɨman/ + Verb construction.

CHAPTER 5

PRODUCTION EXPERIMENT

5.1 Introduction

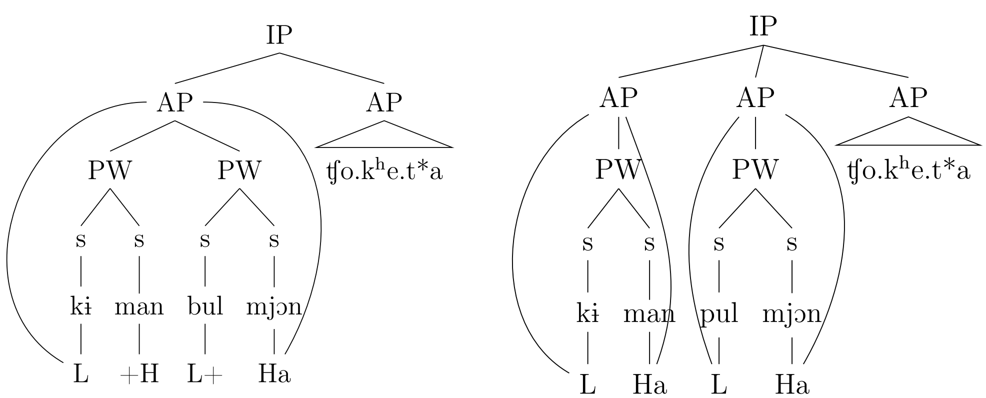
This chapter presents a pilot production experiment that investigates the acoustic implementations of the /kɨman/ + Verb construction. For an example of this construction, see (1), which is repeated from Example (2) in §4.1.

(1) /kɨman pul-mjɔn ʧoh-kes*-ta/

up to that degree/amount blow-COND be good-FUT-DECL

a. Grammatical in a context where a child has been blowing something, e.g., a balloon, to the extent that the balloon is about to pop (i.e., 'stop blowing')

b. Grammatical in a context where a child has not been blowing balloons, though the child has been asked to do so, and the speaker wants to say that it is about time to blow them (i.e., 'start blowing')



(a) Stop blowing (b) Start blowing

Figure 5.1: Prosodic trees for the sentences in (1)

In Chapter 4, I have proposed the prosodic structures in Figure 5.1 (repeated from Figure 4.3). These prosodic structures have been proposed, based on the observations made in Gim (2004), which are summarized in Table 5.1 (repeated from Table 4.2). To summarize the proposal presented in §4.5, I have argued that difference in prosodic phrasing of /kɨman/ and the following verb is derived from the fact that different word is under focus, depending on the meaning of /kɨman/. When /kɨman/ means 'stop', /kɨman/ is under focus, which causes post-focal dephrasing, causing the verb to be prosodically phrased together with /kɨman/. When /kɨman/ means 'start', on the other hand, the following verb is under focus, and as a result the verb starts a new AP.

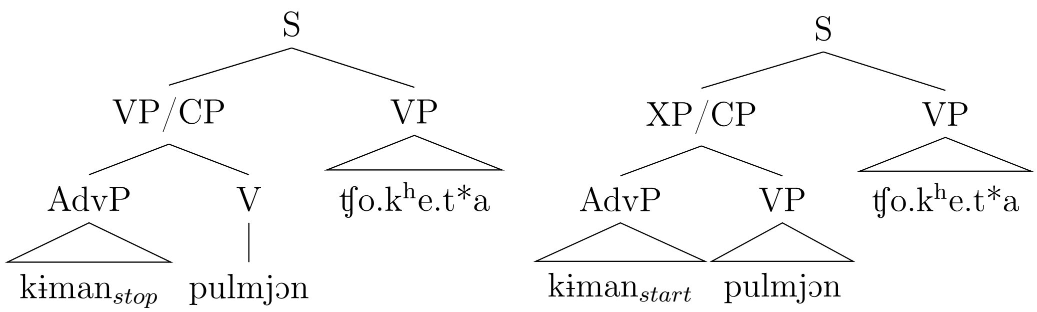
In Chapter 4, I have also discussed the prosodic phrasing of the verb following /kɨman/ and the sentence-final matrix predicate. In this section, we limit the discussion to the prosodic phrasing of /kɨman/ and the following verb, as that seems to be what is consistently conditioned by the meaning of /kɨman/.

While the prosodic structures capture Gim (2004)'s observations, the observations themselves are subject to investigation, since they were based on the author's impressions. The first goal of this chapter is to test if /kɨman/ + Verb sentences differ in prosodic phrasing with acoustic data collected from a pilot production experiment. For details of expected acoustic implementations of these prosodic structures, see §5.2.

Meaning of /kɨman/	Prosodic phrasing	Word under focus
Stop	[/kɨman/ Verb]	/kɨman/
Start	[/kɨman/] [Verb]	Verb

Table 5.1: Summary of the observations on p. 62 in Gim (2004)

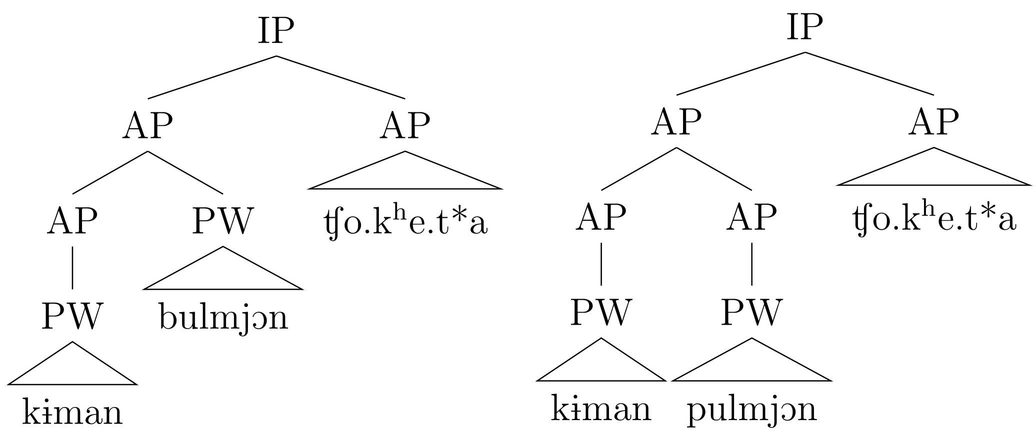
In §4.6.2, I have also proposed alternative prosodic structures for the sentences in (1) in Figure 5.3 (repeated from Figure 4.9), based on the proposal that the sentences in (1) have syntactic structures in Figure 5.2 (repeated from Figure 4.8).



(a) 'It would be good if (you) could stop blowing (it).'

(b) 'It would be good if (you) could start blowing (it).'

Figure 5.2: Syntactic structures for the two /kɨman/ sentences, repeated from Figure 4.8



(a) Stop blowing

(b) Start blowing

Figure 5.3: The recursive AP proposal for the sentences in (1)

These alternative prosodic structures differ from Figure 5.1 in that they involve recursive APs–an AP dominating another AP, violating the non-recursivity assumption of the Strict Layer Hypothesis (Selkirk, 1986). I have explained that these recursive prosodic structures have not been widely considered as viable options for representing the prosodic phrasing in Seoul Korean (cf. Baek & Yun (2018) and Lee (2022)). One potential reason for this may be that there is a tendency to prioritize F0 evidence for diagnosing an AP right edge (i.e., Ha tone), and to treat segmental evidence for diagnosing an AP left edge (i.e., Lenis voicing) essentially as noise in the data. Many authors (e.g., Jun, 1993, 1994; Cho & Keating, 2001) have assumed that Lenis voicing is optional within an AP and thus is not a reliable cue for prosodic structure. However, as I have shown that segmental realization is not optionally conditioned by the prosodic structure in previous chapters, I argue that it is worth investigating whether there is acoustic evidence for recursive prosodic structures. The second goal of this chapter is therefore testing if the acoustic data collected from the pilot experiment provide support for these alternative prosodic structures. For this goal, I investigate F0 and syllable duration to see if there is evidence for recursive prosodic structures. I discuss how F0 and syllable duration may be used to diagnose if there is a recursive AP in the prosodic structure in §5.2.3.

In addition to F0 and syllable duration, I also investigate how prosodic phrasing is implemented in terms of intensity. In Chapter 3, I have shown that the acoustic correlate for prosodic phrasing may differ depending on the laryngeal category of the PW-initial syllable onset. I have argued that prosodic phrasing, specifically whether a PW starts a new AP, is implemented in terms of variation in F0 when a PW starts with an Aspirated obstruent (stop and affricate), but in terms of variation in intensity when a PW starts with a Lenis obstruent. Following this finding, the third goal of this chapter is to investigate whether there is a difference in how prosodic phrasing of /kɨman/ + Verb is implemented, depending on the laryngeal category (Lenis vs. Aspirated) of the verb-initial obstruent.

Methodologically, this chapter uses a data-driven approach to investigate if two sentences differ in prosodic structures, rather than relying on a perceptual transcription system. In a typical prosodic analysis, researchers first transcribe the collected data with an existing prosodic transcription system (e.g., K-ToBI (Jun, 1993, 2000)), and uses acoustic data (some parameterization of F0 data, e.g., maximum pitch excursion (Lee & Lee, 2013)) to validate their transcriptions. Success of this approach depends on the assumption that transcriptions of the data are accurate and that the parameterization of the acoustic signal is done in a way that reflects the difference between the prosodic structures. For instance, following this approach, researchers would first have to label the data collected from a production experiment, using K-ToBI labels. Let's for the sake of argument that most of Stop Verb-ing sentences were labeled as 'L +H L+ Ha', while most of Start Verb-ing sentences were labeled as 'L Ha L Ha'. Essentially, this means that researchers have completely relied on the perceptual judgment of transcribers to determine whether there is a prosodic difference between the two meanings of the /kɨman/ sentence. Prosodic transcriptions are routinely followed by some parameterization of acoustic data (e.g., measuring pitch excursions), but this step is only to validate the transcriptions, rather than to directly compare how two sentences differ acoustically. Finding the adequate parameterization of the acoustic data is another challenge, and it necessarily involves some assumptions about how the assumed prosodic structure is expected to be implemented. For instance, Tremblay et al. (2019) has assumed that the slope of F0 fall is the acoustic correlate of whether there is an AP boundary: a steep slope indicates a boundary, while a gradual one does not. Lee & Lee (2013) on the other hand has assumed that a larger pitch excursion indicates an AP boundary while a smaller one does not.

In this chapter, I use Functional Data Analysis (FDA) (Ramsay & Silverman, 2005) to investigate how the two sentences in (1) differ in the shapes of F0 and intensity contours. Differences in shapes of F0 and intensity contours are then interpreted to reflect differences in prosodic structure. FDA is a data-driven approach that allows researchers to rely on the data to identify the nature of variation in the data, without making any strong prior assumptions. In the analysis, the variation in the data is first captured with a set of so-called Principal Component (PC) functions, each of which represents a variation in shape identified from the set of input curves. Since the analysis parameterizes the entire contour as a function, rather than taking point-wise measurements, it allows for a more nuanced comparison of two contours. Each input curve (e.g., F0 values measured over time from a sentence) is represented with a set of coefficient value for these PC functions. Two sentences in (1) may be compared using these PC coefficients.

The PC functions therefore serve as both an atheoretic, data-driven representation of the original acoustic data, and coefficients of these functions serve as a means of comparing the two meanings of /kɨman/. This differs from the traditional approach where representation of the data is based on transcribers' subjective impression, and comparing representations of the data (i.e., ToBI labels) depends on the assumptions of how these tonal labels are acoustically implemented. For details, see §5.3.

To summarize, this chapter has three research questions outlined in (2).

(2) Three research questions of the current chapter

a. Difference in phrasing: Do the two meanings of /kɨman/ + Verb differ in prosodic phrasing, as proposed in Figure 5.1 (and in Figure 5.3)?

b. Recursive APs: Are the recursive prosodic structures in Figure 5.3 supported by the acoustic data?

c. Segment specific acoustic implementation: Is the prosodic phrasing acoustically implemented differently depending on the laryngeal category of the verb-initial obstruent, as proposed in Chapter 3?

To preview the findings, I show that the two meanings of /kɨman/ + Verb are reflected in differences in prosodic phrasing, which is in turn implemented differently depending on the laryngeal category of the verb-initial obstruent (2a and 2c). I show that the F0 contour differs between the two meanings when the verb starts with an Aspirated obstruent, but the intensity contour differs when the verb starts with a Lenis obstruent. Finally, I argue that neither F0 nor syllable duration provides sufficient evidence for the recursive prosodic structures in Figure 5.3 (2b).

This chapter proceeds as follows. §5.2 discusses phonetic correlates. Methods are described in §5.3. §5.4 presents results. §5.5 provides a chapter summary and discussions.

5.2 Phonetic correlates of prosodic structures

In §4.4.1, I have discussed the expected acoustic implementations of the proposed prosodic structures in Figure 5.1. Assuming the sentences in (1) have the prosodic structures in Figure 5.1, they are expected to have phonetic realizations of segments and tones summarized in Table 5.2.

	Stop V-ing (Fig. 5.1a)	Start V-ing (Fig. 5.1b)
Verb initial lenis voicing	[**b**]	[**p**]
Tones	L +**H** **L**+ Ha	L **Ha** **L** Ha

Table 5.2: Comparison of acoustic implementation of the two prosodic structures in Figure 5.1 repeated from Table 4.3

In this section, I discuss how the expected acoustic difference between the two /kɨman/ sentences varies when the laryngeal category of the verb-initial segment varies. Three kinds of phonetic correlates will be discussed: tonal implementation of the prosodic structure (§5.2.1), prosodically-conditioned acoustic realization of the verb-initial segment measured as change in the intensity contour (§5.2.2), and pre-boundary lengthening (§5.2.3).

5.2.1 Tones

As reviewed in Chapter 1, the most studied phonetic correlate for the prosodic structures in Seoul Korean in the past literature has been the tonal contour of the Accentual Phrase (e.g., Jun, 1993, 2000, 2006, 2007, 2011). In particular, the initial and final tones of the AP are important because once they are identified, they serve the function of delimiting AP boundaries. These tones will therefore be crucial in testing if the /kɨman/ + Verb construction has the prosodic structures in Figure 5.1.

While the K-ToBI labels make it seem obvious that the prosodic structures can be determined based on the tones, there are complications as some tonal labels are similar to each other. Due to this potential ambiguity, while the original K-ToBI model of Seoul Korean (Jun, 1993) assumes that the AP-medial tones ('+H' and 'L+') are phologically distinct from the initial (L/H) or the final (La/Ha) tones, more recent works use more coarsely defined binary tonal categories (L/H) without specifying the diacritics that mark the position in the AP (e.g., Yun & Lee, 2022; Hatcher et al., 2024). As summarized in Table 5.2, the sentences in (1) are expected to differ in the tonal sequences and it is an empirical question whether there is a measurable difference between the two expected tonal sequences: 'L +H L+ Ha' and 'L Ha L Ha' (cf. Lee & Lee, 2013).

5.2.1.1 Tone on the verb-initial syllable

One of the most notable phonetic correlates of AP tones is that the initial tone varies as a function of the laryngeal category of the initial segments. As a reminder, when the AP starts with a fortis or aspirated segment, the initial tone is H, and in all other cases, it is L. This is summarized in Table 5.3 which is repeated from Table 1.1 in Chapter 1. Segments that induce an H tone will be referred to as 'H-segments', and ones that induce an L tone will be referred to as 'L-segments'.

AP-initial Tone	Consonant types	Phones
H	Aspirated	/p^h, t^h, k^h, $ʧ^h$/
	Fortis	/p*, t*, k*, ʧ*/
	Fricatives	/s, s*, h/
L	Lenis	/p, t, k, ʧ/
	Nasals	/n, m/
	Vowels	/a, e, i, o, u, ɨ, .../

Table 5.3: H/L inducing segments

This tonal variation due to the AP initial segment suggests that in the /kɨman/ + Verb construction, if the verb starts with an H segment, the resulting tonal contours

Sentence meaning	kɨman	Verb	-mjɔn
kɨman$_{start}$	L Ha	**H**	Ha
kɨman$_{stop}$	L +H	**L+**	Ha

Table 5.4: Expected tones when the verb starts with an H-segment, in the two /kɨman/ sentences

between the two /kɨman/ meanings will be largely different starting from the verb-initial syllable. In the prosodic structures in Figure 5.1, the verb in the kɨman$_{start}$ sentence starts a new AP, but not in the kɨman$_{stop}$ sentence. Therefore, when the verb starts with an H segment, the F0 on the verb-initial syllable in the kɨman$_{start}$ will be much higher than the verb-initial syllable in the kɨman$_{stop}$ sentence, as summarized in Table 5.4. Such verbs that start with an H-segment will be referred to as 'H-verbs', and ones that start with an L-segment will be referred to as 'L-verbs'.

5.2.1.2 Phonetic implementation of /L +H L+ Ha/ vs. /L Ha L Ha/

The tonal contours for the two /kɨman/ sentences for L-verbs are expected to be much more similar to each other, than the ones for H-verbs. This is especially true when the verb phrase is short (2-3 syllables), as it is in the sentences in (1).

When the verb phrase is longer, it may be easier to diagnose the prosodic difference between the two /kɨman/ sentences. For example, Jun & Oh (1996) found that when verb phrase is longer than 4 syllables, the F0 contours are measurably different between two different phrasings of the same sequence of words. For an example, I review the Yes-no/WH question ambiguity. As mentioned in previous chapters, in Korean, a Yes-No question and a WH question can have the same sequence of words, which are prosodically disambiguated (Cho, 1990), as in (3) (repeated from (1) in §4.1). Jun & Oh (1996) argued the indefinite pronoun (/ɔntʃe/ 'any time' in (3a)) and the verb phrase form separate APs in a Yes-No question, while the WH word (/ɔntʃe/ 'when' in (3b)) and the verb form a single AP (Jun & Oh, 1996) in a WH question.

(3) /atʃumɔni-nɨn ɔntʃe ɔtʃirɔwɔ-jo?/

madam-TOP when/any time be dizzy-HON

a. Yes-No question: Is there any time that you feel dizzy, madam?

b. WH question: When do you feel dizzy, madam?

Jun & Oh (1996) diagnosed the difference in phrasing by showing they have different tonal contours. In a production experiment, they showed that the tonal fall from the second syllable of the ambiguous WH word /ɔntʃe/ to the first L target had a different duration as a function of question type. This is illustrated in Table 5.5. When the question was a Yes-No question (3a), there was an immediate fall in the first syllable of the verb, i.e., /ɔ/ in /ɔ̲tʃirɔwɔ-jo/, which suggested that the verb initiated a new AP, and the AP-initial 'L' was anchored on the verb-initial syllable. On the other hand, when the question was a WH question (3b), the L target was identified in the penultimate syllable of the verb, i.e., /wɔ/ in /ɔtʃirɔw̲ɔ̲-jo/, because that was where the AP-medial 'L+' was expected to be anchored. They also found that the two question types differed in the boundary tone type. H% was the most frequent boundary tone type for Yes-No questions, and LH% was for WH questions.

Yes-No question						WH question						
Indefinite	Verb					WH-word		Verb				
σ σ	σ̲	σ	σ	σ	σ	σ	σ	σ	σ	σ	σ̲	σ
L Ha	L	+H	...	L+	H%	L	+H	...			L+	LH%

Table 5.5: Syllables and tones for the two question types when the verb phrase is long

The prosodic phrasing of the two question types is similar to the expected difference in prosodic phrasing for the /kɨman/ + Verb construction. Intuitively, in a Yes-No question, the focus may be thought to be carried by the verb, since the answer depends on the truth value of the verb. In a WH question, the WH-word may be under focus, as it is what the information that the speaker seeks. Ishihara (2002) has argued that WH-words bear prosodic prominence in Japanese because WH-words are inherently under focus in WH-questions, and PWs following WH-words are deaccented. This difference in the information structure may be responsible for difference in prosodic phrasing, as demonstrated in Chapter 4 for /kɨman/ + Verb sentences. I leave investigating the potential syntactic difference in the two question types, and whether there is a unified semantic or syntactic analysis for the ambiguity of question types and the ambiguity of the /kɨman/ + Verb construction, to future work.

Yun & Lee (2022) investigated the same question type ambiguity, but when the verb phrase was shorter. Recall that when the verb is longer, the L target after the H target in the peninitial syllable is expected in the verb-initial syllable for Yes-No questions, and in the verb penultimate syllable for WH-questions. When the verb is disyllabic, however, the verb-initial syllable is the verb penultimate syllable as shown in Table 5.6. They argued that when a short verb phrase formed a single AP with the preceding disyllabic WH word, in a WH question, the penultimate 'L+' target was not realized.

Yes-No question		WH question	
Indefinite	Verb	WH-word	Verb
$\sigma\ \sigma$	$\underline{\sigma}\ \ \sigma$	$\sigma\ \ \sigma$	$\underline{\sigma}\ \ \sigma$
L Ha	L H%	L +H	... LH%

Table 5.6: Syllables and tones for the two question types when the verb phrase is short

However, there seems to be a confound in the question type ambiguity construction they investigated, which is that the sentence-final tones also differed by the question type. While the boundary tone on the verb was 'LH%' when the question was a WH question, it was 'H%' when the question was a Yes-No question. Therefore, it is not clear the difference in the tonal contour was due to the fact that there is a difference in the prosodic phrasing of the two PWs, or that there is a difference in the boundary tone. While I agree that Jun & Oh (1996)'s findings support that these two question types have different prosodic phrasings, I argue that the difference in the boundary tone seems to be a confound which makes it difficult to test if the tonal contours differ solely due to the difference in prosodic phrasing, when the verb after WH-word is short.

In the /kɨman/ + Verb construction, this confound of the boundary tone is avoided since both /kɨman/ sentences end with a conditional suffix /mjɔn/ bearing the AP-final tone 'Ha', and followed by the matrix predicate /ʧoh-kes*-ta/ ('be good-FUT-DECL'). If there is a difference in prosodic phrasing between the two meanings of /kɨman/ + Verb, the expected tonal sequences of these two sentences, given the prosodic structure in Figure 5.1 are [L +H L+ Ha]$_{AP}$ when /kɨman/ means 'stop', and [L Ha]$_{AP}$ [L Ha]$_{AP}$ when /kɨman/ means 'start', as shown in Table 5.7.

Sentence meaning	kɨman	Verb-mjɔn
kɨman$_{start}$	L **Ha**	**L** Ha
kɨman$_{stop}$	L +**H**	**L**+ Ha

Table 5.7: Expected tones when the verb starts with an H-segment, in the two /kɨman/ sentences

We expect to see a measurable difference in the F0 contours between the two sentences, if the prosodic phrasing is implemented in terms of the difference in tonal sequences (Ha L vs. +H L+). Namely, if the tonal sequence for /kɨman$_{start}$/ sentence is '[L Ha]$_{AP}$[L Ha]$_{AP}$', we would expect to see a larger fall (Lee & Lee, 2013) or a steeper fall (Tremblay et al., 2019) between /man/ and the verb-initial syllable in a /kɨman$_{start}$/ sentence, compared to a /kɨman$_{stop}$/ sentence.

5.2.2 Domain-initial strengthening and domain-medial continuity lenition

Besides the difference in tonal contours, we also expect to see a difference in the segmental realization of the verb-initial segment, if the two meanings of /kɨman/ + Verb differ in prosodic phrasing.

As reviewed in Chapter 1 and Chapter 2, the general pattern that consonants exhibit is that they are produced with a stronger articulation in the AP-initial position to mark the prosodic boundary (Domain-initial strengthening), and produced with a lenited articulation in the AP-medial position to signal the continuation of the current prosodic constituent (i.e., 'continuity lenition' (Kingston, 2008; Ennever et al., 2017; Katz & Pitzanti, 2019)). All segments (consonants and vowels) are expected to exhibit such prosodically-conditioned variation in Korean (Cho, 2022), though segments may vary in how much they are prosodically conditioned. In this chapter, I focus on the following well-studied cases of prosodically conditioned lenition patterns in Korean: continuity lenition of lenis stops, VOT shortening of aspirated stops (Cho & Jun, 2000), optional /h/ deletion (Jun, 1993), and the realization of nasal consonants (as opposed to their AP-initial denasalized counterparts) (Cho & Keating, 2001; Jang et al., 2018). This is summarized in Table 5.8.

Segment type	Acoustic realization in the AP-medial position
Lenis	Lenition
Aspirated	Aspiration shortening
/h/	Optional deletion
Nasal	More nasal (denasalized AP-initially)

Table 5.8: AP-conditioned segmental realization

I test whether the change in the intensity contour is appropriate in quantifying the difference in the strength of segmental articulation for these segments. Lenition of lenis stops, /h/ deletion, and nasal realization are predicted to be captured by the shallower change in the intensity contour, since the AP-medial realizations of these segments may be characterized as being more sonorant and vowel-like compared to their AP-initial strong realizations, i.e., voiceless lenis stops, glottal fricative, and denasalized nasals. As for aspirated stops, while previous acoustic investigations focused on their VOT changes conditioned by prosodic position (Cho & Jun, 2000), given that the AP-medial aspirated stops are produced with a weaker articulation

(Cho & Jun, 2000), they may exhibit a shallower change in the intensity contour as well. Given the assumptions about the prosodic structures in Figure 5.1, the verb-initial segments are predicted to have a shallower change in the intensity contour in the kɨman$_{stop}$ sentence, compared to the onse in the kɨman$_{start}$ sentence.

5.2.3 Acoustic implementation of the recursive AP proposal

I have also proposed that the prosodic structures of the two sentences in (1) may involve recursive APs, as shown in Figure 5.3, repeated in Figure 5.4.

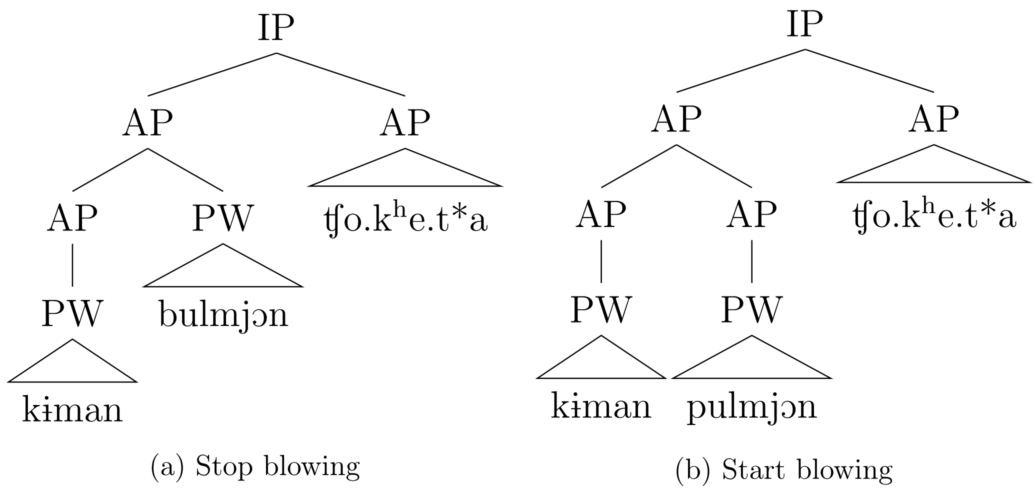


Figure 5.4: The recursive AP proposal for the sentences in (1) (repeated from Figure 5.3)

In terms of domain-initial strengthening and domain-medial continuity lenition, the recursive AP proposal is expected to be implemented in the same way as the non-recursive counterpart. In Figure 5.4, while the verb-initial segment is AP-medial when /kɨman/ means 'stop', it is at the AP-initial position when /kɨman/ means 'start'. As a result, the verb-initial segment is expected to be produced with a weaker articulation in the kɨman$_{stop}$ sentence, compared to the kɨman$_{start}$ sentence. As a result, we do not have a way to determine whether the recursive prosodic structure in Figure 5.4 is supported by the acoustic data, based on the segmental realization of the verb-initial segment alone.

