

On the prosody of reported speech in Seoul Korean

Makaila Johnston

Supervisor: Dr. Emily Elfner

A Major Research Paper submitted
in partial fulfillment of the requirements for the degree of
Master of Arts in Linguistics and Applied Linguistics

Graduate Program in Linguistics and Applied Linguistics

York University

Toronto, Ontario

August 2025

Table of Contents

Chapter 1: Introduction.....	4
Chapter 2: Background.....	6
2.1 Korean prosody.....	6
2.1.1 Prosodic Hierarchy.....	7
2.1.1.1 Accentual Phrase.....	8
2.1.1.2 Intermediate Phrase.....	10
2.1.1.3 Intonational Phrase.....	11
2.2 Reported speech types.....	12
2.3 Reported speech in Korean.....	15
2.3.1 Direct Speech vs. Indirect Speech in Korean.....	18
2.4 Previous research on reported speech and prosody.....	19
2.4.1 Mimicry and Pitch Variation.....	20
2.4.2 Frames, Boundaries, and Pauses.....	22
2.4.3 Additional Cues.....	24
2.4.4 Conclusion.....	25
2.5 Research questions.....	25
Chapter 3: The prosody of reported speech in Seoul Korean.....	28
3.1 Methodology.....	28
3.1.1 Participants and Tasks.....	29
3.1.1.1 Task one: Dialogue.....	30
3.1.1.2 Task two: Q&A.....	30
3.1.1.3 Task three: Fill-in-the-blank.....	31
3.2 Results.....	32
3.2.1 Methodology.....	32
3.2.1.1 Dialogue Task.....	32
3.2.1.2 Q&A.....	33
3.2.1.3 Fill-in-the-Blank.....	34
3.2.2 Direct and Indirect speech comparisons.....	35
3.2.3 Direct speech.....	37
3.2.4 Indirect speech.....	46
3.2.5 Occurrence of cues in Direct and Indirect speech.....	51
3.3 Summary and Discussion.....	52

Chapter 4: Match Theory Analysis.....	56
4.1 Direct speech.....	57
4.2 Indirect speech.....	68
4.3 Conclusion.....	71
Chapter 5: Conclusion.....	72
Notes.....	74
Appendix A.....	75
Appendix B.....	78
Appendix C.....	79
Appendix D.....	85
References.....	90

Chapter 1: Introduction

A romantic description of prosody is to call it the melody of speech. Though this is not altogether wrong, a linguistic definition would state that prosody is concerned with the suprasegmental features of languages— with intonational features such as tonal structure, pitch, duration, and intensity (Féry, 2016). Elements such as pitch, tonal structure, and intensity imply that prosody is restricted to spoken languages. This is, in fact, not the case; Sandler et al. (2020) claim that prosody exists in signed languages, signalled through manual or non-manual gestures, such as repetition or duration of a sign. Prosody can be organized into a hierarchical structure, which Féry (2016) defines as “[...] the parsing of continuous speech” into “organized prosodic domains.” Though moras, syllables, and feet are all a part of the prosodic structure (Féry, 2016), these lower-level prosodic constituents are not of interest for the current study. Instead, the interest lies in the higher-level prosodic constituents, namely the phonological phrase and intonational phrase. The role of prosody in speech is multifaceted. It can aid in disambiguating structurally ambiguous strings, indicate different sentence types through pitch contours, indicate prominence, convey direct versus indirect speech acts, and signal turn-taking in conversation (Hirschberg, 2002). It is the communication of direct and indirect speech acts which is the focus of the current study.

This paper is concerned with the intonational phonology and prosodic structure of Seoul Korean, specifically with the use of reported speech. There is a structural element in Korean that indicates “quote,” and the interaction of this structural element, or rather morpheme, with intonational elements of the utterance is the focus of this paper. This morpheme (glossed as QUOT below) is illustrated in the examples of indirect and direct reported speech in Korean in (1).

(1) a. Indirect speech:

yecin ssi-nun cikum ka-ntako haysseyo

Yejin HON-TOP now go-QUOT do

“Yejin said that she’s going now.”

b. Direct speech:

yecin ssi-nun cikum ka lako haysseyo

Yejin HON-TOP now go QUOT do

“Yejin said ‘I’m going now.’”