It has not been explicitly studied how recursive prosodic structures are tonally implemented in Seoul Korean (Baek & Yun, 2018; Lee, 2022). I hypothesize that recursive prosodic structures in Figure 5.4 may be tonally implemented as in Table 5.9.

Meaning of /kɨman/	Stop	Start
Syllables	kɨ man pul mjɔn	kɨ man pul mjɔn
Tones	$[[\text{L Ha}]_{AP}$ L+ $\text{Ha}]_{AP}$	$[[\text{L Ha}]_{AP}$ $[\text{L Ha}]_{AP}]_{AP}$

Table 5.9: Hypothesized tonal implementations of recursive prosodic structures in Figure 5.4

If we hypothesize that AP right edge is marked with a Ha tone, we would see Ha tones on both /man/ and /mjɔn/ for both meanings of /kɨman/ + Verb. However, when /kɨman/ means 'start', the verb-initial syllable is expected to have an AP-initial L tone, while it is expected to have an AP-medial L+ tone, as it is the penultimate syllable of the higher AP. If we hypothesize that an AP initial L tone is realized as a lower F0 than an AP-medial L+ tone, we would expect to see a difference in the F0 contour between the two meanings of /kɨman/ + Verb. In particular, the fall from /man/ to the verb-initial syllable is expected to be larger in the kɨman$_{start}$ sentence, compared to the kɨman$_{stop}$ sentence. However, this acoustic difference (Ha +L vs. Ha L) is very similar to what is expected if the sentences have non-recursive prosodic structures (Ha L vs. +H L+), as discussed in §5.2.1.2. Therefore, the transition between /man/ and the verb-initial syllable does not differentiate the recursive prosodic structures from non-recursive ones.

On the other hand, if we hypothesize that when two AP edges are aligned, such an edge is realized differently from when there is a single AP edge, we would expect to see a higher rise at the end of the verb (i.e., a higher Ha in the kɨman$_{start}$ sentence). Baek & Yun (2018) has shown that when a syntactic constituent boundary aligns with multiple AP edges, predicted from the syntactic structure of the sentence, such an edge is realized with a 'stronger boundary', acoustically defined as an insertion of a pause, pitch reset, and a complex boundary tone (HL or LH). In Figure 5.4, while

/mjɔn/ is aligned with a single AP edge in the kɨman$_{stop}$ sentence, it is aligned with two AP edges in the kɨman$_{start}$ sentence.

In addition to the tone on /mjɔn/, there may be differences in syllable duration depending on the meaning of /kɨman/ + Verb. In the non-recursive AP proposal, the final syllable of each PW (/man/ and /mjɔn/) is either aligned with an AP boundary or a PW boundary. Jun (1993) has shown, however, that there is no pre-boundary lengthening at the AP right edge (Jun, 1993). On the other hand, in Baek & Yun (2018), it has been shown that when there are multiple AP edges, there is pre-boundary lengthening at the right edge. This predicts that the difference between the two PW-final syllables (duration of /mjɔn/ - duration of /man/) may be greater in the kɨman$_{start}$ sentence than in the kɨman$_{stop}$ sentence. In the recursive AP proposal, in the kɨman$_{stop}$ sentence, both /man/ and /mjɔn/ are aligned with a single AP boundary, while in the kɨman$_{start}$ sentence, /mjɔn/ is aligned two AP right edges, and /man/ is aligned with a single AP right edge.

In summary, the recursive AP proposal predicts that the two /kɨman/ + Verb sentences differ in the F0 height of tone on /mjɔn/, and the difference in syllable durations of PW final syllables (/man/ and /mjɔn/). In §4.5, I have argued that if we assume the recursive AP structures, we do not have to make the assumption that /kɨman/ or the following verb is under focus, in order to explain the difference in prosodic phrasing of the two sentences. In contrast, if we do not find acoustic support for the recursive AP proposal, we may have to assume there is difference in the information structure of the two /kɨman/ sentences, as it is the only way to account for a difference in prosodic phrasing.

5.2.4 Research questions reformulated

Based on the discussion in §5.2, I reformulate the three research questions in (2) as follows:

(4) a. Difference in phrasing: Do the two meanings of /kɨman/ + Verb differ in prosodic phrasing, as proposed in Figure 5.1?

(i) LHaHHa (start) vs. L+HL+Ha (stop): Is there a noticeable shape difference in the F0 contour between the two /kɨman/ sentences when the following verb starts with an H-segment?

(ii) LHaLHa (start) vs. L+HL+Ha (stop): Is there a shape difference (larger excursion or steeper slope) in the F0 contour between the two /kɨman/ sentences when the following verb starts with an L-segment?

(iii) Continuity lenition: Is the verb-initial segment produced with a weaker articulation in the kɨman$_{stop}$ sentence, compared to the kɨman$_{start}$ sentence?

b. Recursive APs: Are the recursive prosodic structures in Figure 5.3 supported by the acoustic data?

(i) High tone on /mjɔn/: Is the tone on /mjɔn/ higher in the kɨman$_{start}$ sentence than in the kɨman$_{stop}$ sentence?

(ii) Syllable duration: Is the difference in syllable duration between /mjɔn/ and /man/ larger in the kɨman$_{start}$ sentence than in the kɨman$_{stop}$ sentence?

c. Segment specific acoustic implementation: Is the prosodic phrasing acoustically implemented differently depending on the laryngeal category of the verb-initial obstruent, as proposed in Chapter 3?

These research questions will be addressed in the production experiment presented in the next section. If the answers to the three questions in (4a) are positive, then the proposed prosodic structures in Figure 5.1 are supported. If the answers to the two questions in (4b) are positive, then the recursive AP proposal is supported. If the answer to question (4c) is positive, then the corpus-based results of Chapter 3 that prosodic phrasing may be acoustically implemented differently, depending on the laryngeal category, will have been replicated with more controlled lab-based data. Results are presented in §5.4.

5.3 Methods

This section describes the methods for the production experiment. §5.3.1 presents the stimuli sentences, §5.3.2 describes the recording. §5.3.3 discusses data processing and acoustic measurements. §5.3.4 introduces how the time-series (F0 and intensity) data were parameterized using Functional Principal Component Analysis (FPCA) (Ramsay & Silverman, 2005; Aston et al., 2010; Cheng et al., 2010; Gubian et al., 2010; Zellers et al., 2010; Gubian et al., 2015, 2019; Hughes et al., 2023), in order to be data-driven in parameterizing the difference between the two /kɨman/ + Verb sentences, rather than transcribing the data with theory-dependent ToBI labels.

5.3.1 Stimuli

			Stem length		
Tone	Category	Phones	Mono	Di	Total
H	Aspirated	/pʰ, tʰ, kʰ, ʧʰ/	10	7	17
	Fortis	/p*, t*, k*, ʧ*/	8	11	19
	Fricatives	/s, s*/	4	6	10
		/h/	2	0	2
		Subtotal	24	24	48
L	Lenis	/p, t, k, ʧ/	15	37	52
	Nasals	/n, m/	8	22	30
	Vowels	/a, e, i, o, u, ɨ, .../	2	14	16
		Subtotal	25	73	98
		Total	49	97	146

Table 5.10: Number of monosyllabic and disyllabic verb stems by verb-initial segment type

A total of 146 Korean verb stems were chosen. The full list of verb stems is available in Appendix A: Verbs.[1] The verb stems are summarized in Table 5.10 by the verb-initial segment type (H/L-segment) and the verb-stem length in syllables (mono/disyllable). There are disproportionately many disyllabic verb stems that started with an L-segment in the Korean lexicon, and hence more disyllabic L-verbs were chosen than disyllabic H-verbs.

(5) a. /kɨman VERB-mjɔn ʧoh-kes*-ta/

kɨman$_{stop}$ VERB-COND be good-FUT-DECL
It would be good if you could stop V-ing.

b. /kɨman VERB-mjɔn ʧoh-kes*-ta/

kɨman$_{start}$ VERB-COND be good-FUT-DECL
It would be good if you could start V-ing.

These verb stems are recorded in the carrier sentences in (5). For each sentence, three repetitions were recorded, rendering 876 (146 stems * 2 sentences * 3 repetitions) sentences in total. Recording was made in a sound proof booth at the UMass Amherst Phonetics Lab.

5.3.2 Recording

The sentences were recorded in a single session with two blocks: Stop V and Start V. To produce the sentence with the intended meaning of /kɨman/, before each sentence, a semantically related sentence was produced with three repetitions. Adding the target and the paraphrases, a total of 1,752 (876 * 2) sentences were recorded. These paraphrases are in (6).

[1]Some of these verbs were adjectival verbs, such as /ʧ*a-/ 'be salty'. I assume these are not different from regular verbs in the syntactic structure in any way, and therefore not different in the prosodic structure either.

(6) a. /kjesok VERB-mjɔn konɾanha-ta/

continue VERB-COND be troubling-DECL
It would be troubling if you continued to V (= Stop V-ing)

b. /itʃe VERB-mjɔn tʃoh-kes*-ta/

now VERB-COND be good-FUT-DECL
It would be good if you now V-ed (= Start V-ing).

In each block, the sentences were presented on PowerPoint slides in a pseudo-randomized order. Each verb stem had a pair of slides containing the paraphrase and the target. For example, in the first block to record Stop V sentences, the sentence in (6a) was shown with one of the verb-stems. Three repetitions of (6a) were recorded. On the next screen, the target sentence in (5a) was presented, and three repetitions of the target sentence were recorded.

The paraphrases were therefore meant to serve as the reminder for the meaning of the /kɨman/ + Verb sentence. Unlike the /kɨman/ + Verb construction, the meaning of the paraphrase sentences were not dependent on the prosodic phrasing of the preverbal word and the verb phrase. Therefore, the paraphrases were not necessarily recorded with the same prosodic phrasing, and therefore with the same tonal contour, as the target sentence. Three repetitions of each experimental sentence were preceded by three repetitions of the paraphrase sentence, which did not need to have the same prosodic phrasing. This was intended to elicit a particular tonal melody, rather than a natural prosodic structure mapped from the meaning of the sentence.

This is a pilot experiment with an apparent limitation that the speaker is myself. In a future production experiment, a more concrete design is needed to elicit the sentences from the speaker, possibly providing the contexts in which each sentence may be produced.

5.3.3 Data processing and measurements

The recordings were cut into sentence-sized wav files. For each wav file, there was a text file containing the content of the wav file written in Korean script. These wav-text pairs were fed to a web-based Forced Aligner developed for Korean (Yoon, 2024). The Forced Aligner then output TextGrid files containing segmentations for Eojoels and individual phones. These TextGrid files were manually corrected for alignments. Using the same syllabification algorithm described in Chapter 2 phone intervals were merged to form syllable intervals, assuming the maximum onset principle. Syllable durations were measured by measuring TextGrid intervals.

F0 and intensity were measured every 10 ms over each wav file. For F0, minimum and maximum F0 were set to 50 Hz and 300 Hz, which were decided after examining the wav files for the pitch range. For intensity, the minimum pitch range was set to 75 Hz, and it was band-pass filtered with the range 0-400 Hz, following the setting for a male speaker used in Chapter 2.

5.3.4 Functional Principal Component Analysis

The F0 and intensity contours were parameterized using Functional Principal Component Analysis (FPCA), to be able to analyze the contours without having to assume the prosodic structure of the utterance and assigning K-ToBI labels on the data. As discussed in Chapter 1, the traditional approach that first assigns prosodic transcription labels on acoustic data suffers from a few limitations.

First, the labels are constrained by the particular theory of intonation that the researcher works with. The same F0 contour may be labeled differently depending on the theoretical assumptions, and the difference in labels may lead to different conclusions about the prosodic structure of a signal. For instance, as discussed earlier, the same F0 contour may be transcribed differently if one works under the assumption that APs can be recursive (Baek & Yun, 2018; Lee, 2022) or if one works under the assumption that APs cannot be recursive (Jun, 2000).

Second, the labels are assigned by the subjective impressions of the transcriber (Jun, 2022, p.163), and therefore they often need further validation such as checking interlabeler agreement, and checking how they differ in some acoustic measurement. Therefore, a given set of tokens from two linguistic categories (e.g., two meanings of /kɨman/) are parameterized twice before they are compared, first as a theory-dependent set of labels assigned by individual transcribers who may have different perceptual grammars, and second as an acoustic measurement, again chosen by the researcher, either to maximize the contrast between the transcription labels, or to reflect how the transcription labels are expected to differ according to the theory. Both of these parameterizations introduce a high degree of researcher freedom, which may bias the interpretation of the empirical data.

Although it has been reported that trained listeners disagree less in their label decisions compared to naive listeners (Jun et al., 2000), the fact that training changes label decisions poses questions of whether the resulting labels reflect how naive listeners perceive the auditory signal in a naturalistic situation. Moreover, the possibility that there are interlabeler disagreement reduces the replicability of a particular analysis. Additionally, contours may vary in a number of dynamic ways and the chosen acoustic measurements (e.g., the size of F0 excursion) may not fully reflect the actual ways in which two types of contours differ.

Finally, using the K-ToBI prosodic transcription system suggests the prosodic structures mainly differ in terms of F0. The labels 'Ha L' and '+H L+' for instance imply that there are tonal difference between a prosodic structure that has an AP boundary and one that does not. However, the impression that there is a prosodic boundary may not strictly depend on the difference in F0 contours. In principle, since the labeling is based on the perception of the audio signal, two sentences that have measurably different F0 contours may be labeled as having the same prosodic structure, and two sentences that have F0 contours that do not differ in any measurable way, may still be labeled with different ToBI labels, if they happen to differ in some other aspect, e.g., in their intensity contour.

In this study, I attempt to overcome these shortcomings by parameterizing the F0 and intensity contours in a data-driven way, rather than assigning theory-dependent labels based on subjective impressions, paying attention to a limited set of aspects in the data. To achieve this, I use Functional Principal Component Analysis (FPCA). FPCA is one of the statistical techniques grouped under the term Functional Data Analysis, which is developed to apply classic statistical tools (e.g., Principal Component Analysis) to time-series data, represented in terms of functions (Ramsay & Silverman, 2005). §5.4.1 presents more details of how FPCA is conducted.

Intensity contours, as well as F0 contours, are investigated because Intensity captures the expected lenition of segments in prosodic constituent medial position (i.e., 'continuity lenition' (Kingston, 2008; Ennever et al., 2017; Katz & Pitzanti, 2019)). Wagner & McAuliffe (2019) also demonstrated that intensity was a good cue for predicting the prosodic phrasing in data collected in a production experiment which tested the prosodic difference between sentences with [[A and B] or C] and [A and [B or C]] constructions. An interesting aspect of this study is that they analyzed intensity as a phonetic correlate of the prosodic phrasing, not the segmental realization. This is in line with the 'continuity lenition' perspective that the acoustic realization of the segment is a result of being in the medial position in a prosodic constituent.

FPCA has been used in a growing amount of phonetic research, varying in the type of phonetic variables investigated (F0, formants, or articulatory contours obtained with electromagnetic articulatography (EMA)). For example, it has been used to analyze segmental variation, such as the difference in formant contours between a diphthong and a hiatus in Spanish (Gubian et al., 2015) and CV coarticulation in

Polish (Gubian et al., 2019). It has also been used to investigate suprasegmental variation, such as the difference between question and statement in Neapolitan Italian (Gubian et al., 2010), prenuclear tones in Standard Southern British English (Zellers et al., 2010), tonal contraction in Taiwan Mandarin (Cheng et al., 2010), and phonetic implementation of phonologically different high tone spans in Luganda (Hughes et al., 2023).

5.4 Results

This section presents how the /kɨman/ + Verb sentences differ as a function of the sentence meaning. §5.4.1 presents the differences in tonal contours. §5.4.2 presents the differences in syllable duration. Finally, §5.4.3 presents the differences in the verb-initial segment realization.

5.4.1 FPCA with F0 contours

The typical workflow for conducting an FPCA is summarized in (7). These steps are discussed with a concrete example using the F0 contours for H_{short} sentences in §5.4.1.1. Providing detailed descriptions of each step is outside of the scope of this dissertation. The steps are taken from Gubian et al. (2015), which provides a detailed step-by-step tutorial.

(7) The workflow of Functional Principal Component Analysis (Gubian et al., 2015)

a. Preprocessing (Speaker normalization)

b. Smoothing and Landmark registration

c. FPCA

d. Class analysis (i.e., Regression model fitting)

e. Reconstruction of mean contour

5.4.1.1 Demonstration with F0 contours for H-verb sentences

The F0 contours for H-verb sentences are investigated for the following research question (7a-i): Does the H-verb initial syllable have a higher F0 in the kɨman$_{start}$ sentence than in the kɨman$_{stop}$ sentence? If the answer is yes, then it suggests that the verb starts a new AP in the kɨman$_{start}$ sentence.

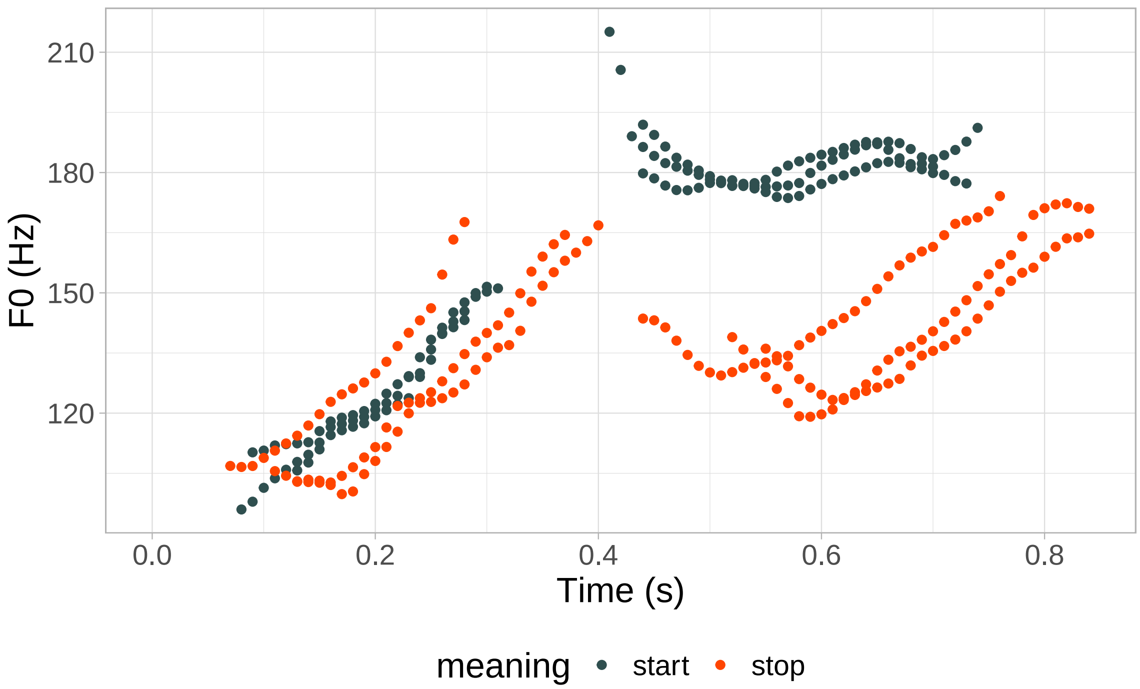


Figure 5.5: Raw F0 measurements from three repetitions of /kɨman pʰulmjɔn/ per each sentence meaning. Gap in F0 contour is due to the closure and aspiration at the beginning of the verb /pʰulmjɔn/.

In a typical workflow for FPCA, the raw contours are preprocessed to normalize for variation due to unrelated factors such as apparent pitch range difference according to participants' gender. As for the current pilot experiment, since all data came from a single speaker, this step is skipped.

Figure 5.5 presents raw F0 measurements per each sentence meaning measured from one of the H-verb stems, i.e., /kɨman pʰulmjɔn/. It is already noticeable that the F0 contour for the 'start' sentence rises to an H target and stays high, while the contour for the 'stop' sentence goes down to an L target before it goes up again.

This is predicted from the proposal that the verb starts a new AP in the ‘start’ sentence. The measurements from the matrix predicate after the verb phrase will not be considered here as they are expected to be irrelevant for distinguishing the sentence meaning.

These individual curves have different total durations. To be parameterized using the same PC functions, these contours need to be projected onto a fixed time interval. As they are time-normalized, an important landmark is registered. A ‘landmark’ refers to a particular time point that is linguistically meaningful (e.g., a syllable or word boundary). The relevant landmark is the PW boundary bewteen /kɨman/ and the verb. This time point is registered, meaning the time interval is warped such that the contours are synchronized with respect to the landmarks. Smoothing of the contours and the time-warping with respect to the landmark is coded using `landmarkUtils` package (Gubian, 2024) in R (R Core Team, 2023).

The smoothed and landmark registered contours are the input to the FPCA which was implemented with the `MFPCA` package (Happ-Kurz, 2022) in R (R Core Team, 2023). FPCA outputs the mean curve contour ($\mu(t)$), with a small number (n) of Principal Component (PC) curves (PC_1, PC_2 ... PC_n). Each PC curve represents a variation in shape identified from the set of input curves. As in standard Principal Component Analysis, PCs are ordered by the amount of variation that each PC captures in the data. PC1 by definition represents the shape component that the input curves vary in the most. FPCA provides a model of a given curve with multiple PC curves and the weights for each PC curve (s_1, s_2 ... s_n), named PC scores. A given input curve ($f(t)$) is approximated by a linear combination of weighted PCs added to the mean curve as in Equation 5.1. This linear combination is exactly the same as ordinary PCA, except each PC is a function of time, rather than vectors of real numbers.

$$f(t) \approx \mu(t) + s_1 * PC_1 + s_2 * PC_2 ... s_n * PC_n \tag{5.1}$$

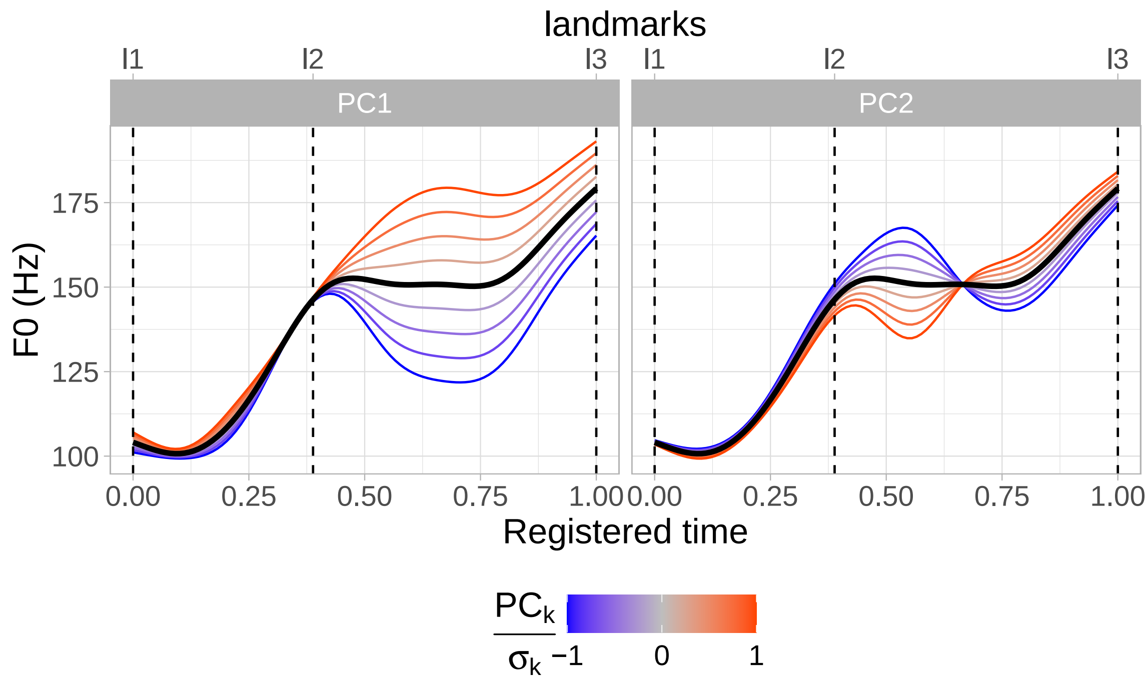


Figure 5.6: PC1 and PC2 from F0 contours of H-verb sentences before /ʧoh-kes*-ta/ (i.e., /kɨman/ Verb stem /mjɔn/) Black: mean contours, Blue: Mean contour minus quarters of standard deviation. Red: Mean contour plus quarters of standard deviation. Dashed lines: three landmarks, i.e., beginning of /kɨman/, the beginning of the verb, and the end of the verb. PC1 accounts for 65.9% variance, PC2 12.5% variance in the data set. Red and blue contours are formed by adding or subtracting quarters of standard deviation of each PC to the mean contour.

The two PCs found from the H-verb sentences are presented in Figure 5.6. The PC curves are shown in a normalized time scale (0-1) named 'Registered time'. The time scale takes into account that there are two PWs for all contours: /kɨman/ and the verb and the contours are synchronized using the landmarks. The dashed lines labeled 'l1', 'l2', and 'l3' mark the landmarks: the beginning of /kɨman/, the beginning of the verb, and the end of the verb, respectively. In both panels, the black thick contour is the mean contour, i.e., $\mu(t)$ in Equation 5.1. In addition to this mean contour, 8 additional contours are plotted for each PC, which are formed by adding or subtracting quarters of standard deviation of each PC, as the legend indicates. The red ones are formed by adding quarters of standard deviation of each PC to the mean contour, and the blue ones are formed by subtracting them from the mean contour. The contours with the strongest color are therefore the contours that vary from the mean contour by one standard deviation to either side of the PC dimension.

Figure 5.6 shows that PC1 describes whether the contour rises and stays high after the verb or whether it goes down before going up again. This PC accounts for 65.9% of the variance in the data. PC2 describes something similar to PC1, in terms of the contour shape, but it does not describe the pitch level difference after the verb, in the way PC1 describes it. PC2 only accounts for 12.5% of the variance.

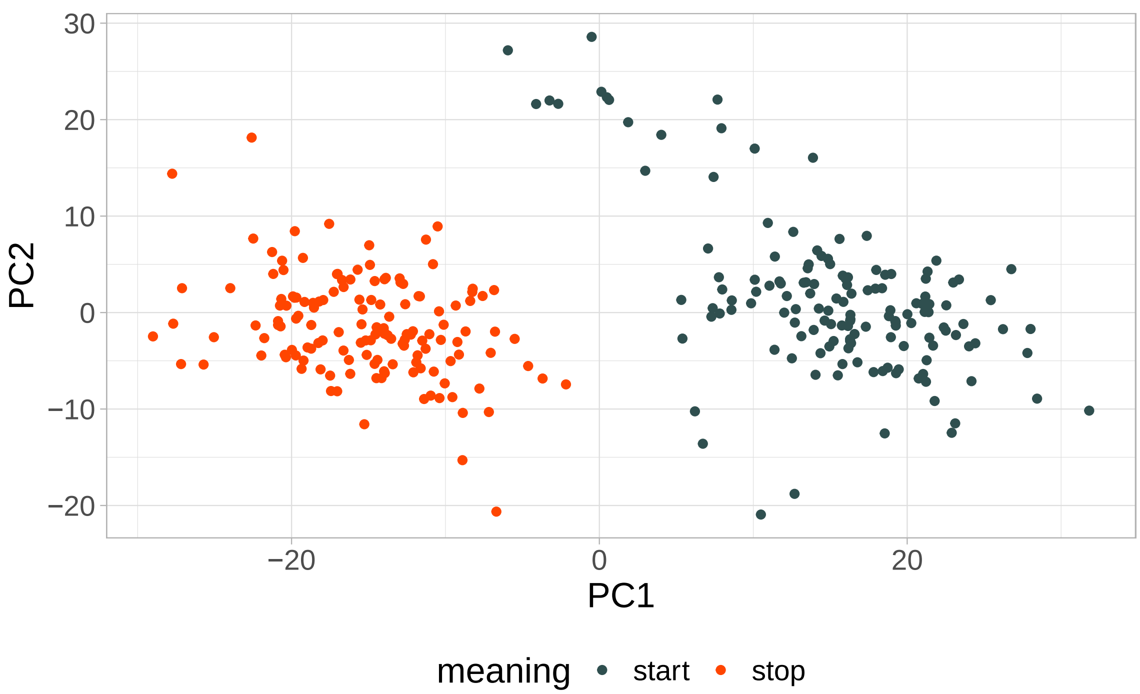


Figure 5.7: Scatter plot showing individual F0 contours in PC1 and PC2, colors represent the meaning of the contours.

The input curves, i.e., F0 contours of H-verb sentences are parameterized as points in the PC space defined with the first two PCs. Figure 5.7 plots individual contours in their PC scores for the first two PCs. 'PC1' will be used in place of s_1 here for convenience, though technically 'PC_1' refers to a function, and s_1 refers to the weight assigned to this function, when approximating a given input curve $f(t)$. The color of a point represents the meaning of the contour. As it is clear from the figure, F0 contours for kɨman$_{start}$ and kɨman$_{stop}$ sentences are separated on PC1. The ones with the 'start' meaning in black are clustered in the right half of PC1, while the ones with the 'stop' meaning in red are clustered in the left half of PC1. Being right on this dimension corresponds to being closer to the red contour and being left on this dimension corresponds to being closer to the blue contour in Figure 5.6.

A standard statistical tool such as regression analysis can be used to test whether PC1 can be predicted by the meaning of the sentence. In this study, a linear mixed effects regression model was fitted to test whether there was a reliable difference between the two meanings of /kɨman/ in terms of PC1, when other factors were considered such as the verb-initial consonant type (Ctype), stem length of the verb (Stem.length), and which repetition is the token from (1, 2, or 3) (Rep) and random variation for each item (a random intercept for each verb item). The model was fitted using `lme4` (Bates et al., 2015) and the statistical significance was computed with `lmerTest` (Kuznetsova et al., 2017) in R (R Core Team, 2023). The model formula is in (8). The same formula will be used for other FPCA analyses presented in the later sections, unless otherwise specified. The predictor 'Ctype' will be different for each FPCA, depending on which subset of the data is being investigated. As the data here consists of F0 contours for H-verb sentences, Ctype is a categorical predictor with three levels coding the categories of H-segments: aspirated, fortis, and /s, s*, h/.

(8) Statistical model formulas for FPCA in this study

a. Linear mixed effects regression

PC1 ~ Meaning * Ctype + Meaning * Stem.length + Meaning * Rep + (1|item)

b. Posthoc pairwise comparison for model in a.

emmeans(model, pairwise ∼ Meaning|Ctype + Meaning|Stem.length + Meaning|Rep)

Rep	Stem.length	Ctype	Estimate	SE	t.ratio	p.value
Repetition 1	Disyllable	Aspirated	33.9	1.31	25.908	<.0001
		Fortis	31.5	1.27	24.736	<.0001
		/s, s*, h/	32.0	1.39	23.004	<.0001
	Monosyllable	Aspirated	35.2	1.23	28.541	<.0001
		Fortis	32.7	1.20	27.228	<.0001
		/s, s*, h/	33.2	1.39	23.883	<.0001
Repetition 2	Disyllable	Aspirated	30.9	1.31	23.562	<.0001
		Fortis	28.4	1.27	22.323	<.0001
		/s, s*, h/	28.9	1.39	20.793	<.0001
	Monosyllable	Aspirated	32.1	1.23	26.047	<.0001
		Fortis	29.7	1.20	24.672	<.0001
		/s, s*, h/	30.1	1.39	21.673	<.0001
Repetition 3	Disyllable	Aspirated	28.8	1.31	21.952	<.0001
		Fortis	26.3	1.27	20.667	<.0001
		/s, s*, h/	26.8	1.39	19.275	<.0001
	Monosyllable	Aspirated	30.0	1.23	24.334	<.0001
		Fortis	27.5	1.20	22.916	<.0001
		/s, s*, h/	28.0	1.39	20.154	<.0001

Table 5.11: Result of the pairwise posthoc comparisons for the model fit with the formula in (8) testing whether two meanings of /kɨman/ differ significantly on PC1 across levels of Rep, Stem.length and Ctype.

As a reminder, the research question that we address here is: Does the H-verb initial syllable have a higher F0 in the kɨman$_{start}$ sentence than in the kɨman$_{stop}$ sentence? There may be differences in the size of the difference between the two meanings, varying by Ctype, Stem.length, and Rep, but these effects are supplementary. Consequently, instead of directly reporting the model result, the model was submitted to a posthoc pairwise comparison test, using `emmeans` (Lenth, 2018), which calculated 'estimated marginal means'. Only the result of this pairwise comparison test will be reported. The formula for this is in (8b). The estimated marginal means addressed the following question: Is the model predicted PC1 significantly different between the two meanings (Start - Stop) across levels of 'Ctype', 'Rep', and 'Stem.length'? The answer to this question is directly relevant for addressing the current research question. Since the contours for 'Start' have higher values in PC1 (see Figure 5.7), we expect the difference between the two meanings (Start - Stop) to be positive. The results are in Table 5.11.

The 'estimate' column in Table 5.11 shows the difference in the marginal means between PC1 for the 'Start' tokens and PC1 for the 'Stop' tokens, for each Stem.length and Ctype variable. Table 5.11 is divided into three parts for each repetition. In each part, the top half shows the estimated marginal mean difference for disyllbic stems and the bottom half shows the estimated marginal mean difference for monosyllabic stems. Across all combinations of Rep, Stem.length and Ctype, the estimated marginal mean difference between the PC1 for the 'Start' and 'Stop' are positive, and these differences are all statistically significant ($p < 0.05$). The results indicate that no matter which repetition the curves were from, what the verb stem length was, and what the verb-initial segment was, the F0 contours from the 'Start' sentences had significantly higher PC1 values than the ones from the 'Stop' sentences, which confirms the observation from Figure 5.7 that PC1 provides a good explanation of how the F0 contour shape differs by the sentence meaning.

Finally, once the PC that distinguishes the two meanings of /kɨman/ was identified, average contour ($\mu(t)$) per meaning was reconstructed using the category specific mean for the chosen PC score, i.e., PC1. The curves in Figure 5.8 show the reconstructed curves for each meaning, using the marginal means for PC1 averaging across other factors. The resulting contours can be conceived as representative models of F0 contours of kɨman$_{stop}$ + H and kɨman$_{start}$ + H. Using the K-ToBI notation, these contours may be analyzed as having 'L +H L+ Ha' and 'L Ha H Ha', respectively, which supports the proposal that the verb starts a new AP in the kɨman$_{start}$ sentence.

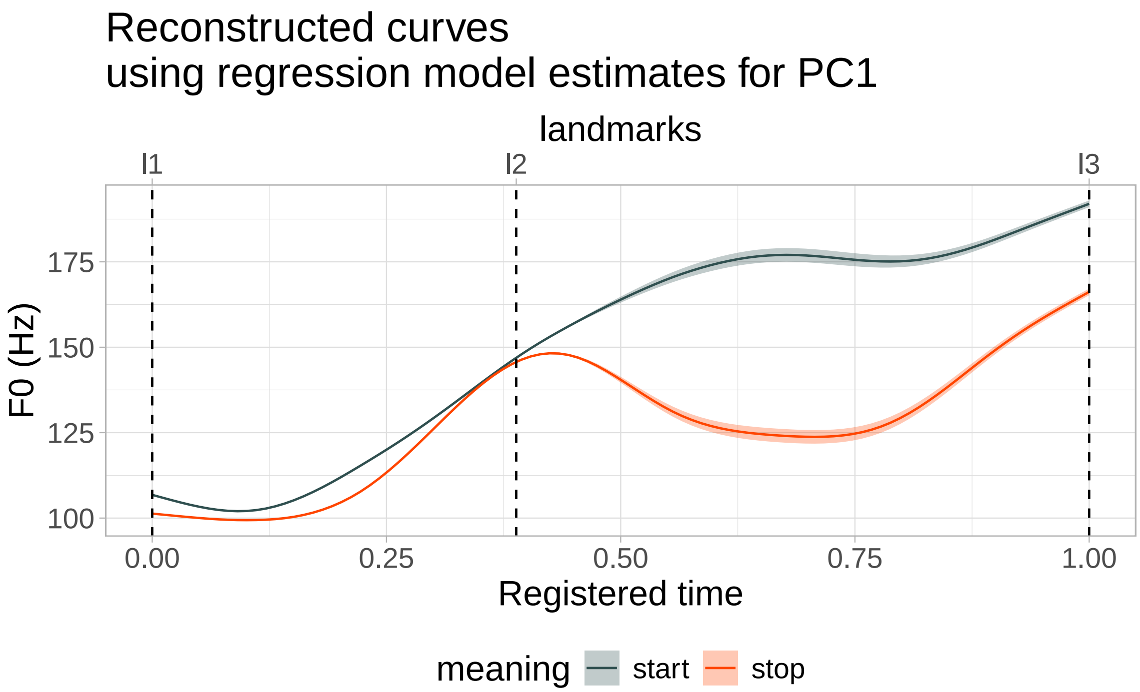


Figure 5.8: Reconstructed curves with marginal means for PC1 for each sentence meaning, averaging across Ctype and Stem.length.

5.4.1.2 Phonetic implementation of expected L +H L+ Ha vs. L Ha L Ha

The second research question we address is related to the F0 contours for L-verb sentences: Is there a difference in the transition between /man/ and the verb-initial syllable, depending on the meaning of /kɨman/? In the proposed prosodic structures, we expect the verb to start a new AP in the kɨman$_{start}$ sentence, but not in the kɨman$_{stop}$ sentence. These prosodic structures are expected to be realized with different tonal melodies: L Ha L Ha for kɨman$_{start}$ and L +H L+ Ha for kɨman$_{stop}$. We expect to find a smaller or more gradual transition in the kɨman$_{stop}$ sentence compared to the kɨman$_{start}$ sentence, which reflects the difference in the tonal melodies.

The two PCs found from the L-verb sentences are presented in Figure 5.9. PC1 describes the overall height difference and the late timing of the L target after the beginning of the verb ('l2' in the figure). This PC1 seems to describe the expected difference between L Ha L Ha and L +H L+ Ha tonal melodies. As a lower PC1 (blue) indicates a later alignment of the L target, it corresponds to the L +H L Ha melody. We expect the F0 contours for kɨman$_{stop}$ sentences to have a lower PC1 value compared to the ones for kɨman$_{start}$ sentences. This PC accounts for 38.1% of the variance in the data. Notably, PC1 found for L-verb sentences (38.1%) does not describe as much variance as PC1 found for H-verb sentences (65.9%). This suggests that the variation in the data is less systematic. PC2 describes the range of pitch movement on /kɨman/ and accounts for 21.8% of the variance.

Another potential way these F0 contours might vary was in the height of the verb-final H target. In the recursive AP proposal, recall that the right edge of the verb following kɨman$_{start}$ is aligned with two AP right edges, while the right edge of the verb following kɨman$_{stop}$ is aligned with only one AP right edge. I have hypothesized that this may lead to a boost in H-target at the end of the verb in kɨman$_{start}$ sentences, compared to kɨman$_{stop}$ sentences. This difference has been described as one that possibly If this was a systematic variation in the data, we would expect to see one of the PCs describing this variation. However, it seems that neither PC1 nor PC2 describes this kind of variation. This suggests that the recursive AP proposal is not supported by the F0 data.

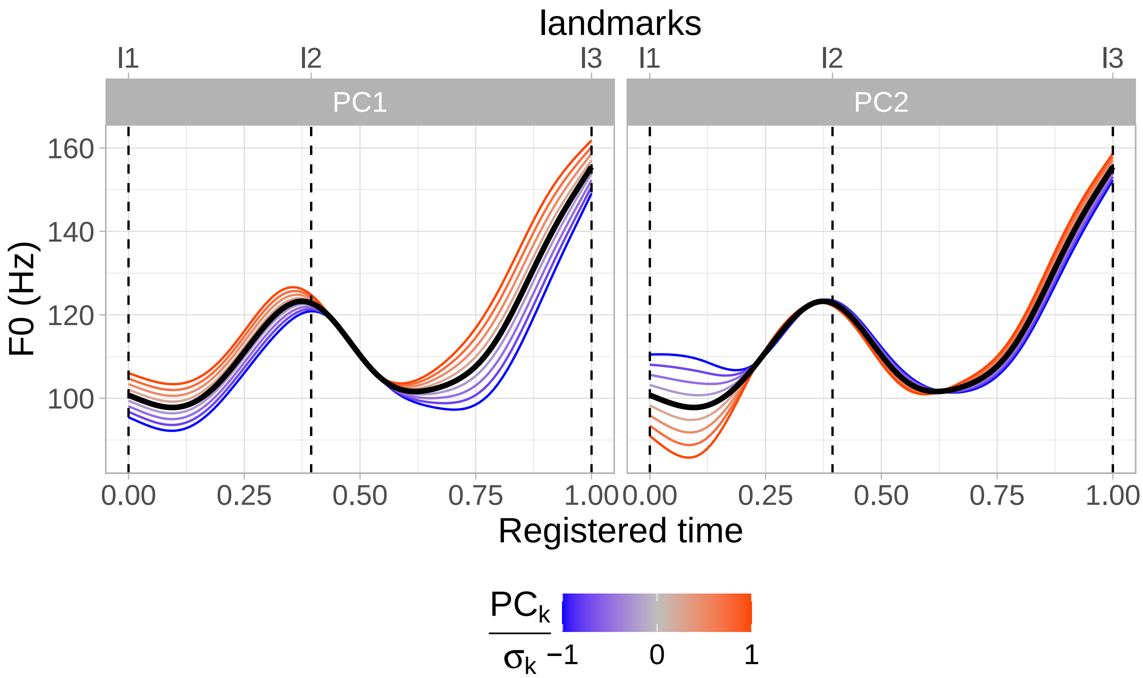


Figure 5.9: PC1 and PC2 from F0 contours of L-verb sentences. PC1 accounts for 38.1% variance, PC2 21.8% variance in the data set.

Rep	Stem.length	Ctype	Estimate	SE	t.ratio	p.value
Repetition 1	Disyllable	Lenis	4.282	0.740	5.789	<.0001
		LFillers	6.932	0.744	9.313	<.0001
	Monosyllable	Lenis	2.239	0.911	2.456	0.0144
		LFillers	4.889	0.966	5.062	<.0001
Repetition 2	Disyllable	Lenis	1.422	0.740	1.922	0.0551
		LFillers	4.072	0.744	5.471	<.0001
	Monosyllable	Lenis	-0.622	0.911	-0.682	0.4955
		LFillers	2.029	0.966	2.101	0.0362
Repetition 3	Disyllable	Lenis	1.560	0.740	2.109	0.0354
		LFillers	4.210	0.744	5.657	<.0001
	Monosyllable	Lenis	-0.483	0.911	-0.530	0.5962
		LFillers	2.167	0.966	2.244	0.0253

Table 5.12: Result of the pairwise posthoc comparison testing whether two meanings of /kɨman/ differ significantly (Start - Stop) on PC1 across levels of Stem length and Ctype.

As before, a linear mixed effects regression model was fitted to test whether there was a reliable difference between the two meanings of /kɨman/ in terms of PC1, when other factors were considered such as the verb-initial consonant type (Ctype), stem length of the verb (Stem.length), repetition number of the token (Rep) and random variation for each item (a random intercept for each verb item). Ctype was coded with two levels: lenis and LFiller. LFiller included all of the L-segments other than lenis stops, i.e., nasals and vowels. The model was submitted to a pairwise posthoc comparison test and the result is in Table 5.12. In most rows, the estimate was positive, meaning F0 contours for kɨman$_{stop}$ were lower in PC1, meaning they had later alignment of the L target, as expected.

The estimates were significant ($p < 0.05$) when the verb stem was disyllabic for both kinds of Ctype values, with the exception of lenis verbs in Repetition 2, which was marginally significant ($p = 0.0551$). When the verb stem was disyllabic, the L target was expected to be aligned to the initial syllable in kɨman$_{start}$ sentences, and to the penultimate verb-second syllable in kɨman$_{stop}$ sentences, as shown in Table 5.13. When the verb stem was monosyllabic, on the other hand, the estimated difference between Start and Stop meanings was consistently significant when the verb-initial segment was not lenis (LFiller). When the verb-initial segment was lenis, however, the estimated difference between Start and Stop meanings was only significant in Repetition 1 ($p = 0.0144$), but not in Repetitions 2 and 3 ($p = 0.4955$ and 0.5962, respectively).

	kɨman$_{start}$		kɨman$_{stop}$	
Stem length	kɨman	Verb	kɨman	Verb
Disyllable	σ σ	$\underline{\sigma}$ σ σ	σ σ	σ $\underline{\sigma}$ σ
	L Ha	L ... Ha	L +H	... L+ Ha
Monosyllable	σ σ	$\underline{\sigma}$ σ	σ σ	$\underline{\sigma}$ σ
	L Ha	L Ha	L +H	L+ Ha

Table 5.13: Syllables and tones for the two question types when the verb phrase is long.

These category specific estimated marginal means of PC1 were used to reconstruct F0 contours for each Ctype and Stem.length values. This is in Figure 5.10. The difference between the two meanings is very subtle in all four panels, and in particular, the two F0 contours for monosyllabic lenis verbs are not distinguished.

To answer the research question, therefore, there seems to be some evidence that suggests the AP initial L in the kɨman$_{start}$ sentence is aligned earlier than the AP-medial L+ in the kɨman$_{stop}$ sentence, but it is a very subtle difference, and may not be significant when the verb stem is monosyllabic starting with a lenis stop. This miniscule difference in production raises questions about whether this is perceptually salient for listeners. This is addressed in the discussion section (§5.5).

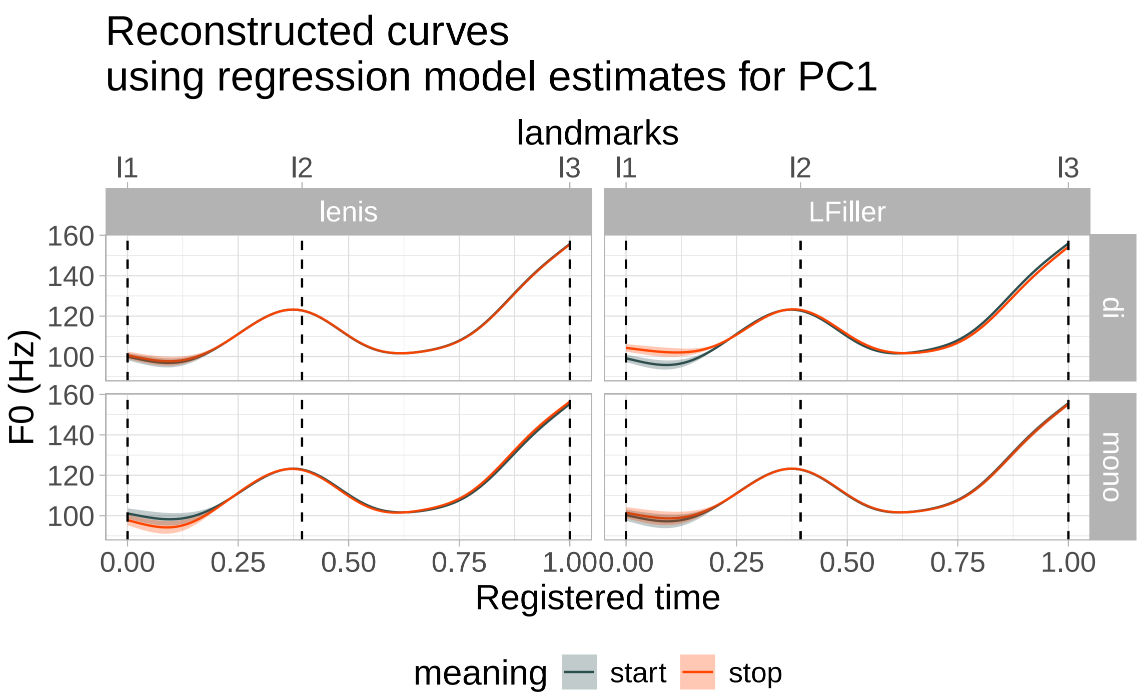


Figure 5.10: Reconstructed F0 curves with marginal means for PC1 for each sentence meaning, for each value of Ctype and Stem.length. The top rows contain contours when Stem.length is disyllablic, and the bottom rows contain contours when Stem.length is monosyllable. The left column corresponds to when Ctype is lenis, and the right column corresponds to when Ctype is LFiller: nasals and vowels.

5.4.2 Syllable duration: /man/ vs. /mjɔn/

Another prediction that the recursive AP proposal makes is that the difference in duration between the verb-final syllable /mjɔn/ and /man/ is greater in the kɨman$_{start}$

sentence compared to the kɨman$_{stop}$ sentence. In this proposal, /kɨman/ in both sentences is phrased as an AP on its own. The difference between the two is that in the kɨman$_{start}$ sentence, the right edge of the verb is aligned with two AP right edges, while in the kɨman$_{stop}$ sentence, it is aligned with only one AP right edge. If we assume that the number of AP right edges decides how much pre-boundary lengthening occurs (Baek & Yun, 2018), we would expect to see a greater difference in duration between /mjɔn/ and /man/ in the kɨman$_{start}$ sentence compared to the kɨman$_{stop}$ sentence.

To test this prediction, a linear mixed effects regression was fitted on the syllable durations of /man/ and /mjɔn/ in H-verb and L-verb sentences. The predictors were Meaning (a binary coded variable with 'Start' as the reference), PW.final.syllable (a binary coded variable with '/man/' as the reference), and the interaction of the two. A random intercept for each verb item was included. The result is in Table 5.14.

Fixed Effects	Estimate	SE	t value	Pr(> |t|)
(Intercept)	186.491	2.191	85.112	<2e-16 ***
Meaning (= Stop)	35.842	2.576	13.913	<2e-16 ***
PW.final.syllable (= /mjɔn/)	18.495	2.576	7.179	1.29e-12 ***
Meaning:PW.final.syllable	-5.717	3.643	-1.569	0.117

Table 5.14: Linear mixed effects regression on syllable durations of /man/ and /mjɔn/ with the formula: duration ∼ Meaning (stop) * PW.final.syllable (/mjɔn/) + (1|item).

Fixed effects of Meaning and PW.final.syllable were both positive and statistically significant, indicating that both /man/ and /mjɔn/ syllables had longer durations in the kɨman$_{stop}$ sentence than in the kɨman$_{start}$ sentence, and in general, /mjɔn/ was longer than /man/ in both sentences. The fact that /mjɔn/ was longer than /man/ was an unexpected results, based on the proposals for the prosodic structures in Figure 5.1. In the original proposal, both /man/ and /mjɔn/ are AP-final syllables in the kɨman$_{start}$ sentence, and thus not expected to differ in duration. This will be discussed further in §5.5.

However, the results did not support the recursive AP proposals either, since crucially the interaction between Meaning and PW.final.syllable was not significant, meaning that the difference between /mjɔn/ and /man/ was not different across the sentence meanings. The lack of interaction in syllable duration, and the lack of evidence suggesting a higher tone on /mjɔn/ for the kɨman$_{start}$ sentence do not support the recursive AP proposal.

5.4.3 FPCA with intensity contours

The last research question to address is whether the verb-initial segment in the kɨman$_{start}$ sentence is pronounced with a stronger articulation than the one in the kɨman$_{stop}$ sentence. As mentioned in §5.2.2, four types of consonants are investigated: lenis obstruents, aspirated obstruents, /h/, and nasals (/n, m/). If the verb starts a new AP, the verb initial consonant is expected to be realized with a stronger articulation. Acoustically, this means that the intensity is expected to drop more dramatically from the preceding sonorant consonant (/n/ in /kɨman/) to achieve a more complete closure for the consonant. Aspiration of lenis and aspirated obstruents, /h/ frication, and denasalization of nasals may all be described as having a more extreme intensity change compared to the preceding sonorant segment.

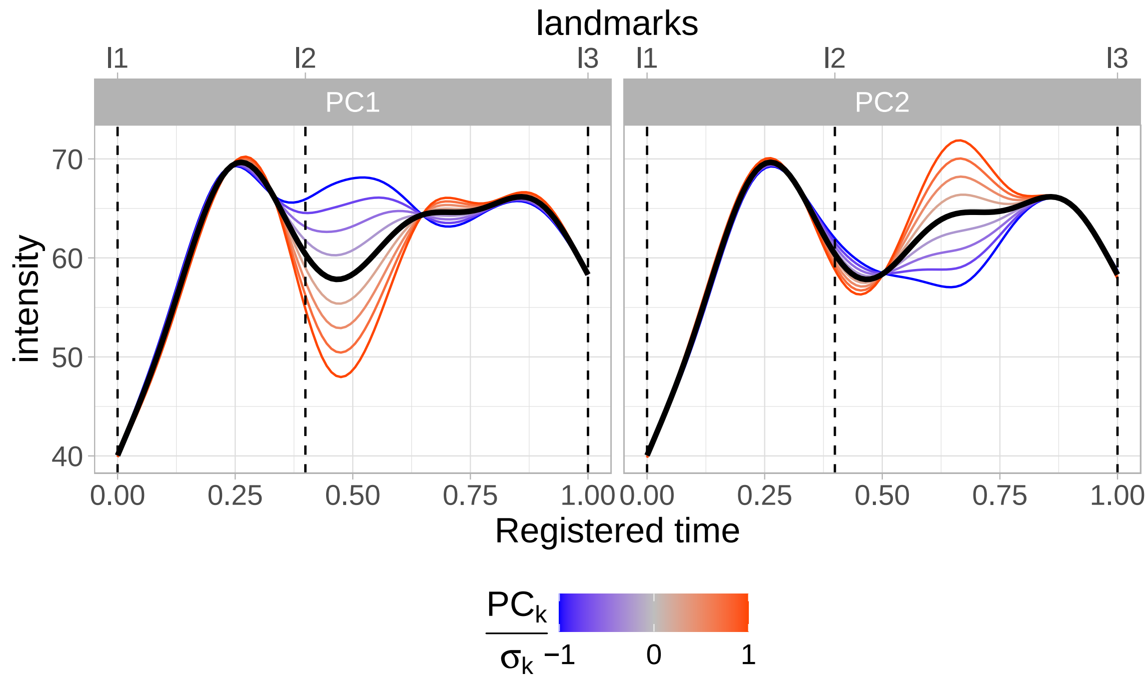


Figure 5.11: PC1 and PC2 from intensity contours of sentences with verbs that start with aspirated obstruents, /h/, lenis obstruents, or nasal consonant. PC1 accounts for 45.9% variance, PC2 24.0% variance in the data set.