How this structural element may influence the intonational elements, and vice versa, is of great interest. Consequently, this paper is not only concerned with prosody, but the interface between syntax and phonology. My goal is to uncover the extent that prosody is used to convey that a segment of an utterance is a reproduction of speech produced at another time, in another context, by either the current speaker or someone else. In the context of Korean, how will a structural element, the quotative morpheme, influence the use of reported speech prosody? It seems likely that the presence of what could be considered a structural “unquote” (indicating the end of the reported speech) within the utterance would reduce the need for the speaker to use other tools, such as prosody, to inform the listener that what they’re hearing is a quotation. A description of the prosody of reported speech in Korean may provide insight into how prosodic elements interact with syntax, and how prosody may be used in disambiguating not a *syntactically* ambiguous structure, but rather a disembodied element of a conversation, attributed to another time (and potentially another speaker).

This paper is organized as follows. Chapter 2 provides a background on Korean prosody and reported speech. Section 2.1 discusses the Korean prosodic hierarchy, and the characteristics of the Accentual Phrase, Intermediate Phrase, and Intonational Phrase. Section 2.2 covers

reported speech, defining the two types of reported speech being investigated in this paper.

Section 2.3 describes reported speech in Korean, and Section 2.4 provides a literature review on reported speech prosody. The research questions for this paper are addressed in Section 2.5.

Chapter 3 concerns the experiments for this paper, with Section 3.1 discussing methodology, and Section 3.2 providing the results, with a summary in Section 3.3. Chapter 4 contains my analysis of the prosody of reported speech in Korean, where I propose that a recursive prosodic structure based on Match Theory succeeds in predicting the location and type of boundaries used to flag reported speech within an utterance. Chapter 5 provides a conclusion to the paper.

Chapter 2: Background

2.1 Korean prosody

This paper will examine the prosody of reported speech in Seoul Korean. There are six major dialect zones in the Korean Peninsula: the northwestern dialects, the northeastern dialects, the central dialects, the southwestern dialects, the southeastern dialects, and the Jeju dialect (Brown & Yeon, 2015). “Standard” Korean is based on the Seoul variety, which is part of the central dialects. Seoul Korean does not have lexical pitch accent (Jun, 2005a; Jeon, 2015) unlike other varieties of Korean with pitch accents, such as the Kyungsang dialect (Jun, 2005a) or Yanbian Korean (Jun & Jiang, 2019). What this means is that the interpretation of a lexical item is not dependent on the stress or tonal pattern of the word. Whether Seoul Korean has fixed stress at the word level, or phrasal stress, is debated (Jun, 2005a). Seoul Korean can be considered a

phrase language, as it does not have lexical tones or stress (Féry, 2016). As no other varieties are being explored in this paper, “Korean” will refer to Seoul Korean henceforth.

2.1.1 Prosodic Hierarchy

Work done by Jun (1996, 1998, 2005a, 2005b) proposes that the prosodic structure of Korean includes three phrase levels above the prosodic word (PWd): the Accentual Phrase (AP), the Intermediate Phrase (iP), and the Intonational Phrase (IP) (Jeon, 2015). Earlier work done by Jun (1996, 2005a) only included the Accentual Phrase and Intonational Phrase, but was later revised to include the Intermediate Phrase (Jun, 2005b).

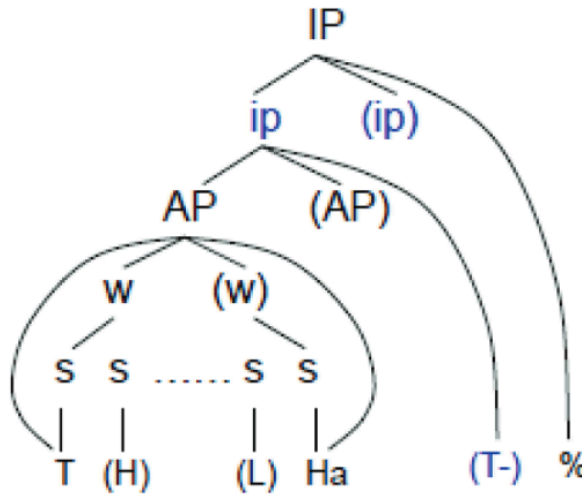


Figure 1. Korean prosodic hierarchy (Jun & Jiang, 2019)

This model of the prosodic hierarchy of Seoul Korean is assumed to be organized following the Strict Layer Hypothesis (Jun, 2005a; Nespor & Vogel, 1986; Pierrehumbert & Beckman, 1988; Selkirk 1986). The prosodic phrases that are relevant in the prosody of reported

speech in Korean are likely APs and IPs, discussed in more detail in Section 2.5. Thus, the following sections discuss the Accentual Phrase, Intermediate Phrase, and Intonational Phrase in greater detail.