The two PCs found from intensity contours of the sentences with verbs starting with the four types of consonants are presented in Figure 5.11. PC1 described how much intensity drops at the beginning of the verb ('l2') and it accounted for 45.9% of the variance in the data. PC1 was higher (red) when intensity dropped more at the beginning of the verb. Therefore, we expect the kɨman$_{start}$ sentence to have a higher PC1 than the kɨman$_{stop}$ sentence, and the difference between them to be a positive value.

To test the prediction, a linear mixed effects regression was fitted for PC1. The dependent variable was PC1, and the predictor was Ctype and Rep (Repetition), and random intercept for each item was included. Stem.length was not included because there were no /h/ disyllable stems in the data (see Table 5.10 for verb items). The model was then submitted to a pairwise posthoc comparison test, to check if the kɨman$_{start}$ sentence had a higher PC1 than the kɨman$_{stop}$ sentence, across different values of Ctype and Rep. The result is in Table 5.15. The estimates were all positive values and they were significant ($p < 0.05$) across value of Ctype and Rep.

Rep	Ctype	Estimate	SE	t.ratio	p.value
Repetition 1	Aspirated	0.529	0.251	2.106	0.0357
	/h/	4.583	0.645	7.101	<.0001
	Lenis	6.008	0.177	33.965	<.0001
	Nasal	0.523	0.206	2.534	0.0116
Repetition 2	Aspirated	0.776	0.251	3.093	0.0021
	/h/	4.831	0.645	7.485	<.0001
	Lenis	6.255	0.177	35.366	<.0001
	Nasal	0.771	0.206	3.734	0.0002
Repetition 3	Aspirated	0.821	0.251	3.270	0.0011
	/h/	4.876	0.645	7.554	<.0001
	Lenis	6.300	0.177	35.618	<.0001
	Nasal	0.815	0.206	3.950	0.0001

Table 5.15: Result of the pairwise posthoc comparison testing whether two meanings of /kɨman/ differ significantly (Start - Stop) on PC1 across levels of Ctype.

Intensity contours for each Ctype were computed using the results of the estimated marginal means. They are presented in Figure 5.12. As shown in Table 5.15, the amount of difference between the two meanings was different across Ctypes. Lenis had the largest difference between the meanings, followed by /h/, nasals, and finally aspirated. It seems that a lack of an oral constriction for /h/ allows for considerable reduction in intensity in the kɨman$_{stop}$ sentence, while nasals, even when they are denasalized (in the kɨman$_{start}$ sentence), are still produced with a high intensity, possibly because they are still voiced.

Going back to the research question, intensity contours for sentences with verbs that started with aspirated, /h/, lenis, and nasals were investigated. As predicted, the verb-initial segment in the kɨman$_{start}$ sentence was stronger than the one in the kɨman$_{stop}$ sentence for all segments investigated, though they differed how much they differed by the sentence meaning.

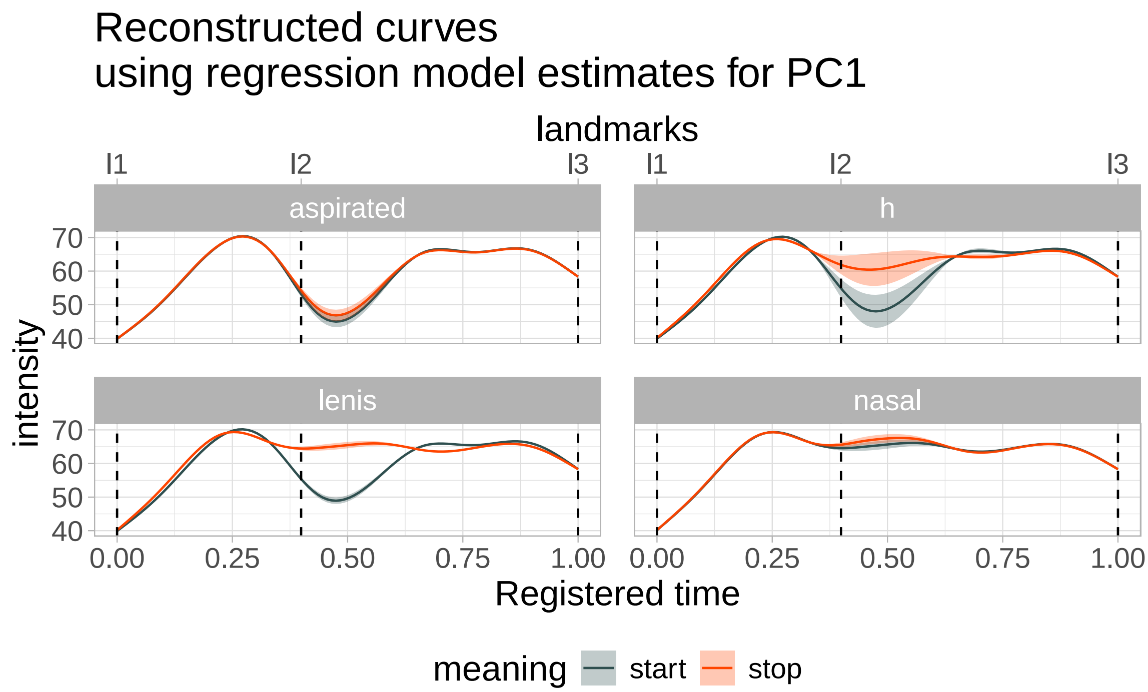


Figure 5.12: Reconstructed Intensity curves with marginal means for PC1 for each sentence meaning, for each values of Ctype (aspirated, /h/, lenis, and nasal)

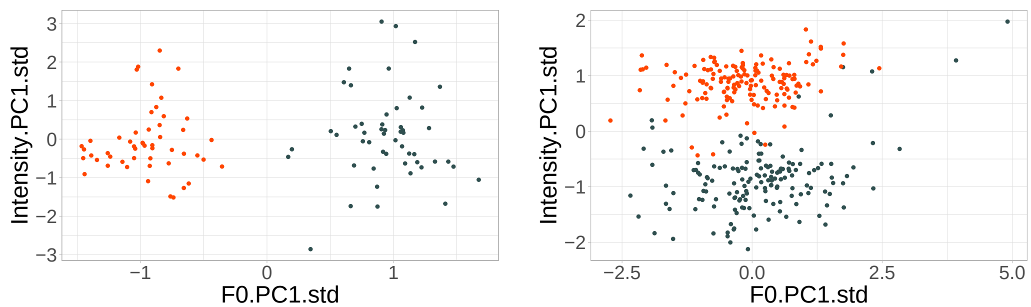
5.4.4 Segment-specific phonetic correlates for prosodic phrasing

In the K-ToBI model, the prosodic structures are always marked by delimiting AP-initial and final tones. The medial syllables were assigned to different tonal labels (+H and L+) primarily for phonetic reasons, and previous results show they are phonetically different compared to initial/final tones (Lee & Lee, 2013). Extending the corpus-based findings presented in Chapter 3, this study presents a case where AP phrasing is not phonetically implemented in terms of a difference in tonal contours, but in terms of a difference in intensity contours, which can be analyzed as segmental lenition. In particular, when the verb stem was monosyllabic and started with a lenis segment, there was no measurable difference between the F0 contours (§5.4.1.2), but there was a large difference between the intensity contours (§5.4.3).

The finding here does not suggest that the F0 contours of the monosyllabic lenis verb stems need to be or will be labeled with the same sequence of tones in the

K-ToBI system, if these contours are transcribed by K-ToBI transcribers. If the F0 contours of two /kɨman/ sentences are transcribed with the same labels, it suggests that they do not differ in the prosodic structures, just just when the verb is monosyllabic and starts with a lenis. Instead, it is most likely that they would still be transcribed differently: 'L +H L+ Ha' for the kɨman$_{stop}$ sentence and 'L Ha L Ha' for the kɨman$_{start}$ sentence. The finding suggest that the medial tones may be assigned because the transcriber already knows from other information, such as the intensity contour, that these syllables are AP-medial, even when they are not measurably different from the initial/final tones. Put differently, the prosodic structures are phonetically implemented with multiple dimensions and tonal labels represent the prosodic structures phonologically, but they may not necessarily show how prosodic structures differ phonetically.

To illustrate this, sentences with Lenis verbs and Aspirated verbs are investigated further. For each subset, F0 contours and intensity contours were parameterized using FPCA, and PC1s are standardized to be on the same scale. The PCs looked similar to the ones presented in earlier corresponding subsections and they are not repeated here. Individual contours are plotted in the scatter plot in Figure 5.13. There is a clear difference between lenis and aspirated segments in how they differentiate the sentence meanings. Aspirated verbs mainly differ by sentence meaning in terms of F0 variation, while Lenis verbs mainly differ in terms of intensity variation. Since they have the same syntactic structures, it is not likely that they have different prosodic structures. However, the results show the way the prosodic structures differ phonetically depends on the laryngeal category of the PW-initial segment.



(a) /kɨman/ Aspirated

(b) /kɨman/ Lenis

Figure 5.13: Scatter plot showing /kɨman/ Verb in the standardized F0 cue (z-transformed PC1 found from their F0 contours) and the standaradized Intensity cue (z-transformed PC1 found from their Intensity contours). Left panel shows data for Aspirated-initial verbs and right panel shows data for Lenis-initial verbs.

5.5 Discussion

In this study, three phonetic correlates of prosodic structures were investigated: F0, duration, and intensity, to address how prosodic structures differ between the two meanings of /kɨman/ + Verb sentences. I repeat the research questions in (9).

(9) a. Difference in phrasing: Do the two meanings of /kɨman/ + Verb differ in prosodic phrasing, as proposed in Figure 5.1?

(i) §5.4.1.1 LHaHHa (start) vs. L+HL+Ha (stop): Is there a noticeable shape difference in the F0 contour between the two /kɨman/ sentences when the following verb starts with an H-segment?

(ii) §5.4.1.2 LHaLHa (start) vs. L+HL+Ha (stop): Is there a shape difference (larger excursion or steeper slope) in the F0 contour between the two /kɨman/ sentences when the following verb starts with an L-segment?

(iii) §5.4.3 Continuity lenition: Is the verb-initial segment produced with a weaker articulation in the kɨman_{stop} sentence, compared to the kɨman_{start} sentence?

b. Recursive APs: Are the recursive prosodic structures in Figure 5.3 supported by the acoustic data?

  (i) §5.4.1.2 High tone on /mjɔn/: Is the tone on /mjɔn/ higher in the kɨman$_{start}$ sentence than in the kɨman$_{stop}$ sentence?

  (ii) §5.4.2 Syllable duration: Is the difference in syllable duration between /mjɔn/ and /man/ larger in the kɨman$_{start}$ sentence than in the kɨman$_{stop}$ sentence?

c. §5.4.4 Segment specific acoustic implementation: Is the prosodic phrasing acoustically implemented differently depending on the laryngeal category of the verb-initial obstruent, as proposed in Chapter 3?

First, the two meanings of /kɨman/ + Verb sentences were found to differ in prosodic phrasing, as proposed in Figure 5.1. This was clearly demonstrated with the F0 contours of sentences with H-verbs (§5.4.1.1). While the F0 contour for the kɨman$_{start}$ sentence had a higher F0 after the verb, which may be interpreted as an AP-initial H tone, due to the verb-initial H-segment (Jun, 1993), the F0 contour for the kɨman$_{stop}$ sentence had a lower F0 after the verb, which may be interpreted as an AP-medial L+ tone.

Second, the F0 contours for sentences with L-verbs (§5.4.1.2) also differed by the sentence meaning, though the difference was more subtle and was mostly limited to disyllabic verb stems that started with nasals or vowels. The found difference was in terms of the timing of the L target after /kɨman/, as reported in Tremblay et al. (2019). However, specifically for lenis stops, this difference was not significant, when the verb stem was monosyllabic.

Third, the sentences with lenis stops as verb-initial segments were found to differ in terms of intensity contour, instead. In addition to lenis stops, aspirated, /h/, and nasals were found to vary in terms of intensity change as a function of sentence meaning. This may be interpreted as suggesting these consonants exhibiting 'continuity lenition' (Kingston, 2008; Ennever et al., 2017; Katz & Pitzanti, 2019), though the extent to which they were lenited differed.

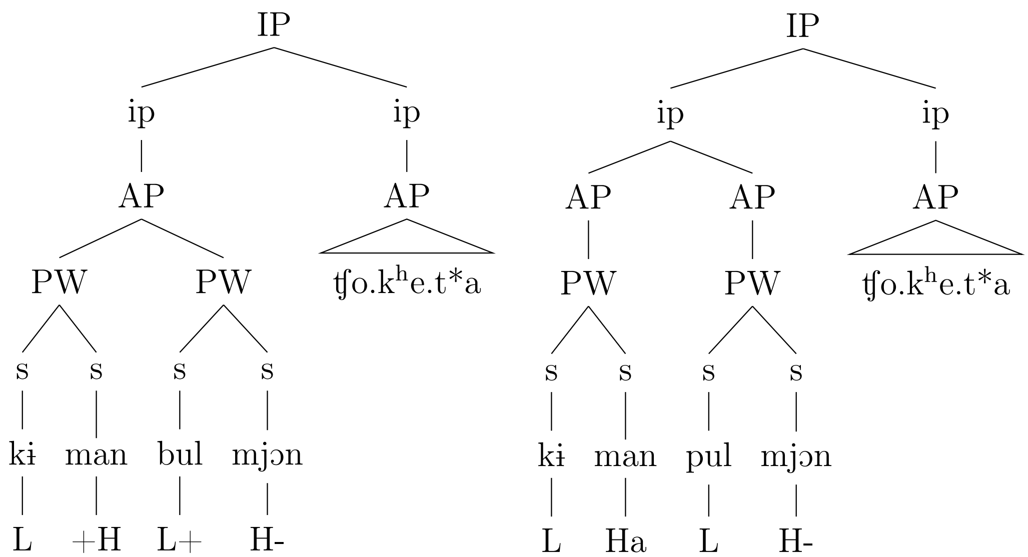
Fourth, the results did not support the recursive AP proposal in Figure 5.3. If the prosodic phrasing was recursive, we would expect the tone on /mjɔn/ to be higher in the kɨman$_{start}$ sentence than in the kɨman$_{stop}$ sentence, since /mjɔn/ is aligned with two AP right edges in the former, but only one AP right edge in the latter. This structural difference would also predict a difference in syllable duration between /mjɔn/ and /man/, specifically in the kɨman$_{start}$ sentence. However, neither of these predictions were borne out. This suggests that the prosodic structures were not recursive in the two /kɨman/ sentences, or at least that there is no empirical evidence to posit recursive structures at least for the sentences investigated here. This suggests that the prosodic phrasing difference reported here is due to the difference in the underlying information structure, as discussed in Chapter 4. Future studies could further investigate how the two meanings of /kɨman/ may lead to a difference in information structure, and therefore to a difference in prosodic phrasing.

This chapter presented a data-driven approach, FPCA, to investigate how prosodic structures differ between a pair of syntactically ambiguous sentences. Unlike a traditional approach that first labels the data with ToBI labels, the current analysis used PC curves to capture the prosodic characteristics of the sentences. The found PC curves were then investigated to see if the two syntactic structures differ in these particular shape characteristics. While K-ToBI labels prioritize the tonal evidence, the current approach allowed for a more objective comparison of F0 contours. Since FPCA can be applied to any time series data, it also allowed for the investigation of intensity contours, which revealed that the two meanings of /kɨman/ + Verb sentences differ in terms of intensity contours when the verb starts with a lenis stop. In §5.4.4, it was clearly shown that the prosodic phrasing difference between the two meanings of /kɨman/ + Verb sentences was phonetically implemented differently, depending on the laryngeal category of the verb-initial segment. This finding replicated the corpus-based findings presented in Chapter 3, and it suggests that the prosodic structures are phonetically implemented with multiple dimensions, and that tonal labels may not always reflect how prosodic structures differ phonetically. I argue that the K-ToBI tonal labels should therefore be understood as phonological symbols that reflect the prosodic structure, rather than reflecting how they are phonetically implemented in terms of F0 contours.

The fact that the tonal contours differed minimally for lenis verbs poses a perceptual question of what listeners pay attention to when they try to disambiguate the sentence meaning of the /kɨman/ + Verb construction. Previously, Gubian et al. (2010) tested impressionistically that the F0 contours for question/statement in Neapolitan Italian reconstructed using the PCs were good in quality, in that when PC score values at the two ends of a PC dimension that distinguished questions from statements, the resulting F0 contours were indeed unambiguous. When a PC score value was chosen in the ambiguous mid point of this continuum, however, the resulting F0 contour sounded ambiguous. Their results suggested that FPCA can be used to create controlled stimuli for perception experiment, to test whether listeners interpret a particular F0 contour with one meaning versus another. The reconstructed F0 contours described in this chapter were used to create stimuli for a perception study presented in Chapter 6.

Finally, I make revisions to the original proposal for prosodic structures in Figure 5.1, based on the unexpected finding that /mjɔn/ was longer than /man/ in both sentences (§5.4.2). In the original proposal presented in Figure 5.1, since /man/ and /mjɔn/ are both AP-final syllables in the kɨman$_{start}$ sentence, from the structure there is no reason why /mjɔn/ must be longer than /man/. In the kɨman$_{stop}$ sentence, while /man/ is PW-final (and not AP-final), /mjɔn/ is AP-final. However, given the expectation that AP-final syllables do not exhibit preboundary lengthening (Jun, 1993), we do not expect a duration difference between these two syllables in the kɨman$_{stop}$ sentence, either.

There are two potential explanations. First, /mjɔn/ may be longer than /man/ because it has an extra glide /j/. Second, there may be an intermediate phrase (ip) dominating the AP(s), as shown in Figure 5.14. By introducing an ip node, lengthening of /mjɔn/ may be analyzed to be due to being ip-final.



(a) Stop blowing (b) Start blowing

Figure 5.14: Prosodic trees for the sentences in (1) which explain why /mjɔn has a longer syllable duration than /man/ in both sentences. /mjɔn/ is longer due to ip-final lengthening.

The ip was first introduced in the K-ToBI model to explain a bigger prosodic juncture at the end of a relative clause (Jun, 2006, 2007). Given that /mjɔn/ (a conditional suffix) is at the end of a CP syntactically, just like the end of a relative clause, it suggests that 'CP's map to 'ip's in Syntax-prosody mapping. Further studies are needed to investigate whether /mjɔn/ is simply longer than /man/ due to segmental reasons, or due to other reasons such as the syntactic positions of these morphemes. Introducing the level of 'ip' may also change the syntax-prosody mapping proposed in Chapter 4 as it increases the number of ways PWs may be prosodically phrased (into APs and ips). I leave this for future research.

CHAPTER 6

PERCEPTION EXPERIMENT

6.1 Introduction

This chapter presents a perception experiment that tests how listeners interpret the /kɨman/ + Verb construction discussed in the previous two chapters, Chapter 4 and Chapter 5, when it is presented auditorily.

In the experiment, listeners were asked two questions. The first question was about which lexical item they heard in the position of the verb ('Choose the verb you heard'). The second question was about which meaning of /kɨman/ they interpreted the utterance as ('Choose the action that the speaker wants (proceed or discontinue)'). The first question tests how listeners decode the segmental realization of the verb-initial segment, while the second question tests how listeners decode the prosodic structure of the utterance. For each question, they were asked to answer a binary-forced choice question, choosing one out of two lexical items that differed in the phonemic category of the verb-initial segment (e.g., /pul/ 'to blow' or /p^hul/ 'to solve') for the lexical identification question, and one out of two meanings of the /kɨman/ + Verb construction (i.e., 'Start Verbing' or 'Stop Verbing') for the sentence interpretation question. These options are summarized in (1) and (2).

(1) Two choices for a lexical identification question for an audio-manipulated utterance containing /kɨman/ + Verb

a. kɨman Lenis verb (e.g., /pul/ 'to blow')

b. kɨman Non-lenis verb (e.g., /p^hul/ 'to solve')

(2) Two choices for a sentence interpretation question for an audio-manipulated utterance containing /kɨman/ + Verb

a. kɨman$_{start}$ Verbing

b. kɨman$_{stop}$ Verbing

This experiment investigates how listeners decode the acoustic signal when the F0 contour cue mismatched with the expected segmental realization of the verb-initial lenis obstruent. The F0 contour was manipulated such that it either signaled an AP boundary between /kɨman/ and the following verb, or not. The segmental realization of the verb-initial lenis obstruent was then manipulated to create a mismatch with the F0 contour. When the F0 contour signaled an AP boundary after /kɨman/, the verb-initial lenis obstruent was manipulated to signal no boundary, and vice versa.

When the F0 contour mismatches the segmental realization, listeners may choose to decode the acoustic signal in one of the following ways. First, they could prioritize the F0 cue over the segmental realization. This prediction follows from the view that segmental cues are uninformative and unreliable, and thus possibly not used in perception (Jun, 1994; Arvaniti & Baltazani, 2005). In Chapter 3 and Chapter 5, I have shown that when it comes to distinguishing whether there is an AP boundary before a lenis obstruent or not, F0 cues may be less informative compared to how lenis obstruent is realized. However, it is still an empirical question whether listeners would still prioritize F0 cues, despite their potential lack of distinctiveness in perception. I will refer to this pattern as 'F0 prioritization'.

Second, listeners may instead prioritize the segmental realization over the F0 cue. For instance, when listeners hear an F0 contour signaling an AP boundary, while the following lenis obstruent is not what they would expect in this context (i.e., if the lenis is voiced), they may interpret the utterance as having no prosodic boundary, prioritizing the segmental realization. This is a plausible pattern given previous findings in the literature that suggest listeners are sensitive to segmental realization in auditory sentence processing (e.g., Scott & Cutler, 1984; Mitterer et al., 2021; Choi & Mazuka, 2001, 2003; Yoo, 2020; Farinella & Lee, 2024). I will refer to this pattern as 'Segment prioritization'.

Third, listeners may process the sentence differently, by identifying the lexical item as having a different phoneme category. One way to resolve the mismatch between the F0 contour and the segmental realization is to categorize the segment as a different phoneme. For instance, as reviewed in Chapter 1, lenis and aspirated obstruents have no difference in terms of VOT in the AP-initial position (e.g., Kang & Guion, 2008; Bang et al., 2018), and no difference in terms of F0 in the AP-medial position (Lee, 2018). Therefore, depending on the F0 contour, a given lenis obstruent may be interpreted as an aspirated obstruent, provided that there exists a minimal pair word starting with an aspirated obstruent, rather than a lenis one. For details of this pattern, see §6.2.1. I will refer to this pattern as 'lexical reanalysis', as listeners taking this strategy are reanalyzing the segmental realization in the signal as conveying an alternative phoneme category, based on the F0 contour.

Another question that this experiment is designed to test concerns whether listeners are biased toward interpreting /kɨman/ + Verb as 'Stop Verb-ing' or 'Start Verb-ing'. In previous studies that conducted prosodic disambiguation experiments, listeners have shown to be biased towards one of the two meanings of the syntactically ambiguous sentence (Mitterer et al., 2021; Yun & Lee, 2022; Farinella & Lee, 2024). Previous research on /kɨman/ has reported mixed results regarding this question (Gim, 2004; Tan, 2023), as will be discussed in §6.2.

To preview the results presented in §6.4, the results suggest there is a significant perceptual role of segmental realization in signaling prosodic structure, again providing evidence that F0 contour is not the only thing that is relevant for the implementation of prosodic structure. In particular, the results supported the second pattern, 'segment prioritization', as a strong realization of lenis obstruent increased the likelihood that listeners interpreted the sentence as meaning 'Start Verbing', consistently for all F0 contour patterns, compared to a weak realization. This finding aligns with the results of Chapter 3 and Chapter 5, which showed that the F0 contour was not very distinct between the two prosodic structures that differed in prosodic phrasing, while the segmental realization was more robustly differentiated. It also supports the prediction that follows from the results of Chapter 2, which argues that lenis obstruent is not optionally affected by the prosodic structure, and the acoustic realization of lenis obstruent therefore may be used as a perceptual cue, contrary to what has been assumed in previous literature (Jun, 1994). The results also suggested that listeners took the third strategy 'lexical reanalysis' only when the F0 contour signaled an AP boundary but not when it did not. As for the second research question, listeners were biased towards understanding /kɨman/ + Verb as 'Stop Verb-ing'.

§6.2 reviews the findings presented in this dissertation so far, which motivate the design of the current perception experiment. It also reviews relevant literature on the interdependency of prosodic structure and phoneme identification. Most of the studies mentioned in §6.2 have already been reviewed in Chapter 1 so I briefly summarize them here, focusing on the role of prosodic structure in phoneme identification and the role of segmental realization in prosodic disambiguation. Previous studies reviewed in Chapter 1 have shown that the identification of a phoneme category is affected by the computation of the prosodic context of the acoustic signal (Kim & Cho, 2013; Steffman, 2019; Steffman et al., 2022) and the interpretation of the sentence is affected by the acoustic realization of the segment within the prosodic context (Scott & Cutler, 1984; Mitterer et al., 2021; Choi & Mazuka, 2001, 2003; Yoo, 2020; Farinella & Lee, 2024). §6.2 also addresses the second research question and presents mixed results in previous studies regarding whether listeners may be biased towards interpreting /kɨman/ + Verb as 'Stop Verb-ing' or 'Start Verb-ing'. §6.3 presents the methods for the experiment. §6.4 presents the results. §6.5 discusses the results, limitations of the current experiment, and future directions.

6.2 Background

The perception experiment in this chapter is motivated by production findings in Seoul Korean presented in the previous chapters of this dissertation. In two corpus studies, I have shown that prosodic structure in Seoul Korean is not just encoded with changes in F0, but also via segmental realization when the syllable onset is a lenis obstruent. In particular, Chapter 2 has shown that the acoustic realization of lenis obstruents in Seoul Korean is prosodically conditioned, and lenis obstruents in the initial position of a PW are always produced with a stronger acoustic realization compared to ones in the medial position. Crucially, it has shown that voicing is only one of the ways that the domain-initial strengthening (or equivalently, domain-medial lenition) is acoustically implemented, and the 'strength' of an acoustic realization may be measured in a higher-dimensional phonetic space, by measuring voicing, degree of reduction, and duration. This result supports the idea that the acoustic realization of lenis obstruents can serve as a reliable cue for listeners in identifying prosodic boundaries.

Chapter 3 has shown the tonal cue, parameterized as the change in F0 within a disyllable window, may be less informative in signalling a prosodic boundary when the second syllable onset is a lenis obstruent, compared to the segmental realization. Following up on this result, Chapter 5 has additionally shown that in a pilot production experiment, there is no measurable difference between two tonal contours that may be prosodically transcribed as 'L +H L+ Ha' and 'L Ha L Ha', when the third syllable starts with a lenis obstruent. Instead, the difference between the two sentences seems to be encoded by the realization of lenis obstruent, i.e., the change in the intensity contour. Findings from these chapters suggest that Seoul Korean listeners may be more influenced by the realization of lenis obstruent than the F0 contour when decoding the acoustic signal into phonological representation, when the F0 contours of two prosodic structures are very similar and potentially nondistinguishable in the acoustic signal (e.g., 'L +H L+ Ha' and 'L Ha L Ha'). However, it still remains to be tested whether listeners are sensitive to a very small difference between the two F0 contours in perception.

As reviewed in Chapter 1, there are two experimental methods to test how listeners decode an acoustic signal into a phonological representation containing prosodic structure and phonemes: a post-boundary segment identification experiment and a prosodic disambiguation experiment. In the following subsections, I will review these methods focusing on the prosodic phrasing of Seoul Korean and how lenis and aspirated contrast is maintained. Then, I will review previous findings relevant for the second research question, i.e., whether listeners may be biased towards interpreting /kɨman/ + Verb as 'Stop Verb-ing' or 'Start Verb-ing'.

6.2.1 Role of prosodic structure in phoneme identification

Previous perceptual experiments show that listeners categorize the same acoustic stimuli as different phonemes depending on how they perceive the prosodic structure (Kim & Cho, 2013; Steffman, 2019; Steffman et al., 2022). In Chapter 1, these experiments have been called the 'contrast-shift experiments'.

To repeat briefly, in a contrast-shift experiment, participants are presented with a segment (e.g., a stop) manipulated along some phonetic dimension (e.g., VOT) and aksed to make a phoneme category judgment, where the context before the segment is manipulated such that in one condition the acoustic signal flags a prosodic boundary while in the other condition, it does not. Participants generally show a significant shift in their perceptual boundary, as a function of the post-segment acoustic manipulation. If the acoustic manipulation signals a prosodic boundary, they in turn expect the post-boundary segment to be realized with the acoustic attributes of a domain-initial segment (e.g., a longer VOT for a voiceless stop in English (Kim & Cho, 2013)).

In the context of the /kɨman/ + Lenis/Aspirated verb used in the current study, there are two conditions summarized in Table 6.1 for which listeners may perceive a Lenis token as Aspirated, taking the Lexical reanalysis strategy.

				Verb	
		kɨ	man	σ	σ
(a)	Tone	L	Ha	H	Ha
(b)	Tone	L	+H	L+	Ha

Table 6.1: Two conditions where listeners may misperceive a verb-initial strong lenis as Aspirated

First, if the F0 contour conveys the tonal sequence 'L Ha H Ha' where the first two tones are associated with /kɨman/, and the verb-initial lenis token is realized as strong (as in an AP-initial position from the kɨman$_{start}$ sentence), it may be perceived as aspirated, if there exists a lexical item that starts with an aspirated obstruent, as shown in (a) in Table 6.1. This prediction is based on results of previous perception studies that have shown Seoul Korean listeners pay attention to F0 in distinguishing lenis and aspirated, rather than VOT, in the AP-initial position (Lee, 2018; Choi et al., 2020). The VOT merger of Lenis and Aspirated stops has likely taken place completely in listeners' grammars if they are born after year 1980 (Bang et al., 2018) and the F0 is likely to be the primary cue for lenis/aspirated contrast. In other words, if a strong lenis obstruent which has a similar VOT to an aspirated obstruent is presented with a high F0, listeners will perceive it as an aspirated obstruent.

Second, if the F0 contour conveys the tonal sequence 'L +H L+ Ha' and the lenis token is realized as strong, it may potentially be perceived as aspirated, if it makes a word, as shown in (b) in Table 6.1. This prediction assumes that Seoul Korean listeners pay attention to VOT in distinguishing lenis and aspirated, rather than F0, in an AP-medial position (Lee, 2018; Choi et al., 2020), where the verb initial syllable, regardless of the laryngeal category of the verb-initial segment, bears the L+ tone. Therefore, if listeners interpret the F0 contour 'L +H L+ Ha' as indicating there is no AP boundary after /kɨman/, they might misperceive a strong lenis obstruent as aspirated, as it is a viable option.

6.2.2 Role of segmental realization in prosodic disambiguation

Another way to test how listeners decode prosodic structure from the acoustic signal is to conduct a 'prosodic disambiguation' experiment (Scott & Cutler, 1984; Mitterer et al., 2021; Choi & Mazuka, 2001, 2003; Yoo, 2020; Yun & Lee, 2022; Farinella & Lee, 2024). In a prosodic disambiguation experiment, listeners are tested on their interpretation of a pair of syntactically ambiguous sentences that is prosodically disambiguated. The same sequence of words is presented with different acoustic realizations that are manipulated to encode different prosodic structures. Multiple acoustic manipulations may be crossed in the experimental design to test which part of the acoustic signal is more influential for listeners' decoding process. I will review relevant studies focusing on the role of lenis obstruent realization in prosodic disambiguation (Choi & Mazuka, 2001, 2003; Yoo, 2020; Yun & Lee, 2022; Farinella & Lee, 2024).

First, Choi & Mazuka (2001) showed that when lenis stop was voiced while F0 indicated an AP boundary, listeners still interpreted the acoustic signal as having an AP boundary 76% of the times (p.183). However, they did not test whether listeners would perceive an AP boundary if lenis was voiceless while F0 did not indicate an AP boundary.

Yoo (2020) more recently tested the role of lenis stop realization when F0 was artificially monotonized. In Yoo's prosodic disambiguation experiment, there was no significant difference in listeners' judgements between the condition where a voiceless lenis stop was spliced into an AP-medial position and the condition where a voiced lenis stop was spliced into an AP-initial position. Both conditions affected participants' interpretation of the sentence in the expected direction, but the degree that they changed the interpretation did not differ significantly (p.238). However, the stimuli in Yoo's study was artificially monotonized for F0 and therefore the results may not reflect what listeners do in natural speech perception.

Yun & Lee (2022) tested how listeners disambiguated a syntactically ambiguous question in Korean. As WH-words in Korean are homophonous with indefinite pronouns (e.g., who and anyone), a WH-word followed by a verb may be ambiguous between a WH question and a Yes-No question. Jun & Oh (1996) argued that the two question types differed in prosodic phrasing of the WH-word and the subsequent verb. In a Yes-No question, the two PWs are prosodically phrased separately into two APs, whereas in a WH question, the two PWs are phrased together into a single AP.

Yun & Lee (2022) argued that their results indicated that listeners interpreted the mismatch between the F0 contour and the expected realization of the verb-initial segment differently, based on the prosodic context. In their experiment, verbs always started with either a nasal or a lenis alveolar stop, both of which have AP-conditioned acoustic realization: denasalized vs. nasal and voiceless vs. lenited. When the F0 contour of the utterance originally produced as a Yes-No question was manipulated such that it did not signal an AP boundary (expected for a WH question) but the segment did (expected for a Yes-No question), listeners' likelihood of interpreting the utterance as a Yes-No question decreased. In contrast, when the F0 contour of the utterance originally produced as a WH question was manipulated such that it did signal an AP boundary (expected for a Yes-No question) but the segment did not (expected for a WH question), listeners' likelihood of interpreting the utterance as a WH question did not change.

Yun & Lee (2022) argued that the reason for the asymmetry between the role of the post-WH L tone may be due to the asymmetry in how listeners interpret the verb-initial segment in the prosodic context. They argued that their results meant that the AP-initial realization of consonants does not necessarily create a percept of an AP boundary, in the absence of an accompanying F0 cue (i.e., no AP boundary in an Yes-No question F0 contour) while the AP-medial realization of consonants may always create a percept of an absence of an AP boundary, even in the presence of an accompanying F0 cue (i.e., an AP boundary in an Yes-No question F0 contour). They further argued that this was expected given that lenis obstruents can optionally remain voiceless in the AP-medial position, while 'voicing in the AP-initial position is strictly prohibited in Korean' (p. 42). Their results are summarized in Table 6.2. Rephrasing their arguments with the aforementioned decoding strategies, they argued that a mismatch between an F0 contour for a no-boundary ($[_{AP}$ ... +H **L+** ...]) and a segmental realization for a boundary (strong lenis) was resolved in factor of the F0 cues. In other words, listeners took the 'F0 prioritization' strategy. In contrast, they argued that a mismatch between an F0 contour for a boundary (Ha $]_{AP}[$ L) and a segmental realization for a no-boundary (weak lenis) is prosodically reanalyzed to indicate no-boundary (i.e., $[_{AP}$ ... +H **L+** ...]). Listeners took the 'Segment prioritization' strategy in this condition.

	... Ha $]_{AP}[$ **L** ...	$[_{AP}$ ... +H **L+** ...]
Strong Lenis	Expected	F0 prioritized, interpreted as $[_{AP}$ ... +H L+ ...]
Weak Lenis	Segment prioritized, interpreted as $[_{AP}$ ... +H L+ ...]	Expected

Table 6.2: Arguments made in Yun & Lee (2022) about the role of segmental realization in prosodic disambiguation. The verb-initial segment, aligned with the L target in bold (L or L+), was a lenis stop. Yun & Lee (2022) argued different decoding strategies were taken, depending on the F0 contour. When the F0 contour signaled an AP boundary (Ha $]_{AP}[$ L), listeners took the 'Segment prioritization' strategy, while when the F0 contour signaled no boundary ($[_{AP}$ ... +H L+ ...]), listeners took the 'F0 prioritization' strategy.

Extending Yoo (2020) and Yun & Lee (2022), Farinella & Lee (2024) tested the role of segmental realization in the prosodic disambiguation of the question type ambiguity. Following Yoo (2020), they spliced the verb-initial syllable preceded by a WH word to create conditions where the acoustic realization of the verb-initial lenis obstruent mismatched with the encompassing F0 contours. They cross-spliced the weak realization of lenis obstruent from the AP medial (i.e., from a WH question) into the AP initial token (i.e., a Yes-No question), and the strong realization of lenis obstruent from the AP initial (i.e., from a Yes-No question) into the AP medial token (i.e., a WH question). They also same-spliced the segments from one question type into another instance of the same question type. They manipulated the stimuli such that other phonetic properties, for instance the duration of the syllables, and the tonal contours, were identical between the cross- and the same-spliced stimuli.

Following Yun & Lee (2022), they hypothesized that given the proposed optionality of lenis obstruent voicing in the AP-medial position, the voiceless realization spliced from the AP-initial position (Yes-No question) and inserted into the AP-medial position (WH question) would not cause any difference in the likelihood of interpreting it as a WH question. On the contrary, the voiced realization spliced from the AP-medial position (WH-question) and inserted into the AP-initial position (Yes-No question) would create a mismatch between the tonal cues suggesting an AP boundary and the segmental realization suggesting an absence of the AP boundary. This mismatch would therefore result in a decreased likelihood of interpreting the question as a Yes-No question.

Their hypotheses were only partially met by their results. They found that both manipulations were effective in decreasing the likelihood of the parsing suggested by the tonal cues. However, this prosodic reanalysis effect was much greater in the latter manipulation (AP-medial segment reversing AP-initial F0 contour), as expected in their hypothesis.

In the current experiment, the expected prosodic difference between the two meanings of /kɨman/ + Verb is similar to the prosodic difference between the two question types. While the question types also differed in the boundary tone on the verb-final syllable (H% for a Yes-No question and LH% for a WH question), the /kɨman/ + Verb construction avoids this confounding factor since the verb is not the IP-final PW and it is followed by a context word /ʧoh-kes*-ta/ ('would be good'). Chapter 5 has shown that the prosodic difference between the two meanings is therefore limited to the presence or absence of an AP boundary between the two PWs.

6.2.3 Listener bias toward a particular meaning given a sequence of words

When we interpret the result of a prosodic disambiguation experiment, we also need to take into consideration that the acoustic signal to the prosodic structure mapping constitutes only one of the steps involved in speech perception.

As discussed in Chapter 1, we assume that the acoustic implementation of the prosodic structure can be optional and probabilistic (Jun, 1993; Arvaniti & Baltazani, 2005), and the mapping between the prosodic structure and the syntactic structure can be a many-to-many function (e.g., Arvaniti, 2011; Im et al., 2023). In a prosodic disambiguation experiment, listeners may be biased towards a particular interpretation of the sentence, regardless of the acoustic signal. For instance, Mitterer et al. (2021) has taken into account in their analysis that listeners may be biased toward a late closure reading compared to an early closure reading, of. a sentence with the form 'Noun1 or Noun2 and Noun3'. Regardless of the acoustic signal, listeners were more likely to interpret the sentence as having a late closure (i.e., (Noun 1 or Noun2) and (Noun3)).

There may be a similar bias in the interpretation of any syntactically ambiguous pair of sentences. In light of this bias, the results of Yun & Lee (2022) and Farinella & Lee (2024) may be reinterpreted, if there is also a syntactic difference between the two types of questions, and there is a bias toward understanding the word sequence 'WH-word + Verb' as a WH question, regardless the acoustic signal. Yun & Lee (2022) in fact reported that their base conditions that model natural F0 contours of each question type were interpreted asymmetrically. While the base condition for a WH question was interpreted as a WH question by their participants 99% of the times, the one for a Yes-No question was interpreted as a Yes-No question by their participants 90% of the times. This bias toward understanding a question as a WH question might have played a role in how the segment mismatch was processed asymmetrically by the participants in these experiments. In both experiments (Yun & Lee (2022) and Farinella & Lee (2024)), listeners seemed to favor interpreting an ambiguous question as a WH question.

We may be able to reinterpret Yun & Lee (2022)'s findings based on the possibility that listeners are overall biased toward understanding the word sequence 'WH-word + Verb' as a WH question. In their experiment, both when a boundary F0 cue (i.e., Ha$]_{AP}$[L) was mismatched with a no-boundary realization of lenis stop (i.e., weak lenis) and when a no-boundary F0 cue (i.e., $[_{AP}$ ... +H L+ ...]) was mismatched with a boundary realization of lenis stop (i.e., strong lenis), listeners were consistently more likely to interpret the sentence as a WH question. Therefore, their results may not necessarily indicate that listeners took different decoding strategies depending on the F0 contour, as they argued.

In the current perception experiment, the second research question asks whether there is a bias for interpreting /kɨman/ + Verb as meaning 'Stop Verbing'. We may find a source of this bias from corpus investigations that count the number of times /kɨman/ is used with the meaning of 'stop' or 'start'. These counts, expressed as probabilities, would indicate the prior probability that the word /kɨman/ is used with a particular meaning, regardless of the acoustic signal. There are two relevant corpus investigations: Gim (2004) and Tan (2023).

The last row of Table 4.1 in §4.2.2 is repeated in Table 6.3. Note that the percentages do not add up to 100% as there were 35 tokens of /kɨman/ whose meaning was not identifiable from the context. Table 6.3 also contains more recent results reported in Tan (2023). Tan (2023) combined the occurences of /kɨman/ collected from various text corpora. Tan reported that there were over 100,000 occurences of the word /kɨman/, and due to practical limitations, only investigated a randomly selected subset of the data which contained 581 occurrences of /kɨman/. In this subset, /kɨman/ meant 'stop' 417 times (71.77%) and meant 'start' 164 times (28.23%).

Source	Stop meaning	Start meaning	Total occurrences
Gim (2004)	206 (18.48%)	874 (78.39%)	1115
Tan (2023)	417 (71.77%)	164 (28.23%)	581

Table 6.3: Counts of /kɨman/ with two meanings identified from corpora in Gim (2004) and Tan (2023)

The two corpus investigations are strikingly different as the proportions of 'stop' and 'start' meanings of /kɨman/ are almost the opposite in the two studies. The source of this difference in the distribution of meanings is unclear because Tan (2023) did not report how the subset was sampled from the larger data. Since two sources report conflicting results, it is unclear whether Korean speakers may prefer to interpret /kɨman/ as 'stop' or 'start', for the ambiguous sequence /kɨman/ + Verb. Without further evidence, the only possible way to interpret the corpus results in these studies may be to conclude that Korean speakers do seem to use /kɨman/ in more than one way, as 18.48% in Gim (2004) and 28.23% in Tan (2023) are both notably higher than 0%. The second research question of this chapter tests whether there is an overall bias toward interpreting /kɨman/ + Verb as 'Stop Verbing' or 'Start Verbing'.

6.3 Methods

Previous results reviewed in §6.2 motivates a perception experiment with a 3X2X2 design. The audio files recorded and investigated for Chapter 5 were manipulated to be used as stimuli for the current perception experiment.

F0	Verb initial seg.	Minimal pair
L Ha L Ha (HaL)	Strong	Available (Asp)
L +H L+ Ha (hl)	Weak	Unavailable (*Asp)
L Ha H Ha (HaH)		

Table 6.4: Experimental conditions

I tested three F0 contours: L +H L+ Ha, L Ha L Ha, and L Ha H Ha, which are intended to signal 'Stop Lenis-Verbing' (e.g., /pul/ 'to blow'), 'Start Lenis/Aspirated Verbing', and 'Start Aspirated-Verbing' (e.g., /p^{h}ul/ 'to solve'), respectively. These F0 contours were crossed with two segmental realization of lenis obstruents in the verb-initial position. The verb-initial syllable with a lenis obstruent onset was spliced from an utterance spoken with the meaning of 'Start Verbing' or 'Stop Verbing'. The acoustic characteristics of the verb-initial lenis obstruent in these two utterance meanings are discussed in detail in Chapter 5. The segmental realization was either same-spliced or cross-spliced, such that audio files in the experimental conditions are all manipulated. Finally, I chose the stimuli verbs such that some verbs had a minimal pair with an Aspirated obstruent verb-initially (e.g., /pul/ ('to blow') or /p^{h}ul/ ('to solve')), while others did not (e.g., /po/ ('to see') or */p^{h}o/) This was expected to play a role in listeners decoding of the acoustic signal as the lexical reanalysis would only be possible for when there was a lexical item available. As a result, there are twelve experimental lists: 3 F0 contours, 2 segmental realizations, and 2 lexical conditions. Table 6.4 summarizes this. The abbreviations in the parentheses will be used to refer to these conditions (e.g., 'HaL', instead of 'L Ha L Ha').

6.3.1 Items

From the list of verb items selected for Chapter 5, 12 Lenis verbs (a verb starting with a Lenis obstruent) were chosen for the Asp condition (minimal pair exists for an Aspirated obstruent verb-initially) and 12 were chosen for the *Asp condition (minimal pair does not exist). As explained in Chapter 5, the verb stems were either monosyllabic or disyllabic. The verbs in Asp and *Asp conditions were chosen such that the number of verbs for each stem-length were evenly divided. The full list of chosen verbs and their meanings are given in Appendix A: List of verbs.

For the *Asp condition, the Lenis verbs were matched with a verb starting with a nasal stop (/m/ or /n/) or a vowel with the same syllable length of the Lenis verb (e.g., /po/ ('to see') vs. /mil/ ('to push')). These Aspirated minimal pairs in the Asp condition and the matched lexical items in the *Asp condition were used as an alternative choice for the lexical question ('which verb did you hear?'). These will be referred to as lexical alternatives.

The lexical alternatives in the *Asp condition could not be minimal pairs of the Lenis verbs because there were no lexical items available in Korean lexicon. Also, making a minimal pair is not possible due to phonotactic reasons because while some lenis verbs in the *Asp condition had a velar place of articulation (e.g., /ka/ 'to go'), /ŋ/ cannot be a syllable onset (*/ŋa/).

6.3.2 Participants

A total of 24 Korean speakers were recruited for this experiment, by word of mouth, and from posts made on online communities. The experiment was IRB approved at UMass Amherst.

Participants were paid at a rate of 6 (or 8700 Korean Won) US dollars per 30 minutes of their participation. They typically used 30 minutes to complete the experiment. If they took longer or shorter than 30 minutes, they were paid according to the rate which was explained at the beginning of the experiment.

Participants were asked where they were born and where they currently live in. Three of them were neither born, nor currently live in Seoul or cities nearby Seoul. They were nonetheless kept in the analysis because they did not differ systematically from the rest of the participants. Since Seoul Korean is a standard dialect used in South Korea, most listeners are familiar with the dialect either from media or from people in their speech community.

The small number of participants and the potential confound of recruiting non-Seoul Korean speakers are both notable limitations of the current experiment, which needs to be addressed in a follow-up experiment in the future.

6.3.3 Procedure

Participants were instructed to use headphones and they did a standard audio test, where they were instructed to choose the quietest sound out of three beeps. All participants passed the audio test.

At the very beginning of the experiment, participants were given the instruction that they would respond to the sentence 'kɨman _____ ʧoh-kes*-ta/' ('It would be great if you could start/stop _____'). They were explained with examples that the sentence was ambiguous between 'Start Verbing' and 'Stop Verbing'. They were however not informed that what disambiguated the meaning was prosody.

The experiment had two task types: a lexical decision task and a sentence comprehension task. Before each task, participants had four practice trials, where the audio was not manipulated, and they were given the correct answer for the question. For the experimental trials, participants were not given any feedback.

In the lexical decision blocks, only the lexical question was asked for each stimuli. The instruction said 'Listen to the audio and choose the word you heard in the position of the blank'. The sentence on the screen was always 'kɨman _____ ʧoh-kes*-ta/'. Two lexical choices appeared in the bottom of the sentence. Lenis verb was always on the left and the lexical alternative was on the right.

In the sentence comprehension blocks, the sentence question was asked for each stimuli. The instruction for the sentence question said 'Listen to the audio and choose the action that the speaker wants'. The sentence on the screen was the same as that for the lexical question: 'kɨman _____ ʧoh-kes*-ta/'. The verb was left blank because for the experimental stimuli, which word was said was also what participants were tested on in the lexical part. The two choices were: 'It is time to proceed with the action' and 'Discontinue it because it is enough'. If the participants chose the former option, they interpreted the sentence as meaning 'Start Verbing' and if they chose the latter, they interpreted it as 'Stop Verbing'.

There were four blocks for each task type and participants were able to take a self-paced rest between blocks. Each block consisted of 12 experimental trials (12 experimental lists) and 20 filler trials. Since there were 12 lexical items chosen for each lexical condition, the same verbs were used in two blocks, once with a strong realization and once with a weak realization. Ideally, each block would have different list of verb items to counterbalance the effect of lexical items, but there were not enough mono/disyllabic verb in Korean lexicon to populate the experimental lists and the number of blocks. Out of 20 filler trials, 12 trials where the audio contained the lexical alternatives. 8 trials contained lexical items that were not the lexical alternatives of the experimental trials. The filler trials were not manipulated acoustically, and therefore could show whether listeners were able to distinguish the two sentence meanings. 12 Verbs (6 Asp and 6 *Asp) were assigned to 6 experimental groups which were counterbalanced using a Latin square design. The order of blocks were also varied by Participant to take into consideration a potential block order effect. Each block had 32 trials and there were 8 (four lexical and four sentence) blocks, resulting in 256 trials in total. As mentioned, the whole experiment typically took 30 minutes to complete.

6.3.4 Acoustic manipulations

This section describes how the audio stimuli were manipulated. Figure 6.1 show what a typical pair of waveforms looked like for the two meanings of /kɨman/. The verb in the example was /pomjɔn/ ('see-COND'). To control for the potential confound of the context word, all stimuli contained the same token of /ʧoh-kes*-ta/ spliced from one of the recordings.

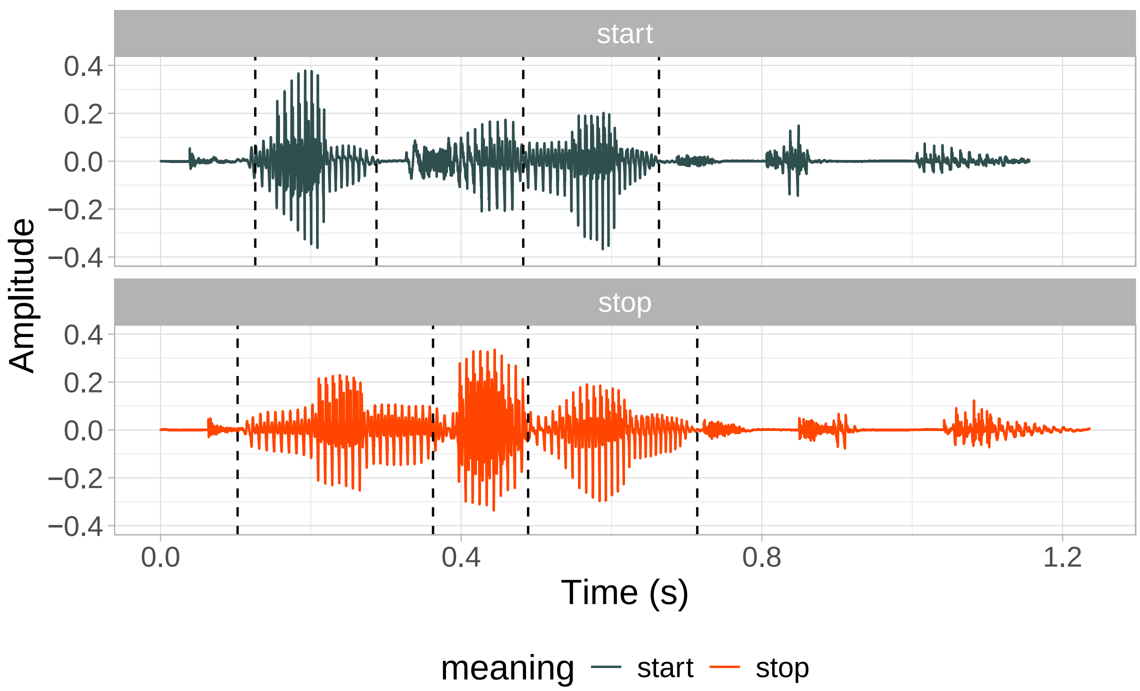


Figure 6.1: Original wav files for /kɨman/ + /pomjɔn/ ('see'). The first four intervals correspond to the four syllables: kɨ.man.po.mjɔn, and the last interval is the context word ʧoh-kes*-ta.

6.3.4.1 Duration normalization

Experimental stimuli were manipulated to match in the syllable duration because syllable durations differed by sentence meaning, as discussed in Chapter 5. The result of duration normalization looked like Figure 6.2. Each interval was either shrunk or elongated by manipulating the Duration tier in Praat (Boersma & Weenink, 2023). The duration measures for /kɨman/ were matched to be the average durations for /kɨ/ and /man/ across stimuli and meanings, which were 100 ms and 215 ms. The verb-final syllable /mjɔn/ also had a fixed duration for all experimental stimuli: 220 ms after a monosyllabic verb stem and 203 ms for a disyllabic verb stem, which were average values for the duration of /mjɔn/ for each verb stem length, across the two meanings. The syllables in the verb stem were adjusted to match the average duration of the two sentence meanings, and therefore varied by each verb stem.

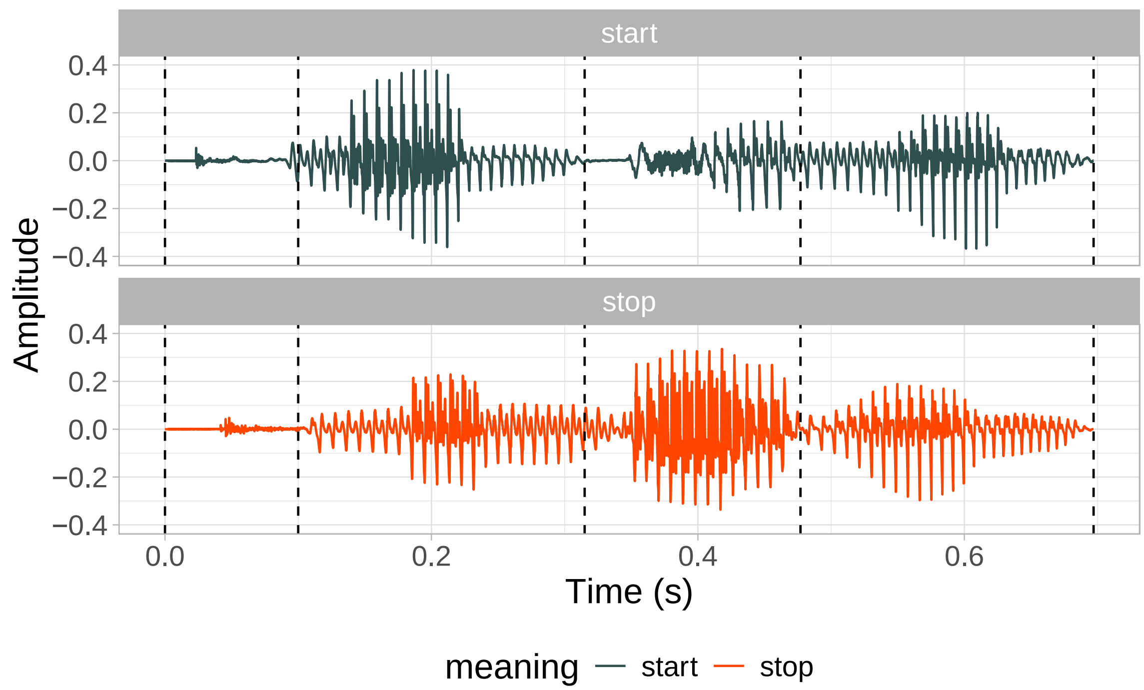


Figure 6.2: Duration normalized version of audio files in Figure 6.1

6.3.4.2 Splicing and F0 manipulation

The verb-initial syllable was spliced to create the mismatch between the F0 contour and the segmental realization. For the unconventional 'L Ha H Ha' contour for Lenis verbs, the same matching audio files were used: Strong realization spliced into an audio originally produced as 'Start Verbing', and Weak realization spliced into an audio originally produced as 'Stop Verbing'.

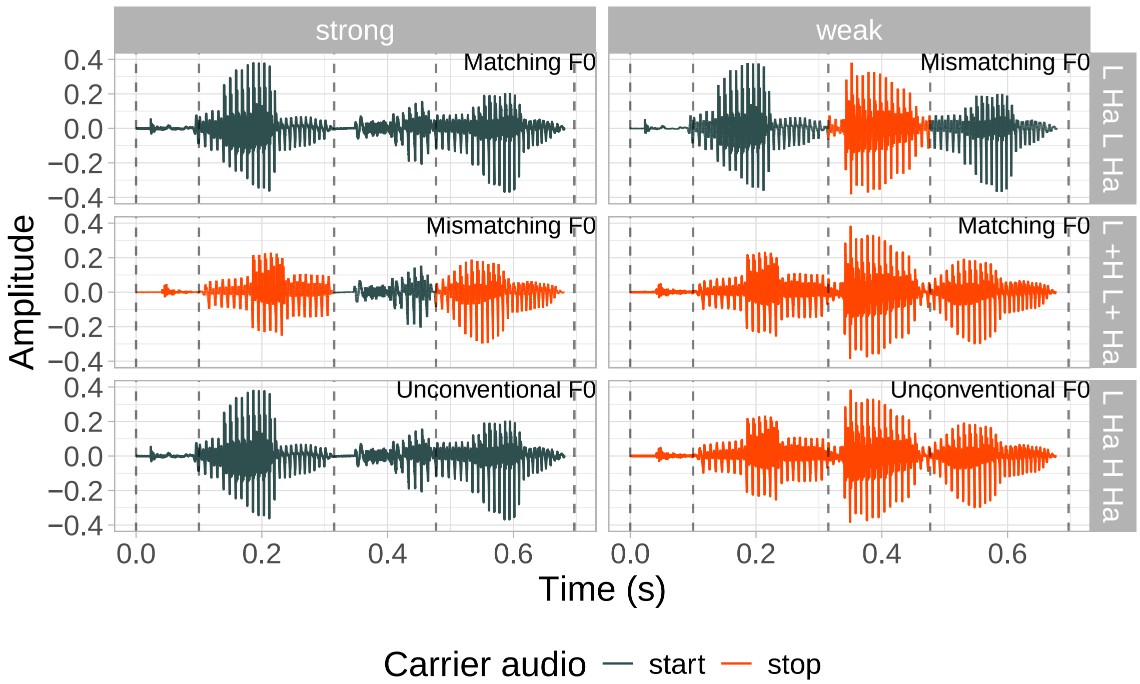


Figure 6.3: Demonstration of how splicing was done on the waveforms of Figure 6.2

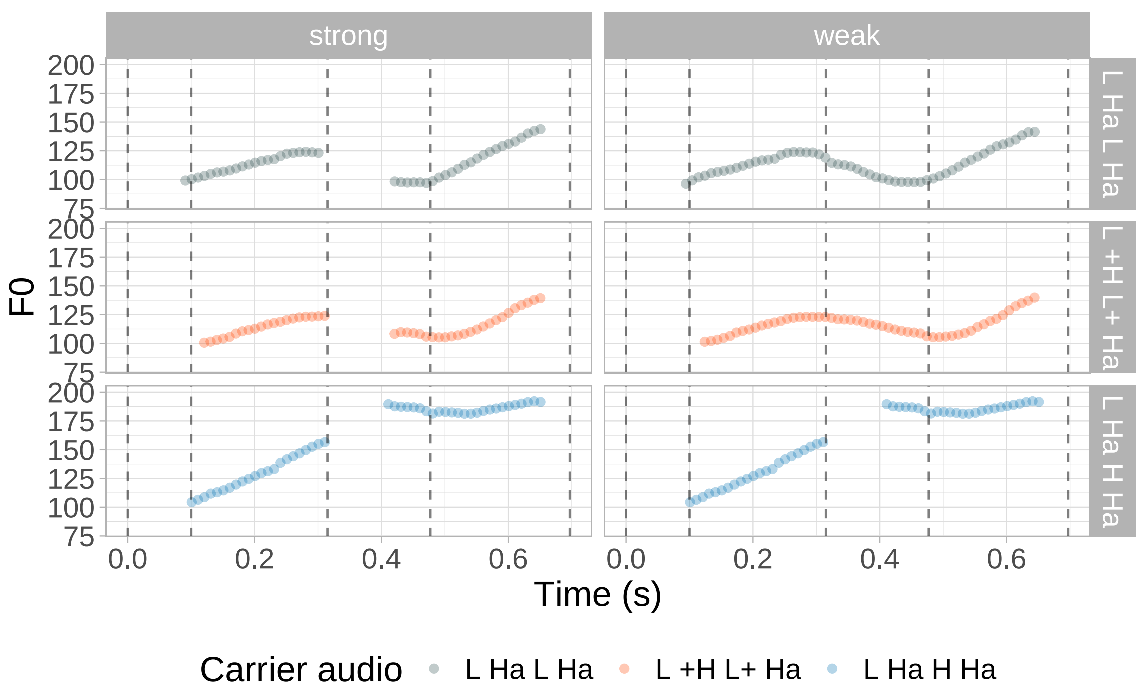


Figure 6.4: Pitch tiers inserted into the waveforms in Figure 6.2

The resulting waveforms for 6 experimental conditions looked like the ones in Figure 6.3. The column headers of Figure 6.3 show the difference between the verb initial syllable (the third interval), and the row headers show the F0 contours that were inserted for each waveform. F0 contours always matched with the amplitude of the syllables surrounding the verb-initial syllable. For instance, the 'L +H L+ Ha' contour was never inserted into an audio originally produced as 'Start Verbing' (Dark grey), and likewise the 'L Ha L Ha' contour was never inserted into an audio originally produced as 'Stop Verbing' (Orange).

F0 contours were defined using an FPCA analysis, which was explained in Chapter 5. Since F0 contours did not differ significantly as a function of sentence meaning when the verb onset was a Lenis obstruent, F0 contours of L-initial filler verbs (nasal or vowel initial) were modeled with FPCA, separately for monosyllabic and disyllabic verbs. F0 contours were defined by reconstructing the F0 contour with the identified first PC, and each F0 contour was generated with 100 points in normalized time and fixed across verb items with the same stem length. The two F0 contours were significantly different, i.e., their PC1 were significantly different by the sentence meaning. To maximize the difference, given the variance in the data, the lower and upper ends of the 95% confidence intervals for PC1 for each sentence meaning was used to reconstruct the two F0 contours: HaL and hl. Finally, the HaH F0 contour was parameterized using the FPCA modeled for Aspirated verbs.

One hundred points in normalized time were inserted into the PitchTier in Praat for each duration normalized verb items. The audio was then resynthesized using the Praat function 'Get resynthesis (overlap-add)'. Figure 6.4 plots the F0 contours that were resynthesized together with the six waveforms of the verb /pomjɔn/ plotted in Figure 6.3.

6.4 Results

This section reports the results of the experiment. The main research question that the experiment is designed to test is how listeners interpret the mismatch between the F0 contour and the segmental realization of the verb-initial lenis obstruent. As for the first question, three possible strategies were proposed in §6.1 to resolve the mismatch: 'F0 prioritization', 'Segment prioritization', and 'Lexical reanalysis'. If participants choose a lexical alternative, especially in Asp condition, it would indicate that they have made a lexical reanalysis decision. If participants differ by the segmental realization or a particular F0 contour in interpreting the utterance as 'Start Verbing', it may indicate that they have made a decision based on the segmental realization, or the F0 contour. In addition, a secondary question that the experiment is designed to investigate is whether there is a bias toward interpreting /kɨman/ + Verb as 'Stop Verbing' or 'Start Verbing', which has been left unresolved in previous corpus investigations, as discussed in §6.2.3.

The result of the experiment had 2304 trials (24 participants * 12 trials * 8 blocks). Participants took longer time in general to respond to a sentence question, but any trial that took longer than 10 seconds were excluded. Any trial that took less than 10 ms were also excluded as it was clear that they were not reacting to the question, but possibly misclicked the keyboard.

6.4.1 Part 1: Lexical identification

The results of Part 1 Lexical identification question are presented in Figure 6.5. As a reminder, these were participants' responses to the question 'which verb did you hear in the blank?'. Panels are divided by three F0 conditions (hl, HaL, and HaH) and two lexical conditions (Asp and *Asp). In each panel, the bars show the average proportion of Lenis response for each of the segmental conditions (Weak and Strong).

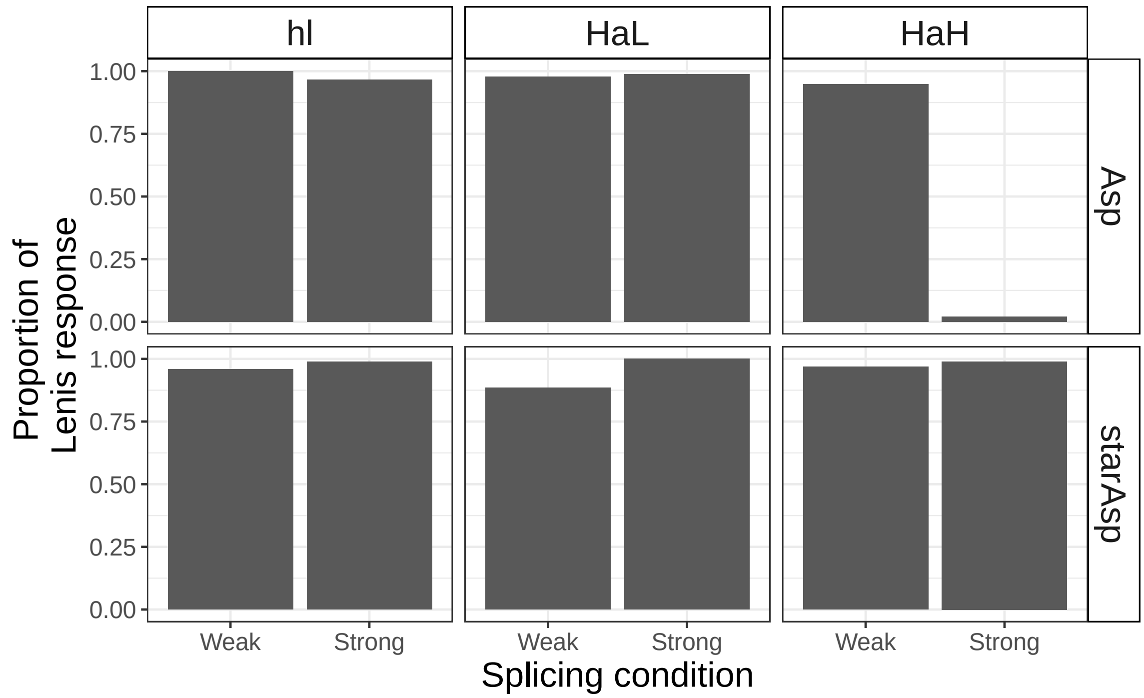


Figure 6.5: Result for the lexical question

As Figure 6.5 shows, participants almost always chose 'Lenis' as the answer for all conditions, except when the F0 contour was 'L Ha H Ha' and the segmental realization was 'Strong', and the verb had a minimal pair starting with an Aspirated obstruent. Since it was a binary forced choice question, this means that they perceived Lenis obstruent as Aspirated in this condition.

While it was predicted that participants might perceive Lenis as Aspirated also in 'hl, Strong, Asp' condition (the right bar of the leftmost top panel), they nearly categorically selected 'Lenis' as response. This will be discussed further in the discussion in §6.5.

6.4.2 Part 2: Sentence interpretation

The results of Part 2 Sentence interpretation question are presented in Figure 6.6. As a reminder, these were participants' responses to the question 'Choose the action that the speaker wants'. The two possible responses were 'It is time to proceed with the action' and 'Discontinue it because it is enough'. Panels are divided in the same way as in Figure 6.5. In each panel, the bars show the average proportion of 'Start' response for each of the segmental conditions (Weak and Strong).

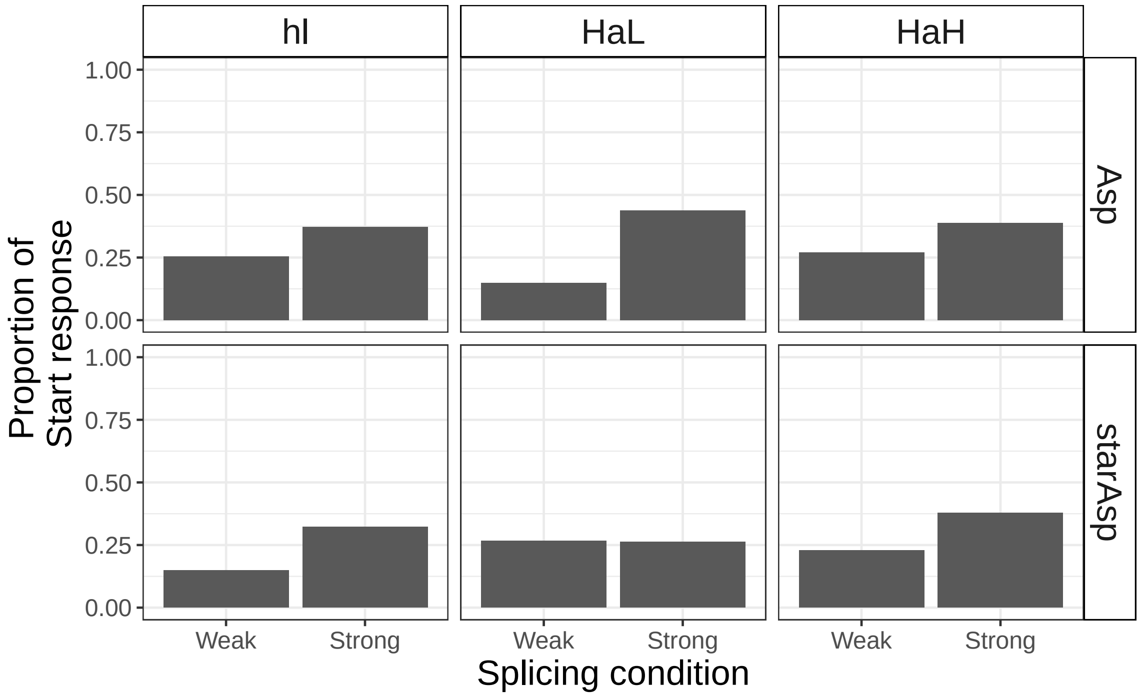


Figure 6.6: Result for the sentence question

While the bar heights differed by panels, there is a notable pattern consistent across panels. Except for the 'HaL, *Asp' panel, participants answered 'Start' more often when the Lenis obstruent was strong, compared to when it was weak.

It is also clear that even the highest bar (HaL, Strong, Asp) only reached slightly above 0.4, meaning participants were biased strongly towards selecting 'Stop' as a response.

6.4.3 Statistical analysis

To further test the effects of F0, Lexical, and Segmental conditions on the sentence interpretation, a mixed effects logistic regression model was fitted. The model was fitted using the '`lme4`' pakage in R (Bates et al., 2015). The p-value were calculated using the '`lmerTest`' package (Kuznetsova et al., 2017). I used an alpha level of .05 for all statistical tests.

Lexical and segmental conditions were sum-coded and F0 conditions were coded with two theoretically-guided Helmert contrasts: one to test the difference between 'L +H L+ Ha' and the two F0 contours with an AP boundary ('L Ha L Ha' and 'L Ha H Ha'), the other to test the difference between the latter two F0 contours. The contrast codings of predictors are summarized in (3).

(3) Contrast coding of predictors

a. Lexical condition: `Lex`

+1/2 (*Asp), -1/2 (Asp)

b. Segmental condition: `Seg`

+1/2 (Strong), -1/2(Weak)

c. F0 condition: `F0`

(i) No boundary vs. Boundary

+2/3 (hl), -1/3 (HaL), -1/3 (HaH)

(ii) Conventional vs. unconventional

+1/2 (HaL), -1/2 (HaH)

In the model, these three predictors were included with all possible interaction terms (i.e., `Lex*Seg*F0`). The response variable `isStart` was a binary variable with a value of 1 if participant's response was 'Start' and otherwise 0.

A list of mixed effects logistic regression models that varied in the complexity of random effects were fitted with the formulas in (4). `Participant` and `Verb` were included to code individual-specific and item-specific differences, `ExperimentalGroup` was included to code random variation due to the mapping of verb items to experimental conditions, `stem length` was included to code the variation due to the verb stem being mono or disyllabic in length. Despite having a small number of levels, `stem length` was included in the random structure rather than a fixed effect, as it was not part of the experimental design that was manipulated. I present a version of the model with `stem length` as a fixed effect in Appendix B: Alternative model. The results of this alternative model do not seem to differ in any meaningful way from the main model results.

(4) Random effect structures tested for model

a. `isStart ~ Lex*F0*Seg`
`+ (1|Participant) + (1|Verb)`
`+ (1|ExperimentalGroup) + (1|Stem length)`

b. `isStart ~ Lex*F0*Seg`
`+ (1|Participant) + (1|Verb)`
`+ (1|ExperimentalGroup)`

c. `isStart ~ Lex*F0*Seg`
`+ (1|Participant) + (1|Verb)`

d. `isStart ~ Lex*F0*Seg`
`+ (1|Participant)`

Model	AIC	BIC	Log-likelihood	p value
(4a)	1295.7	1376.3	-631.86	
(4b)	1293.7	1369.3	-631.86	1
(4c)	1291.9	1362.4	-631.94	0.70
(4d)	1305.2	1370.7	-639.60	$p < 0.001$

Table 6.5: Results of likelihood ratio test analyses between each pair of adjacent models in (4). The selected model was (4c).

Each subsequent model was a version of the previous model with one less random effect term. The order of dropping the random effect term was decided based on the size of the variance explained by each random effect in the first model in (4a). For each pair of adjacent models, the effect of dropping a random intercept term was tested by performing a likelihood ratio test between the original and the simpler version. Table 6.5 summarizes the results of these analyses. The p value column shows whether the model fit significantly changed by dropping a random effect term. If it did not, then it was deemed that the term that was dropped was not justified by the variation in the data.

Only the difference between the model in (4c) and (4d) was significant, meaning there was a significant effect of including the item-specific random variation in the model fit. Consequently, the model in (4c) was selected. I report the result of the model in (4c) in Table 6.6.

	Dependent variable:
	isStart Coefficient (Standard Error)
(Intercept)	-1.079^{***} (0.188)
Seg (+ Strong)	0.798^{***} (0.144)
Lex (+ *Asp)	−0.231 (0.233)
F0	
+ hl vs. - HaL, HaH	−0.149 (0.154)
+ HaL vs. - HaH	−0.250 (0.174)
Interactions	
Lex : hl vs. HaL, HaH	−0.369 (0.307)
Lex : HaL vs. HaH	0.052 (0.348)
Lex : Seg	−0.302 (0.286)
Seg : hl vs. HaL, HaH	0.101 (0.306)
Seg : HaL vs. HaH	0.121 (0.347)
Lex : Seg : hl vs. HaL, HaH	1.205^{**} (0.613)
Lex : Seg : HaL vs. HaH	-1.853^{***} (0.695)
Random effects	
Groups:	Variance (Standard Deviation)
Verb: (Intercept)	0.2023 (0.4498)
Participant: (Intercept)	0.5030 (0.7092)
Observations	1,137
Log Likelihood	−631.937
Akaike Inf. Crit.	1,291.873
Bayesian Inf. Crit.	1,362.379
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 6.6: Result of Model in (4c).

The model indicated that only three effects are significant: the intercept, the strong realization of verb-initial lenis obstruent, and the three-way interactions of the predictors.

The significance of the intercept with a negative coefficient suggests that participants were strongly biased against responding with 'Start'. The significance of the term `Seg` with a positive coefficient suggests that the likelihood of responding with 'Start' increased when the Lenis realization was Strong. Importantly, there was no effect of F0 manipulation, nor the interaction between F0 and Segmental manipulations, which strongly suggest that participants either did not perceive or did not make use of the difference between the F0 manipulations, which were again statistically different in the production experiment (Chapter 5) and even maximized to the extent that was allowed by the variation of the data in the creation of stimuli. This indicates that participants prioritized the segmental realization over the F0 contour in interpreting the utterance as 'Start Verbing' or 'Stop Verbing'.

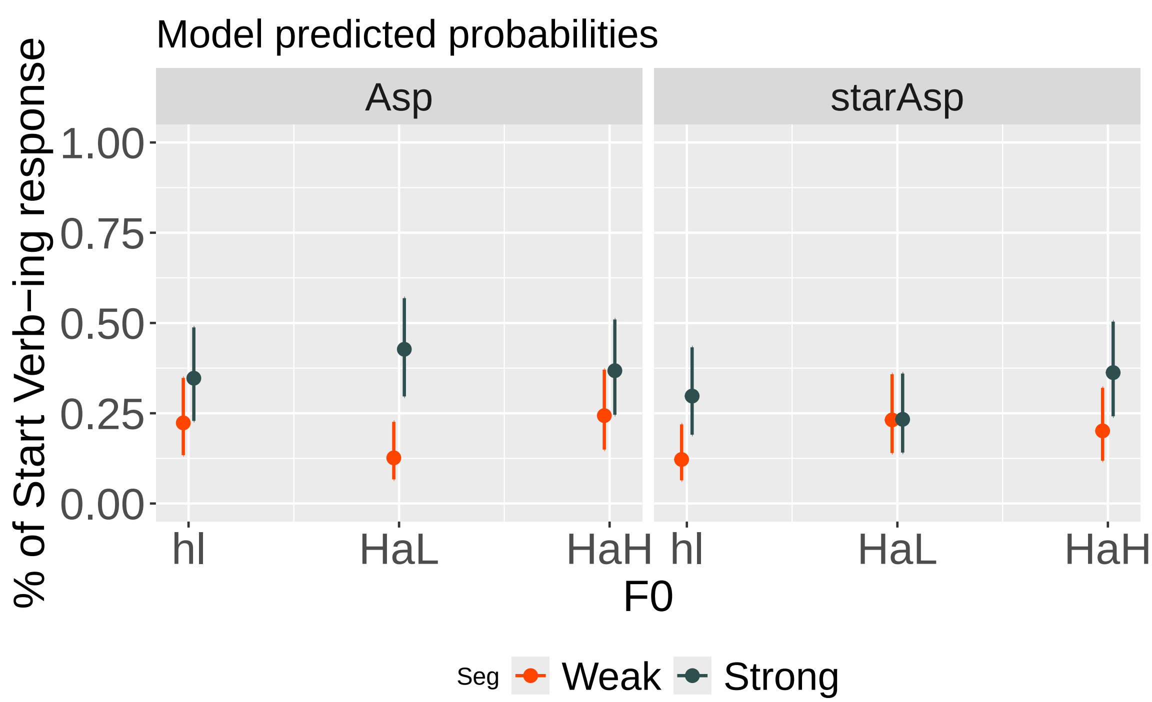


Figure 6.7: Three-way interaction of `Lex:F0:Seg` from the model, on a percentage scale.

The significance of the three-way interactions are easier to interpret with Figure 6.7 plotting the marginal effects for model terms on the percentage scale. The interaction plot was made with the '`plot_model`' function available in the `sjPlot` pakcage (Lüdecke, 2023). The whiskers show the 95% confidence intervals. The three-way interactions were significant because the effect of `Seg` was not significant only in the '*Asp, HaL' condition, as we've previously seen also in Figure 6.6.

6.5 Discussion

This section discusses the findings of the experiment. §6.5.1 provides a summary of results, and discusses limitations of the current experiment and future directions. §6.5.2 discusses the finding that there is a bias toward interpreting /kɨman/ + Verb as meaning 'Stop Verbing'.

6.5.1 Summary of findings

There were two main findings in the current study. First, a lenis obstruent token with a rising F0 contour ('L Ha H Ha') was categorically perceived as an Aspirated obstruent. Second, a strong realization of lenis obstruent increased the likelihood of interpreting the utterance as meaning 'Start Verb-ing'.

The findings suggest that F0 contours were not the primary phonetic correlate, at least in the context of the /kɨman/ + Verb construction under investigation: a disyllable PW (kɨman) followed by another disyllable (or trisyllable when the verb stem was disyllabic), with the second PW starting with a lenis/aspirated obstruent. Instead, listeners only used the segmental realization to differentiate the prosodic structures (i.e., Segment prioritization). This supports the findings of Chapter 3 and Chapter 5, which reported that AP phrasing is acoustically implemented primarily as variation in the realization of lenis obstruents, rather than via the F0 contour.

Alternatively, it is still possible that listeners may not have perceived the difference between the F0 contours, despite the effort of maximizing the difference. In the future, there needs to be a follow-up production experiment testing whether the F0 contours are different in a larger number of participants, and a perception experiment testing whether the differenece of F0 contours is perceivable.

An unexpected finding was that the Strong Lenis obstruent was only perceived as Aspirated when the F0 was rising from the end of /kɨman/, though it was also hypothesized that at least some of the times, it could be perceived as Aspirated, when the F0 contour mismatched with the strong realization. This could be due to the design of the experiment where listeners were asked to pay attention to the verb, whose position was always fixed. This could have made listeners to discard any contextual information about the encompassing F0 contour, but only focus on the realization of the verb itself. Also, while I hypothesized that the decoding of acoustic signal results in a phonological representation containing both the prosodic structure and the segments therein, the experiment tested these two computations separately. A future experiment could have a quaternary forced choice design, where listeners are given all four options: Start/Stop lenis/aspirated. More online experimental methods, such as eye-tracking, may be used to test the effect of prosodic structure in real-time processing of acoustic signal, as in Kim et al. (2018).

Another unexpected finding was that listeners did not differ in their response as a function of the segmental realization when the F0 contour was 'L Ha L Ha', only in the *Asp condition. This was in fact the condition where listeners were predicted to choose 'Start' response most frequently because the segmental realization was expected given the F0 contour. I leave for future studies to explore whether this is a real effect, or a random noise in the current experiment.

6.5.2 Bias toward 'Stop Verbing' interpretation

Finally, there seemed to be an overall bias toward understanding /kɨman/ + Verb as meaning 'Stop Verb-ing'. This is in line with the findings of Tan (2023), rather than Gim (2004).

The bias toward the 'Stop Verb-ing' interpretation seems similar to the bias observed in the interpretation of syntactically ambiguous questions in Korean (Jun & Oh, 1996; Yun & Lee, 2022; Farinella & Lee, 2024), which was discussed in §6.2.3. In both ambiguities, listeners showed a bias toward the interpretation that is compatible with a prosodic phrasing where the ambiguous PW (/kɨman/ or WH-words) and the following verb are prosodically phrased together (e.g., Stop Verb-ing, WH question).

In order to further investigate the bias, there needs to be a study that investigates what derives the difference in prosodic phrasing between two question types. In the case of /kɨman/ + Verb, I have argued in Chapter 4 that the difference in prosodic phrasing is due to the difference in the information structure. It is possible that the same explanation may be applied to the WH-questions, but it is also possible that the difference in prosodic phrasing is due to other factors, such as the syntactic structure of the two question types.

It is worth noting that in Japanese, WH-words are argued to be so-called 'focus-bearing items', which inherently bear contrastive focus (e.g., Ishihara, 2002; Yamashita, 2008). It has been argued that this inherent focus-bearing property of WH-words causes following PWs to be deaccented in Japanese (e.g., Ishihara, 2002; Yamashita, 2008). It is possible that the same may be true for Korean WH-words and /kɨman/, which would explain why they are prosodically phrased together with the following verb.

In both WH/Yes-no questions and /kɨman/ + Verb, the ambiguous word (WH-word or /kɨman/) has two meanings: one where the word is a focus-bearing item and the other where it is not. When they are interpreted as focus-bearing, the WH-word

+ Verb construction is interpreted as a WH-question, rather than a Yes-no question, and /kɨman/ + Verb is interpreted as 'Stop Verbing', rather than 'Start Verbing'. Put this way, the bias toward WH-question and 'Stop Verbing' interpretations may be due to a general tendency to interpret an ambiguous word (/kɨman/ or WH-word) with a focus-bearing meaning. It remains to be investigated how this general tendency may arise in the grammar.

6.6 Appendix A: List of verbs

Verb (meaning)	Lexical alternative (meaning)	Lexical condition
po (see)	mil (push)	*Asp
te (touch)	nal (fly)	*Asp
tol (turn)	mol (drive)	*Asp
ka (go)	al (get to know)	*Asp
ʧol (doze)	nol (play)	*Asp
pewu (learn)	moɨ (gather)	*Asp
puɨ (pour)	muɾɨ (delay)	*Asp
pɔɾi (throw away)	milli (get pushed)	*Asp
tamɨ (put in)	namɨ (stay)	*Asp
tani (commute)	nalli (blow away)	*Asp
kaɾi (cover)	alli (inform)	*Asp
ʧaɾa (grow)	olli (put up)	*Asp
pul (blow)	pʰul (solve)	Asp
ki (crawl)	kʰi (turn on)	Asp
piwu (empty)	pʰiwu (light up)	Asp
tewu (heat)	tʰewu (burn)	Asp
ʧiwu (erase)	ʧʰiwu (get rid of)	Asp
pi (get cut)	pʰi (unfold)	Asp
ke (fold)	kʰe (dig up)	Asp
ʧa (sleep)	ʧʰa (kick)	Asp
puɾɨ (call)	pʰuɾɨ (become green)	Asp
ʧiɾɨ (raise)	ʧʰiɾɨ (host an event)	Asp
ʧewu (put someone to sleep)	ʧʰewu (fill up)	Asp

Table 6.7: List of verbs

6.7 Appendix B: Alternative model

	Dependent variable:
	isStart
Lex+starAsp	−0.213 (0.230)
F0hl.HaL	−0.149 (0.154)
F0HaL.HaH	−0.249 (0.174)
Seg+Strong	0.800*** (0.144)
stem.length+Di	−0.207 (0.230)
Lex+starAsp:F0hl.HaL	−0.374 (0.308)
Lex+starAsp:F0HaL.HaH	0.054 (0.349)
Lex+starAsp:Seg+Strong	−0.294 (0.288)
F0hl.HaL:Seg+Strong	0.101 (0.306)
F0HaL.HaH:Seg+Strong	0.120 (0.347)
F0hl.HaL:stem.length+Di	0.062 (0.304)
F0HaL.HaH:stem.length+Di	−0.047 (0.346)
Seg+Strong:stem.length+Di	−0.105 (0.285)
Lex+starAsp:F0hl.HaL:Seg+Strong	1.207** (0.613)
Lex+starAsp:F0HaL.HaH:Seg+Strong	−1.858*** (0.695)
Constant	−1.070*** (0.187)
Observations	1,137
Log Likelihood	−631.413
Akaike Inf. Crit.	1,298.826
Bayesian Inf. Crit.	1,389.477
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 6.8: An alternative model with `stem.length` as a fixed effect. The results do not differ in any meaningful way from the main model results presented in Table 6.6. The only significant effect is the segmental realization, and the interaction terms involving `Seg`.

CHAPTER 7

CONCLUSION

This chapter concludes the dissertation. In §7.1, I summarize the main findings and discuss future directions for research on phonetic correlates of prosodic structure. In §7.2, I discuss another line of future research, focusing on the syntax-prosody mapping in Seoul Korean.

7.1 Summary of findings

In this dissertation, I investigated the role of segmental realization in signaling prosodic structure, focusing on the realization of lenis and aspirated obstruents in Seoul Korean.

I provided empirical evidence that could challenge two existing claims about how prosodic structure is acoustically implemented in Seoul Korean. These two claims are as follows. First, while lenis obstruents are known to be realized differently depending on the prosodic context, they have also been reported to be only optionally affected by the prosodic context (Jun, 1993, 1994; Han, 2000). Second, the optionality of segmental realization has been one of the reasons why prosodic constituents have been primarily identified and defined based on F0 contours in previous studies (Jun, 1993, 2000). It has been claimed that segmental processes are not as reliable as F0 contours in signaling prosodic structure, and hence they are expected to play a smaller role than F0 cues in speech perception as well (Jun, 1994; Arvaniti & Baltazani, 2005).

This dissertation presents findings that suggest that there may be a greater role for segmental realization in signaling prosodic structure than previously thought.

In Chapter 2, I investigated how lenis obstruents are realized in PW-medial position in spontaneous speech, using a corpus of Seoul Korean (Yun et al., 2015). I found that the realization of lenis obstruents was systematically affected by the prosodic context. While there were fully voiceless lenis obstruents in the PW-medial position, these exceptional tokens were nonetheless found to be more reduced and shorter than fully voiceless lenis obstruents in the PW-initial position.

In Chapter 3, I presented empirical evidence from the same spontaneous speech corpus (Yun et al., 2015) that the AP boundary before a PW starting with a lenis obstruent was more reliably signaled via a change in intensity (i.e., phonetic reduction), than via the F0 contour (i.e., tonal fall). This finding runs counter to the assumption that prosodic structure is primarily signaled via the F0 contours (Jun, 1993, 2000), and that segmental realization is only optionally affected by prosodic structure (Jun, 1994). This finding was also replicated in Chapter 5 with a production experiment.

Unlike for lenis obstruents, I found that the AP boundary before a PW starting with an aspirated obstruent is more reliably signaled via a change in the F0 contour, rather than via a change in intensity. This shows that the primary phonetic correlate of prosodic structure can differ by the segment type. Therefore, the acoustic signal produced by speakers simultaneously contains information about prosodic structure that is conditioned on segmental identity and vice versa in Seoul Korean.

In these investigations, I used a data-driven approach to identify the phonetic correlate of prosodic structure, rather than relying on prosodic transcription. In Chapter 3, I used an unsupervised clustering method to identify the number of prosodic categories justified by the variation in the acoustic data. In Chapter 5, I used Functional Principal Component Analysis (FPCA) to identify the shape components of F0 and intensity contours that are affected by prosodic structure. These unsupervised methods allowed for a more objective analysis of the data, revealing the fact that F0 contours may not always be the most reliable cue for prosodic structure. These methods are potentially useful in further investigations of phonetic correlates of prosodic structure in other languages as well, as they do not require costly manual prosodic transcription. They also open up new possibilities to investigate prosodic structure in large speech corpora that are not prosodically transcribed. Making it possible to study prosody in already existing large corpora allows us to investigate how prosodic structure is acoustically implemented across a larger diversity of languages with more power. This increased power can also make it possible to study how individuals may vary in their use of prosodic cues. In particular, it might be interesting to see whether intensity is a reliable correlate of prosodic phrasing in other languages as well, as intensity has been relatively understudied compared to F0 and duration (cf. Harris & Urua, 2001; Kingston, 2008; Ennever et al., 2017; Katz, 2016; Katz & Pitzanti, 2019; Wagner & McAuliffe, 2019).

Finally, the findings that the realization of lenis obstruent is a reliable cue for prosodic structure were supported by the results of a perception experiment in Chapter 6. The results suggested that listeners were more affected by the segmental realization of lenis obstruents than by F0 contours in making decisions about prosodic structure. The results of the perception experiment also call for further investigation of how listeners may be biased towards interpreting an ambiguous sequence of words as having one prosodic structure over another.

7.2 Future directions in the research on syntax-prosody mapping in Seoul Korean

This dissertation also investigated how multiple prosodic phrasings may be acceptable for a given syntactic structure in Seoul Korean. In Chapter 4, I presented a MaxEnt analysis (Goldwater & Johnson, 2003) to account for empirically observed variation in prosodic phrasing, under broad and contrastive focus, in Seoul Korean

(Cho, 1990; Jun, 1993). Under this framework, I showed how systematic differences in prosodic phrasing of /kɨman/ + Verb could be modeled.

While modeling variation in prosodic phrasing was not the central focus of the dissertation, the kind of analysis I demonstrated for /kɨman/ + Verb opens up several avenues for future research on the syntax-prosody mapping in Seoul Korean. In this section, I discuss three main directions for future research, which may be pursued with both production and perception experiments.

First, it is important to investigate how much variation there is in prosodic phrasing for a given syntactic structure in Seoul Korean. In Chapter 4, my analysis was based on annecdotal reports in previous work that multiple prosodic phrasings are possible for a given syntactic structure (Cho, 1990; Jun, 1993). I chose a set of constraint weights specifically to model this variation. However, in future work, what the distribution of prosodic phrasing actually is for a given syntactic structure needs to be empirically tested. In particular, I propose to conduct more acceptability judgment experiments to test how acceptable different prosodic phrasings are for a given syntactic structure, in addition to standard production experiments where speakers are asked to produce sentences with the most natural prosodic phrasing. In other corners of phonology, we often find that speakers/listeners exhibit gradient acceptability judgments for both attested and unattested linguistic forms (e.g., Albright & Hayes, 2003) and we model how this variation arises in the grammar. In most research on syntax-prosody mapping, however, it is a standard assumption that there is a single best prosodic phrasing for a given syntactic structure (e.g., Selkirk, 1984; Truckenbrodt, 1999; Kim, 2015). Even the studies that report variation in prosodic phrasing (e.g., Jun, 1993; Truckenbrodt, 2002; Lee, 2022) do not discuss whether there is gradient acceptability among the attested phrasings and/or unattested phrasings. In order to test this, we would need to create experimental items where acoustic cues are manipulated to vary prosodic phrasing. Multiple phonetic correlates could be investigated, one of which could be the prosodically conditioned segmental realization.

Second, it would be interesting to consider whether there is phonetic evidence for recursive prosodic structures in Seoul Korean. In Chapter 4, I hypothesized that there may be recursive APs in the prosodic structure of /kɨman/ + Verb, based on the proposed difference in syntactic structure of the two meanings of /kɨman/ + Verb. After laying out linking hypotheses for how recursive phonological structure might be phonetically implemented, I tested this hypothesis in the production data in Chapter 5 and tentatively concluded that there was no clear phonetic evidence for recursive structure, for the /kɨman/ + Verb construction.

However, it seems that the idea that prosodic structures can be recursive has not been widely considered in the literature on prosodic structure in Seoul Korean (cf. Baek & Yun, 2018; Lee, 2022), and there needs to be more work done to investigate this possibility. One potential place to explore first is where there seems to be a phonetic cue mismatch in the acoustic signal in natural productions, that is, when there is an AP-final Ha tone followed by a lenis obstruent that is lenited. Under the standard assumption that a lenis obstruent is optionally conditioned by prosodic structure (Jun, 1993), this cue mismatch would have been prosodically transcribed as containing an AP boundary before the lenis obstruent (i.e., (e.g., [...Ha]$_{AP}$[Lenited Lenis ...]$_{AP}$)). However, a recursive prosodic structure (e.g., [[...Ha]$_{AP}$ Lenited Lenis ...]$_{AP}$) could potentially account for this cue mismatch since the lenited lenis obstruent token here is in the medial position of a higher AP. In order to investigate this further, we would need to conduct a controlled production experiment, where we elicit productions of sentences in Seoul Korean that we expect to be realized with recursive prosodic structures, under the assumptions of syntax-prosody mapping theories (e.g., Match theory (Selkirk, 2011)). In addition, there need to be experimental sentences, where we do not expect to see prosodic recursion. These two sets of sentences could then be compared acoustically to identify potential differences in recursive prosodic structures and non-recursive ones in Seoul Korean.

Finally, a promising direction for future research is to investigate how listeners interpret the acoustic signal in real time in comprehending syntactically ambiguous sentences. In the model of perception proposed by Cho et al. (2007), listeners use acoustic cues to compute prosodic structure and identify segmental contrasts simultaneously. There need to be more steps involved in this process for a listener to interpret the meaning of an utterance. Given that there exists variation in prosodic phrasing for a given syntactic structure, there may also be variation in how a single prosodic phrasing may be compatible with multiple syntactic structures. In addition, among multiple syntactic structures, some may be more likely than others. For instance, in Chapter 6, I discussed that the focus-bearing meaning of /kɨman/ and WH-words in Seoul Korean may lead to a bias in interpreting how ambiguous syntactic structure is interpreted. Considering all these factors together, listeners need to make a joint inference about what prosodic structure (and what phonemes) the acoustic signal is signaling, and what syntactic structure the perceived prosodic structure is compatible with, which is possibly influenced by what syntactic structure is a priori more likely. This process could be investigated with on-line measures of comprehension, such as eye-tracking or EEG, where we can see how and when listeners update and possibly revise their interpretations, as they receive more acoustic input.

BIBLIOGRAPHY

Albright, Adam & Bruce Hayes. 2003. Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition* 90(2). 119–161. doi:10.1016/S0010-0277(03)00146-X. https://linkinghub.elsevier.com/retrieve/pii/S001002770300146X.

Arvaniti, Amalia. 2011. The Representation of Intonation. In Marc Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell Companion to Phonology*, 1–24. Wiley 1st edn. doi:10.1002/9781444335262.wbctp0032. https://onlinelibrary.wiley.com/doi/10.1002/9781444335262.wbctp0032.

Arvaniti, Amalia & Mary Baltazani. 2005. Intonational Analysis and Prosodic Annotation of Greek Spoken Corpora. In Sun-Ah Jun (ed.), *Prosodic Typology*, 84–117. Oxford University PressOxford 1st edn. doi:10.1093/acprof:oso/9780199249633.003.0004. https://academic.oup.com/book/4251/chapter/146102254.

Aston, John A. D., Jeng-Min Chiou & Jonathan P. Evans. 2010. Linguistic Pitch Analysis Using Functional Principal Component Mixed Effect Models. *Journal of the Royal Statistical Society Series C: Applied Statistics* 59(2). 297–317. doi:10.1111/j.1467-9876.2009.00689.x. https://academic.oup.com/jrsssc/article/59/2/297/7079214.

Baek, Hyunah & Jiwon Yun. 2018. Prosodic disambiguation of syntactically ambiguous phrases in Korean. *MIT Working Papers in Linguistics* 88. 89–100.

Bang, Hye-Young, Morgan Sonderegger, Yoonjung Kang, Meghan Clayards & Tae-Jin Yoon. 2018. The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics* 66. 120–144. doi:10.1016/j.wocn.2017.09.005. https://linkinghub.elsevier.com/retrieve/pii/S0095447017302188.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67(1). doi:10.18637/jss.v067.i01. http://dx.doi.org/10.18637/jss.v067.i01.

Beckman, Mary E. 1996. The Parsing of Prosody. *Language and Cognitive Processes* 11(1-2). 17–68. doi:10.1080/016909696387213. https://www.tandfonline.com/doi/full/10.1080/016909696387213.

Beckman, Mary E., Fangfang Li, Eun Jong Kong & Jan Edwards. 2014. Aligning the timelines of phonological acquisition and change. *Laboratory Phonology* 5(1). doi:10.1515/lp-2014-0007. https://www.degruyter.com/document/doi/10.1515/lp-2014-0007/html.

Bellik, Jennifer, Junko Ito, Nick Kalivoda & Armin Mester (eds.). 2023. *Syntax-Prosody in Optimality Theory: Theory and Analyses* Advances in Optimality Theory. Bristol, CT: Equinox Publishing Ltd.

Bennett, Ryan, Emily Elfner & James McCloskey. 2016. Lightest to the Right: An Apparently Anomalous Displacement in Irish. *Linguistic Inquiry* 47(2). 169–234. doi:10.1162/LING_a_00209. https://direct.mit.edu/ling/article/47/2/169-234/630.

Bickel, Balthasar, Kristine A Hildebrandt & René Schiering. 2009. The distribution of phonological word domains: A probabilistic typology. In Janet Grijzenhout & Baris Kabak (eds.), *Phonological Domains*, De Gruyter, Inc.

Boersma, Paul & David Weenink. 2023. Praat: Doing phonetics by computer. http://www.praat.org/.

Bouavichith, Dominique & Lisa Davidson. 2013. Segmental and Prosodic Effects on Intervocalic Voiced Stop Reduction in Connected Speech. *Phonetica* 70(3). 182–206. doi:10.1159/000355635.

Brauer, Markus & John J. Curtin. 2018. Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods* 23(3). 389–411. doi:10.1037/met0000159. https://doi.apa.org/doi/10.1037/met0000159.

Browman, Catherine P & Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6(2). 201–251.

Broś, Karolina, Marzena Żygis, Adam Sikorski & Jan Wołłejko. 2021. Phonological contrasts and gradient effects in ongoing lenition in the Spanish of Gran Canaria. *Phonology* 38(1). 1–40. doi:10.1017/S0952675721000038.

Brugos, Alejna, Mara Breen, Stefanie Shattuck-Hufnagel, Nanette Veilleux & Jonathan Barnes. 2023. Marking prominence: Towards cue-based annotation of prosodic prominence. In Radek Skarnitzl & Jan Volín (eds.), *Proceedings of the 20th International Congress of Phonetic Sciences*, 1673–1677. Guarant International.

Brugos, Alejna, Mara Breen, Nanette Veilleux, Jonathan Barnes & Stefanie Shattuck-Hufnagel. 2018. Cue-based annotation and analysis of prosodic boundary events. In *Speech Prosody 2018*, 245–249. ISCA. doi:10.21437/SpeechProsody.2018-50. https://www.isca-archive.org/speechprosody_2018/brugos18_speechprosody.html.

Buxó-Lugo, Andrés & Duane G. Watson. 2016. Evidence for the influence of syntax on prosodic parsing. *Journal of Memory and Language* 90. 1–13. doi:10.1016/j.jml.2016.03.001. https://linkinghub.elsevier.com/retrieve/pii/S0749596X16000231.

Büring, Daniel. 2009. Towards a typology of focus realization. In Malte Zimmermann & Caroline Féry (eds.), *Information Structure*, 177–205. Oxford University PressOxford 1st edn. doi:10.1093/acprof:oso/9780199570959.003.0008. https://academic.oup.com/book/36175/chapter/314559956.

Chang, Charles B. 2008. The acoustics of Korean fricatives revisited. *Harvard Studies in Korean Linguistics* 12. 137–150.

Cheng, Chierh, Yi Xu & Michele Gubian. 2010. Exploring the mechanism of tonal contraction in Taiwan Mandarin. In *Interspeech 2010*, 2010–2013. ISCA. doi:10.21437/Interspeech.2010-569. https://www.isca-archive.org/interspeech_2010/cheng10b_interspeech.html.

Cho, Hyesun & Edward Flemming. 2015. Compression and truncation: The case of Seoul Korean accentual phrase. *Studies in Phonetics, Phonology, and Morphology* 21. 359–382.

Cho, Taehong. 2001. Effects of Morpheme Boundaries on Intergestural Timing: Evidence from Korean. *Phonetica* 58(3). 129–162. doi:10.1159/000056196.

Cho, Taehong. 2022. The phonetics-prosody interface and prosodic strengthening in Korean. In *The cambridge handbook of korean linguistics cambridge handbooks in language and linguistics*, 248–293. Cambridge University Press Cambridge.

Cho, Taehong & Sun-Ah Jun. 2000. Domain-initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean. *UCLA working papers in phonetics* 57–70.

Cho, Taehong & Patricia A Keating. 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of phonetics* 29(2). 155–190.

Cho, Taehong, James M McQueen & Ethan A Cox. 2007. Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics* 35(2). 210–243.

Cho, Young-mee Yu. 1990. Syntax and phrasing in Korean. *The phonology-syntax connection* 47–62.

Choi, Jiyoun, Sahyang Kim & Taehong Cho. 2020. An apparent-time study of an ongoing sound change in Seoul Korean: A prosodic account. *PLOS ONE* 15(10). e0240682. doi:10.1371/journal.pone.0240682. https://dx.plos.org/10.1371/journal.pone.0240682.

Choi, Tae-Hwan, Nam-Sil Lim & Jeong-Im Han. 2006. Effects of word frequency on a lenition process: Evidence from stop voicing and/h/reduction in Korean. *Speech Sciences* 13(3). 35–48.

Choi, Youngon & Reiko Mazuka. 2001. Are Prosodic Units Universally Important to Children in their Early Language Acquisition? Korean children's Detection and Use of Language-Specific Prosodic Units. *Journal of Cognitive Science* 2. 171–193.

Choi, Youngon & Reiko Mazuka. 2003. Young children's use of prosody in sentence parsing. *Journal of psycholinguistic research* 32. 197–217.

Christophe, Anne, Sharon Peperkamp, Christophe Pallier, Eliza Block & Jacques Mehler. 2004. Phonological phrase boundaries constrain lexical access I. Adult data. *Journal of Memory and Language* 51(4). 523–547. doi:10.1016/j.jml.2004.07.001. https://www.sciencedirect.com/science/article/pii/S0749596X04000816.

Chung, Daeho. 2011. A Constituency-based Explanation of Syntactic Restrictions on Korean Predicates. *Linguistic Research* 28(1). 199–221.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford University Press.

Clayards, Meghan, Michael K Tanenhaus, Richard N Aslin & Robert A Jacobs. 2008. Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* 108(3). 804–809.

Cohen Priva, Uriel. 2017. Informativity and the actuation of lenition. *Language* 569–597.

Colantoni, Laura & Irina Marinescu. 2010. The Scope of Stop Weakening in Argentine Spanish. In Marta Ortega-Llebaria (ed.), *Proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology*, 100–114. Somerville, MA: Cascadilla Proceedings Project.

Cole, Jennifer, Yoonsook Mo & Soondo Baek. 2010. The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes* 25(7-9). 1141–1177. doi:10.1080/01690960903525507. https://www.tandfonline.com/doi/full/10.1080/01690960903525507.

Dalcher, C. 2007. Statistical methods for quantitative analysis of multiple lenition components. In *Icphs*, 857–860. Saarbrücken, Germany: International Phonetic Association.

Davidson, Lisa. 2016. Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics* 54. 35–50.

Davidson, Lisa. 2018. Phonation and laryngeal specification in American English voiceless obstruents. *Journal of the International Phonetic Association* 48(3). 331–356.

De Jong, Kenneth. 1998. Stress-related variation in the articulation of coda alveolar stops: flapping revisited. *Journal of Phonetics* 26(3). 283–310. doi:10.1006/jpho.1998.0077.

DiCanio, Christian, Wei-Rong Chen, Joshua Benn, Jonathan D Amith & Rey Castillo García. 2022. Extreme stop allophony in Mixtec spontaneous speech: Data, word prosody, and modelling. *Journal of Phonetics* 92. 101147.

Downing, Laura J. 2008. Focus and prominence in Chichewa, Chitumbuka and Durban Zulu. *ZAS Papers in Linguistics* 49. 47–65. doi:10.21248/zaspil.49.2008.363. `https://zaspil.leibniz-zas.de/article/view/363`.

Egedi, Barbara. 2009. Adverbial (dis)ambiguities. Syntactic and prosodic features of ambiguous predicational adverbs. In Katalin É. Kiss (ed.), *Adverbs and Adverbial Adjuncts at the Interfaces*, 103–132. Mouton de Gruyter. doi:10.1515/9783110214802.1.103.

Ennever, Thomas, Felicity Meakins & Erich R Round. 2017. A replicable acoustic measure of lenition and the nature of variability in Gurindji stops. *Laboratory Phonology* 8(1).

Ernst, Thomas. 2002. *The Syntax of Adjuncts*. Cambridge University Press.

Ernst, Thomas. 2020. The Syntax of Adverbials. *Annual Review of Linguistics* 6(1). 89–109. doi:10.1146/annurev-linguistics-011619-030334. `https://www.annualreviews.org/doi/10.1146/annurev-linguistics-011619-030334`.

Farinella, Alessa & Seung Suk Lee. 2024. The perception of tonal and segmental cues to phrasing in Seoul Korean.

Feldman, Naomi, Thomas Griffiths & James Morgan. 2009. Learning phonetic categories by learning a lexicon. In *Proceedings of the annual meeting of the cognitive science society*, vol. 31, .

Frazier, Lyn & Keith Rayner. 1982. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14(2). 178–210. doi:10.1016/0010-0285(82)90008-1. `https://linkinghub.elsevier.com/retrieve/pii/0010028582900081`.

Frenck-Mestre, Cheryl, Hyeree Choo, Ana Zappa, Julia Herschensohn, Seung-Kyung Kim, Alain Ghio & Sungryung Koh. 2022. The Online Processing of Korean Case by Native Korean Speakers and Second Language Learners as Revealed by Eye Movements. *Brain Sciences* 12(9). 1–28. doi:10.3390/brainsci12091230. `https://www.mdpi.com/2076-3425/12/9/1230`.

Gim, Jin-Hae. 2004. A Study on the Disambiguation of Distribution of Korean Adverb 'geuman'. *Discourse and Cognition* 11(2). 43–64.

Goldwater, Sharon & Mark Johnson. 2003. Learning ot constraint rankings using a maximum entropy model. In *Proceedings of the stockholm workshop on variation within optimality theory*, vol. 111, 120.

Gorman, Kyle B. 2014. *textgrid: A Python module for interacting with Praat TextGrid files*. `http://github.com/kylebgorman/textgrid/`. Python package.

Gubian, Michele. 2024. *landmarkregutils: Utilities for landmark registration*. `https://github.com/uasolo/landmarkregUtils`. R package version 0.0.0.9000.

Gubian, Michele, Francesco Cangemi & Lou Boves. 2010. Automatic and data driven pitch contour manipulation with functional data analysis. In *Speech Prosody 2010*, paper 954–0. ISCA. doi:10.21437/SpeechProsody.2010-98. `https://www.isca-archive.org/speechprosody_2010/gubian10_speechprosody.html`.

Gubian, Michele, Manfred Pastätter & Marianne Pouplier. 2019. Zooming in on Spatiotemporal V-to-C Coarticulation with Functional PCA. In *Interspeech 2019*, 889–893. ISCA. doi:10.21437/Interspeech.2019-2143. `https://www.isca-archive.org/interspeech_2019/gubian19b_interspeech.html`.

Gubian, Michele, Francisco Torreira & Lou Boves. 2015. Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics* 49. 16–40. doi:10.1016/j.wocn.2014.10.001. `https://linkinghub.elsevier.com/retrieve/pii/S009544701400076X`.

Han, Jeong-Im. 2000. Intervocalic stop voicing revisited. *Speech Sciences* 7(1). 203–216.

Happ-Kurz, Clara. 2022. *Mfpca: Multivariate functional principal component analysis for data observed on different dimensional domains*. `https://github.com/ClaraHapp/MFPCA`. R package version 1.3-10.

Harris, John & Eno-Abasi Urua. 2001. Lenition degrades information: consonant allophony in Ibibio. *Speech, Hearing and Language: work in progress* 13.

Hatcher, Richard, Hyunjung Joo, Sahyang Kim & Taehong Cho. 2024. Focus-induced tonal distribution in Seoul Korean as an edge-prominence language. *Journal of Phonetics* 107. 1–29. doi:10.1016/j.wocn.2024.101353. `https://linkinghub.elsevier.com/retrieve/pii/S0095447024000597`.

Hayes, Bruce & Aditi Lahiri. 1991. Bengali intonational phonology. *Natural Language & Linguistic Theory* 47–96.

Hirst, Daniel J. 2007. A Praat plugin for Momel and INTSINT with improved algorithms for modelling and coding intonation. In *16th international congress of phonetic sciences icphs xvi*, .

Hughes, Cerys, Seung Suk Lee, Alessa Farinella & Kristine M Yu. 2023. Phonetic Implementation of Phonologically Different High Tone Spans in Luganda. In R Skarnitzl & J Volín (eds.), *Proceedings of the 20th International Congress of the Phonetic Sciences*, 1985–1989. Prague, Czech Republic: International Phonetic Association.

Hyde, Brett. 2012. Alignment constraints. *Natural Language & Linguistic Theory* 30(3). 789–836. doi:10.1007/s11049-012-9167-3. http://link.springer.com/10.1007/s11049-012-9167-3.

Im, Suyeon, Jennifer Cole & Stefan Baumann. 2023. Standing out in context: Prominence in the production and perception of public speech. *Laboratory Phonology* 24(1). doi:10.16995/labphon.6417. https://www.journal-labphon.org/article/id/6417/.

Ishihara, Shinichiro. 2002. Invisible but Audible Wh-Scope Marking: WH-Constructions and Deaccenting in Japanese. In *Proceedings of the West Coast Conference on Formal Linguistics (WCCFL) 21*, .

Ito, Junko & Armin Mester. 2012. Recursive prosodic phrasing in Japanese. In Toni Borowsky, Shigeto Kawahara, Takahito Shinya & Mariko Sugahara (eds.), *Prosody Matters: Essays in Honor of Elisabeth Selkirk*, 280–303. London: Equinox.

Jackendoff, Ray S. 1972. *Semantic interpretation in generative grammar.* ERIC.

Jang, Jiyoung, Sahyang Kim & Taehong Cho. 2018. Focus and boundary effects on coarticulatory vowel nasalization in Korean with implications for cross-linguistic similarities and differences. *The Journal of the Acoustical Society of America* 144(1). EL33–EL39.

Jun, Sun-Ah. 1993. *The phonetics and phonology of Korean prosody*: Ohio State University dissertation.

Jun, Sun-Ah. 1994. Lenis Stop Voicing Rule. *Theoretical Issues in Korean Linguistics* 101–114.

Jun, Sun-Ah. 1995. Asymmetrical prosodic effects on the laryngeal gesture in Korean. *Phonology and phonetic evidence: Papers in Laboratory Phonology IV* 235–253.

Jun, Sun-Ah. 1996. Influence of microprosody on macroprosody: a case of phrase initial strengthening. *UCLA Working Papers in Phonetics* 92. 97–116.

Jun, Sun-Ah. 1998. The accentual phrase in the Korean prosodic hierarchy. *Phonology* 15(2). 189–226.

Jun, Sun-Ah. 2000. K-ToBI (Korean ToBI) Labelling Conventions (version 3.1). https://linguistics.ucla.edu/people/jun/ktobi/k-tobi.html.

Jun, Sun-Ah. 2006. Intonational phonology of Seoul Korean revisited. *Japanese-Korean Linguistics* 14. 15–26.

Jun, Sun-Ah. 2007. The intermediate phrase in Korean: Evidence from sentence processing. *Tones and tunes: Experimental studies in word and sentence prosody* 143–170.

Jun, Sun-Ah. 2011. Prosodic markings of complex NP focus, syntax, and the pre-/post-focus string. In *Proceedings of the 28th west coast conference on formal linguistics*, 214–230. Citeseer.

Jun, Sun-Ah. 2014. Prosodic typology: by prominence type, word prosody, and macro-rhythm. In Sun-Ah Jun (ed.), *Prosodic Typology II*, 520–539. Oxford University PressOxford 1st edn. doi:10.1093/acprof:oso/9780199567300.003.0017. https://academic.oup.com/book/27198/chapter/196665443.

Jun, Sun-Ah. 2022. The ToBI Transcription System: Conventions, Strengths, and Challenges. In Jonathan Barnes & Stefanie Shattuck-Hufnagel (eds.), *Prosodic Theory and Practice*, 151–181. The MIT Press. doi:10.7551/mitpress/10413.003.0007. https://direct.mit.edu/books/book/5259/chapter/3572257/The-ToBI-Transcription-System-Conventions.

Jun, Sun-Ah & Sahyang Kim. 2004. Default phrasing and attachment preference in Korean. In *Interspeech 2004*, 3009–3012. Jeju: ISCA. doi:10.21437/Interspeech.2004-755. https://www.isca-archive.org/interspeech_2004/jun04_interspeech.html.

Jun, Sun-Ah, Sook-Hyang Lee, Keeho Kim & Yong-Ju Lee. 2000. Labeler agreement in Transcribing Korean Intonation with K-toBI. In *Interspeech*, 211–214.

Jun, Sun-Ah & Mira Oh. 1996. A prosodic analysis of three types of wh-phrases in Korean. *Language and Speech* 39(1). 37–61.

Jung, Youngim, SunHo Cho, Aesun Yoon & Hyuk-Chul Kwon. 2007a. Prediction of Prosodic Break Using Syntactic Relations and Prosodic Features. In *Annual Conference on Human and Language Technology*, 7–14.

Jung, Youngim, Sunho Cho, Aesun Yoon & Hyuk-Chul Kwon. 2007b. Prediction of prosodic break using syntactic relations and prosodic features. In *Annual conference on human and language technology*, 7–14. Human and Language Technology.

Kang, Kyoung-Ho & Susan G. Guion. 2008. Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *The Journal of the Acoustical Society of America* 124(6). 3909–3917. doi:10.1121/1.2988292. https://pubs.aip.org/jasa/article/124/6/3909/899069/Clear-speech-production-of-Korean-stops-Changing.

Katz, Jonah. 2016. Lenition, perception and neutralisation. *Phonology* 33(1). 43–85. doi:10.1017/S0952675716000038.

Katz, Jonah. 2021. Intervocalic lenition is not phonological: evidence from Campidanese Sardinian. *Phonology* 38(4). 651–692. doi:10.1017/S095267572100035X.

Katz, Jonah & Melinda Fricke. 2018. Auditory disruption improves word segmentation: A functional basis for lenition phenomena. *Glossa: a journal of general linguistics* 3(1). 38.

Katz, Jonah & Gianmarco Pitzanti. 2019. The phonetics and phonology of lenition: A Campidanese Sardinian case study. *Laboratory Phonology* 10(16). 1–40.

Katz, Jonah & Elisabeth Selkirk. 2011. Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language* 87(4). 771–816. doi:10.1353/lan.2011.0076. https://muse.jhu.edu/article/461766.

Keizer, Evelien. 2018. Interpersonal adverbs in FDG: The case of *frankly*. In Evelien Keizer & Hella Olbertz (eds.), *Studies in Language Companion Series*, vol. 205, 47–88. Amsterdam: John Benjamins Publishing Company. doi:10.1075/slcs.205.03kei.

Kim, Hyun-ju & Yeon-ju Jeong. 2019. A brief survey on the polysemy and its historical development of a Korean root kuman (. *Morphology* 21(2). 248–270.

Kim, Jong-Bok. 1999. Korean Short Form Negation and Related Phenomena_ A Lexicalist, Constraint-Based Analysis.pdf. *Language and Information* 3(2). 13–29.

Kim, Jungsun, Weonhee Yun & Ducksoo Kang. 2016. Patterns of consonant deletion in the word-internal onset position: Evidence from spontaneous Seoul Korean speech. *Phonetics and Speech Sciences* 8(1). 45–51.

Kim, Minho, Youngim Jung & Hyuk-Chul Kwon. 2021. Korean Prosody Phrase Boundary Prediction Model for Speech Synthesis Service in Smart Healthcare. *Electronics* 10(19). 2371.

Kim, Miran. 2015. Focus-driven prosodic restructuring patterns: A unified OT account of deaccenting and dephrasing. *Studies in Phonetics, Phonology, and Morphology* 21(1). 25–56. doi:10.17959/SPPM.2015.21.1.25. https://doi.org/10.17959/SPPM.2015.21.1.25.

Kim, Sahyang. 2004. *The role of prosodic phrasing in Korean word segmentation*: University of California, Los Angeles dissertation.

Kim, Sahyang & Taehong Cho. 2009. The use of phrase-level prosodic information in lexical segmentation: Evidence from word-spotting experiments in Korean. *The Journal of the Acoustical Society of America* 125(5). 3373–3386.

Kim, Sahyang & Taehong Cho. 2013. Prosodic boundary information modulates phonetic categorization. *The Journal of the Acoustical Society of America* 134(1). EL19–EL25. doi:10.1121/1.4807431. https://pubs.aip.org/jasa/article/134/1/EL19/614573/Prosodic-boundary-information-modulates-phonetic.

Kim, Sahyang, Taehong Cho & James M McQueen. 2012. Phonetic richness can outweigh prosodically-driven phonological knowledge when learning words in an artificial language. *J. Phon.* 40(3). 443–452.

Kim, Sahyang, Holger Mitterer & Taehong Cho. 2018. A time course of prosodic modulation in phonological inferencing: The case of Korean post-obstruent tensing. *PLOS ONE* 13(8). e0202912. doi:10.1371/journal.pone.0202912. `https://dx.plos.org/10.1371/journal.pone.0202912`.

Kingston, John. 2008. Lenition. In *3rd conference on laboratory approaches to spanish phonology*, 1–31. Cascadilla Proceedings Project.

Klein, Wolfgang, Reinier Plomp, Louis CW Pols et al. 1970. Vowel spectra, vowel spaces, and vowel identification. *Journal of the Acoustical Society of America* 48(4). 999–1009.

Kleinschmidt, Dave F & T Florian Jaeger. 2015. Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological review* 122(2). 148–203.

Kojadinović, Zlatan. 2022. Variation in the prosody of illocutionary adverbs. *Open Linguistics* 8(1). 593–617. doi:10.1515/opli-2022-0209.

Kubota, Ai. 2015. Transforming manner adverbs into subject-oriented adverbs: Evidence from Japanese. *Natural Language & Linguistic Theory* 33(3). 1019–1046. doi:10.1007/s11049-015-9287-7.

Kurumada, Chigusa & Timo B Roettger. 2022. Thinking probabilistically in the study of intonational speech prosody. *Wiley Interdisciplinary Reviews: Cognitive Science* 13(1). e1579.

Kuzla, Claudia, Mirjam Ernestus & Holger Mitterer. 2010. Compensation for assimilatory devoicing and prosodic structure in German fricative perception. In Cécile Fougeron, Barbara Kühnert, Mariapaola D'Imperio & Nathalie Vallée (eds.), *Laboratory Phonology 10*, 731–758. DE GRUYTER MOUTON. doi:10.1515/9783110224917.5.731. `https://www.degruyter.com/document/doi/10.1515/9783110224917.5.731/html`.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26. doi:10.18637/jss.v082.i13.

Kwon, Iksoo. 2013. Evidentiality in Korean conditional constructions. *Linguistics* 51(6). doi:10.1515/ling-2013-0049.

Ladd, D Robert. 2008. *Intonational phonology.* Cambridge University Press.

Ladefoged, P. 2003. *Phonetic data analysis: An introduction to fieldwork and instrumental techniques.* Malden, MA: Wiley-Blackwell.

Lavoie, Lisa M. 2001. *Consonant strength: Phonological patterns and phonetic manifestations*. Routledge.

Lee, Iksop & S Robert Ramsey. 2000. *The korean language*. Suny Press.

Lee, Jiyeon & Hoyoung Lee. 2013. Korean Intonation Patterns from the Viewpoint of f0 Percentage Change. *Language and Speech Sciences* 5(1). 123–130.

Lee, Seung Suk. 2023. Detecting Accentual Phrase Boundaries In Seoul Korean Using Tonal and Segmental Cues. In *Proceedings of the 20th International Congress of the Phonetic Sciences*, 1681–1685. Prague, Czech Republic: International Phonetic Association.

Lee, Seunghun J. 2022. Constituent Structure and Sentence Phonology of Korean. In *The Cambridge Handbook of Korean Linguistics*, 294–317. Cambridge University Press.

Lee, Sook-hyang. 1995. Reduction of velar lenis stops in Korean. *Linguistics* 321–331.

Lee, Yong-cheol. 2017. Prosodic focus in Seoul Korean and South Kyungsang Korean. *Linguistic Research* 34(1). 133–161.

Lee, Yong-cheol & Yi Xu. 2010. Phonetic realization of contrastive focus in Korean. In *Speech Prosody 2010*, paper 033–0. ISCA. doi:10.21437/SpeechProsody.2010-81. `https://www.isca-archive.org/speechprosody_2010/lee10_speechprosody.html`.

Lee, Yoonjeong. 2018. *The Prosodic Substrate of Consonant and Tone Dynamics*: University of Southern California Ph.D.

Lehiste, Ilse, Joseph P. Olive & Lynn A. Streeter. 1976. Role of duration in disambiguating syntactically ambiguous sentences. *The Journal of the Acoustical Society of America* 60. 1199–1202.

Lenth, Russell V. 2018. *emmeans: Estimated marginal means, aka least-squares means*. `https://rvlenth.github.io/emmeans/`. R package version 1.11.1-00001.

Lewis, Anthony Murray. 2001. *Weakening of intervocalic/p, t, k/in two Spanish dialects: Toward the quantification of lenition processes*: University of Illinois at Urbana-Champaign dissertation.

Lüdecke, Daniel. 2023. *sjplot: Data visualization for statistics in social science*. `https://CRAN.R-project.org/package=sjPlot`. R package version 2.8.15.

Mansfield, John Basil. 2015. Consonant lenition as a sociophonetic variable in Murrinh Patha (Australia). *Language Variation and Change* 27(2). 203–225. doi:10.1017/S0954394515000046.

Maye, Jessica, Janet F Werker & LouAnn Gerken. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82(3). B101–B111.

McCarthy, John J & Alan Prince. 1993. Generalized alignment. In *Yearbook of morphology 1993*, 79–153. Springer.

McQueen, James M. & Laura Dilley. 2020. Prosody and Spoken-Word Recognition. In *The Oxford Handbook of Language Prosody*, Oxford University Press. doi:10.1093/oxfordhb/9780198832232.013.33.

Mitterer, Holger, Sahyang Kim & Taehong Cho. 2021. The Role of Segmental Information in Syntactic Processing Through the Syntax–Prosody Interface. *Language and Speech* 64(4). 962–979. doi:10.1177/0023830920974401. `http://journals.sagepub.com/doi/10.1177/0023830920974401`.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Foris Publications.

Noh, Taekyu. 1983. Meaning and scope of adverbs-focusing on 'keuman' and 'iman'. *Hangeul* 180. 131–154.

Norris, Dennis & James M McQueen. 2008. Shortlist B: a Bayesian model of continuous speech recognition. *Psychological review* 115(2). 357.

Oh, Mi-Ra. 2008. Prosody and information structure: Phonetic realizations of focus and topic in korean. *Speech Sciences* 15(2). 7–19.

Parrell, Benjamin & Shrikanth Narayanan. 2018. Explaining Coronal Reduction: Prosodic Structure and Articulatory Posture. *Phonetica* 75(2). 151–181. doi:10.1159/000481099.

Pater, Joe. 2009. Weighted Constraints in Generative Linguistics. *Cognitive Science* 33(6). 999–1035. doi:10.1111/j.1551-6709.2009.01047.x. `https://onlinelibrary.wiley.com/doi/10.1111/j.1551-6709.2009.01047.x`.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot & E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12. 2825–2830.

Pierrehumbert, Janet & Mary E. Beckman. 1988. *Japanese tone structure*. MIT Press Cambridge.

Pierrehumbert, Janet B. 1980. *The Phonology and Phonetics of English Intonation*. Cambridge, Massachusetts: Massachusetts Institute of Technology Doctoral.

Potts, Christopher. 2005. *The Logic of Conventional Implicatures*. Oxford University Press.

Prince, Alan & Paul Smolensky. 1993/2004. *Optimality theory: Constraint interaction in generative grammar*. Blackwell.

Puggaard-Rode, Rasmus, Camilla Søballe Horslund & Henrik Jørgensen. 2022. The rarity of intervocalic voicing of stops in Danish spontaneous speech. *Laboratory Phonology* 13(1). doi:10.16995/labphon.6449.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.

Ramsay, James O. & B. W. Silverman. 2005. *Functional data analysis* Springer series in statistics. New York, NY: Springer 2nd edn.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1). 75–116. doi:10.1007/BF02342617. http://link.springer.com/10.1007/BF02342617.

Ryalls, John, Allison Zipprer & Penelope Baldauff. 1997. A Preliminary Investigation of the Effects of Gender and Race on Voice Onset Time. *Journal of Speech, Language, and Hearing Research* 40(3). 642–645. doi:10.1044/jslhr.4003.642.

Salverda, Anne Pier, Delphine Dahan & James M. McQueen. 2003. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 90(1). 51–89. doi:10.1016/S0010-0277(03)00139-2. https://www.sciencedirect.com/science/article/pii/S0010027703001392.

Schafer, Amy Jean & Sun-Ah Jun. 2002. Effects of Accentual Phrasing on Adjective Interpretation in Korean. In *Sentence processing in east asian languages*, 223–255.

Scott, Donia R. & Anne Cutler. 1984. Segmental phonology and the perception of syntactic structure. *Journal of Verbal Learning and Verbal Behavior* 23(4). 450–466. doi:10.1016/S0022-5371(84)90291-3. https://linkinghub.elsevier.com/retrieve/pii/S0022537184902913.

Selkirk, Elisabeth. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. The MIT Press.

Selkirk, Elisabeth. 1986. On derived domains in sentence phonology. *Phonology* 3. 371–405.

Selkirk, Elisabeth. 1996. The Prosodic Structure of Function Words. *University of Massachusetts Occasional Papers in Linguistics* 18. 439–469. doi:10.1002/9780470756171.ch25. https://onlinelibrary.wiley.com/doi/10.1002/9780470756171.ch25.

Selkirk, Elisabeth. 2005. Comments on Intonational Phrasing in English. In Frota, Sónia, M Vigário & M. J. Freitas (eds.), *Prosodies : With special reference to iberian languages*, 11–58. De Gruyter, Inc.

Selkirk, Elisabeth. 2011. The Syntax-Phonology Interface. In John Goldsmith, Jason Riggle & Alan Yu (eds.), *The Handbook of Phonological Theory*, Oxford: Blackwell Publishing 2nd edn.

Selkirk, Elisabeth & Seunghun J. Lee. 2015. Constituency in sentence phonology: an introduction. *Phonology* 32(1). 1–18. doi:10.1017/S0952675715000020. https://www.cambridge.org/core/product/identifier/S0952675715000020/type/journal_article.

Selkirk, Elisabeth et al. 1982. The syllable. *The structure of phonological representations* 2. 337–383.

Sells, Peter. 2015. Negation and Negative Polarity Items. In Lucien Brown & Jaehoon Yeon (eds.), *The Handbook of Korean Linguistics*, 212–231. Wiley 1st edn. doi:10.1002/9781118371008.ch12.

Shattuck-Hufnagel, Stefanie & Alice E. Turk. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25(2). 193–247. doi:10.1007/BF01708572. http://link.springer.com/10.1007/BF01708572.

Silva, David J. 2006. Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 23(02). 287–308. doi:10.1017/S0952675706000911. http://www.journals.cambridge.org/abstract_S0952675706000911.

Silva, David James. 1992. *The phonetics and phonology of stop lenition in Korean*: Cornell University dissertation.

Silverman, Kim EA, Mary E Beckman, John F Pitrelli, Mari Ostendorf, Colin W Wightman, Patti Price, Janet B Pierrehumbert & Julia Hirschberg. 1992. ToBI: A standard for labeling English prosody. In *Icslp*, vol. 2, 867–870.

Simonet, Miquel, José I. Hualde & Marianna Nadeu. 2012. Lenition of /d/ in spontaneous Spanish and Catalan. In *Interspeech 2012*, 1416–1419. ISCA. doi:10.21437/Interspeech.2012-50.

Smith, Jennifer L. 2008. Markedness, faithfulness, positions, and contexts: Lenition and Fortition in Optimality Theory. In Joaquim Brandão De Carvalho, Tobias Scheer & Philippe Ségéral (eds.), *Lenition and Fortition*, 519–560. Mouton de Gruyter. doi:10.1515/9783110211443.3.519.

Smolensky, Paul & Géraldine Legendre. 2006. *The harmonic mind: From neural computation to optimality-theoretic grammar (Linguistic and philosophical implications)*. MIT Press.

Soh, Hooi Ling. 2009. Speaker presupposition and Mandarin Chinese sentence-final -le: a unified analysis of the “change of state” and the “contrary to expectation” reading. *Natural Language & Linguistic Theory* 27. 623–657.

Steffman, Jeremy. 2019. Intonational structure mediates speech rate normalization in the perception of segmental categories. *Journal of Phonetics* 74. 114–129. doi:10.1016/j.wocn.2019.03.002. https://linkinghub.elsevier.com/retrieve/pii/S0095447018301177.

Steffman, Jeremy, Sahyang Kim, Taehong Cho & Sun-Ah Jun. 2022. Prosodic phrasing mediates listeners' perception of temporal cues: Evidence from the Korean Accentual Phrase. *Journal of Phonetics* 94. 101156. doi:10.1016/j.wocn.2022.101156. https://linkinghub.elsevier.com/retrieve/pii/S0095447022000316.

Swartz, Bradford L. 1992. Gender Difference in Voice Onset Time. *Perceptual and Motor Skills* 75. 983–992.

Szigetvári, Péter. 2008. What and where? In Joaquim Brandao de Carvalho, Tobias Scheer & Philippe Ségéral (eds.), *Lenition and fortition* (Studies in generative grammar 99), 93–129. Berlin: Mouton de Gruyter.

Tan, Jialu. 2023. *A study on the adverb 'keuman' in Korean*: Seoul National University MA thesis.

Torreira, Francisco & Mirjam Ernestus. 2011. Realization of voiceless stops and vowels in conversational French and Spanish. *Laboratory Phonology* 2(2). doi:10.1515/labphon.2011.012.

Tremblay, Annie, Taehong Cho, Sahyang Kim & Seulgi Shin. 2019. Phonetic and phonological effects of tonal information in the segmentation of Korean speech: An artificial-language segmentation study. *Appl. Psycholinguist.* 40(5). 1221–1240.

Truckenbrodt, Hubert. 1995. *Phonological Phrases: Their Relation to Syntax, Focus, and Prominence*: Massachusetts Institute of Technology Ph.D.

Truckenbrodt, Hubert. 1999. On the Relation between Syntactic Phrases and Phonological Phrases. *Linguistic Inquiry* 30(2). 219–255. doi:10.1162/002438999554048. https://direct.mit.edu/ling/article/30/2/219-255/44.

Truckenbrodt, Hubert. 2002. Variation in p-phrasing in Bengali. *Linguistic Variation Yearbook* 2. 259–303. doi:10.1075/livy.2.10tru. http://www.jbe-platform.com/content/journals/10.1075/livy.2.10tru.

Turk, Alice, Satsuki Nakai & Mariko Sugahara. 2006. Acoustic segment durations in prosodic research: A practical guide. *Methods in empirical prosody research* 3. 1–28.

Wagner, Michael & Michael McAuliffe. 2019. The effect of focus prominence on phrasing. *Journal of Phonetics* 77. 100930. doi:10.1016/j.wocn.2019.100930. https://linkinghub.elsevier.com/retrieve/pii/S0095447018300767.

Wang, Chen. 2018. *The Syntax of Le in Mandarin Chinese*. London, UK: Queen Mary University of London dissertation.

Warner, Natasha & Benjamin V. Tucker. 2011. Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America* 130(3). 1606–1617. doi:10.1121/1.3621306.

Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf & Patti J. Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America* 91(3). 1707–1717. doi: 10.1121/1.402450. `https://pubs.aip.org/jasa/article/91/3/1707/893892/Segmental-durations-in-the-vicinity-of-prosodic`.

Xie, Xin, Andrés Buxó-Lugo & Chigusa Kurumada. 2021. Encoding and decoding of meaning through structured variability in intonational speech prosody. *Cognition* 211. 104619.

Yamashita, Hideaki. 2008. Prosody and the syntax of shika-NPIs in Tokyo Japanese and its implications for the theory of grammar. *Toronto Working Papers in Linguistics* 28. 375–394.

Yoo, H & Sun-Ah Jun. 2016. Distribution of Accentual Phrase-medial tones in Seoul Korean [Poster]. In *The acoustical society of america*, Honolulu, Hawaii.

Yoo, Kayeon. 2020. *The production and perception of domain-initial strengthening in Seoul, Busan, and Ulsan Korean*: University of Cambridge dissertation.

Yoon, James Hye-Suk. 1993. Tense, coordination, and the clausal structures of english and korean. *Harvard Studies in Korean Linguistics* 5. 436–448.

Yoon, Tae-Jin. 2024. Korean forced aligner. `https://tutorial.tyoon.net/`.

Yu Cho, Young-mee. 1990. Syntax and phrasing in Korean. In Sharon Inkelas & Draga Zec (eds.), *The Phonology-Syntax Connection*, 47–62. University of Chicago Press.

Yun, Ilsung. 2000. Voicing in intervocalic lax obstruents /p, t, k, c/ of Korean. *Speech Sciences* 7(3). 21–33.

Yun, Jiwon & Hye-Sook Lee. 2022. Prosodic disambiguation of questions in Korean: Theory and processing. *Korean Linguistics* 18(1). 18–47. doi:10.1075/kl.00014.yun. `http://www.jbe-platform.com/content/journals/10.1075/kl.00014.yun`.

Yun, Weonhee, Kyuchul Yoon, Sunwoo Park, Juhee Lee, Sungmoon Cho, Ducksoo Kang, Koonhyuk Byun, Hyeseung Hahn & Jungsun Kim. 2015. The Korean corpus of spontaneous speech. *Phonetics and Speech Sciences* 7(2). 103–109.

Zellers, Margaret, Michele Gubian & Brechtje Post. 2010. Redescribing intonational categories with functional data analysis. In *Interspeech 2010*, 1141–1144. ISCA. doi: 10.21437/Interspeech.2010-361. `https://www.isca-archive.org/interspeech_2010/zellers10_interspeech.html`.