2.1.1.1 Accentual Phrase

The Accentual Phrase in the Korean prosodic hierarchy is above the level of the Prosodic Word (PWd), and is the smallest prosodic unit with a contour (Jeon, 2015). It is approximately word sized, which could be a lexical item plus a particle– for example, a noun with a case marker (*koyangi-ka* ‘cat-NOM’) (Jeon, 2015; Jun, 1996). The number of APs within an IP will change depending on speech rate, where faster speech results in fewer APs (Jun, 1996). The underlying tonal pattern of an AP is THLH, where the laryngeal features of the first segment determine whether the initial tone (T) is high or low (Jeon, 2015; Jun, 1998, 2005a). If the phrase initial segment of an AP is aspirated, tense, /s/ or /h/, the contour will begin with a H tone, otherwise it will begin with a L tone (Jun, 2005b). The data in (2) provides an example of a word with an aspirated initial segment, and one without.

- (2) a. *chingkwu-eykey* ... *cwunta*
 friend-DAT give
 b. *yecin-eykey* ... *cwunta*
 Yejin-DAT give

In (2a), if the word *chingkwu-eykey* ‘friend-DAT’ forms one AP, the expected contour would be HHLH, as it begins with an aspirated consonant. In contrast, if (2b)’s *yecin-eykey* ‘Yejin-DAT’ forms one AP, we can predict the contour to be LHLH, as there is no aspirated or fortis consonant in the initial segment of the AP.

A four-syllable AP will realize all the tones of the contour, but the realization of the tonal contour will vary depending on the number of syllables. When an AP has more than four syllables, the first two tones of the AP are associated with the initial syllables of the AP, and the final two tones are associated with the final two syllables of the AP (Jeon, 2015; Jun, 1998, 2005a). Interpolation between the second tone of the AP and the third tone of the AP will give intervening syllables their surface pitch value. The second tone is loosely associated with the second syllable of the AP in APs with four or more syllables, but one or both middle tones of the contour THLH will be undershot when there are less than 4 syllables (Jun, 2005a). The choice of the tone that is undershot varies, but it is not contrastive, making the conditions of undershooting unclear. This means a speaker may pronounce a three syllable AP as LHH or LLH, without a change in meaning between the two contours. Jun (2005a) also notes that the final tone of the AP contour, H, is sometimes realized as L due to constraints or stylistic variation in both long and short APs. In total, Korean has fourteen surface contours for an AP, illustrated in the schematic below:

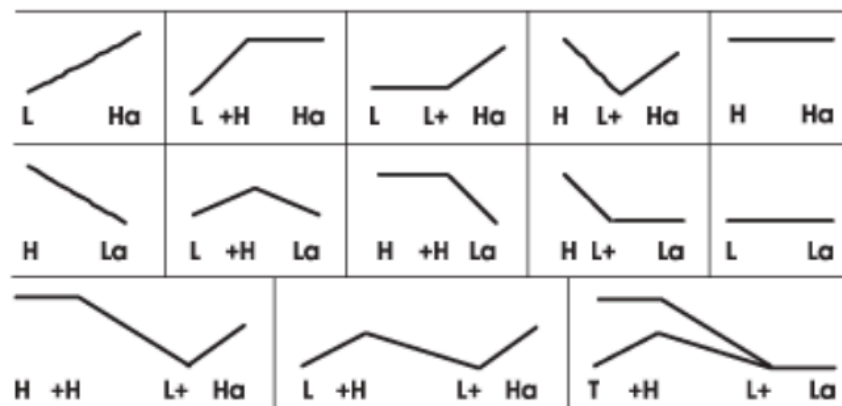


Figure 2. Fourteen tonal patterns of the AP, (Jun, 2005a)

For APs that are IP-final, the final rising tone of the AP contour will be replaced by the boundary tone(s) of the IP (Jun, 1996). APs are typically not lengthened phrase-finally, nor are they followed by a pause (Jun, 2005b). Research conducted by Jun (1996) shows that the AP in Korean is the domain of Lenis Stop Voicing; within an AP (and not across AP boundaries), a lenis obstruent becomes tense after a lenis obstruent.

To conclude, the Accentual Phrase is the smallest unit with an intonational contour, and is higher than the Prosodic Word. The underlying tonal pattern for a four-syllable word is THLH, where T is determined by the laryngeal features of the initial segment. For APs shorter than 4 syllables, medial tones may be undershot, and which tones are chosen to be undershot varies. In APs with more than four syllables, the pitch interpolates between the first and third tone of the contour. IP-final APs will have their final tone preempted by the IP boundary tone(s). The aforementioned fourteen surface contours arise due to the interaction between these numerous factors.

2.1.1.2 Intermediate Phrase

The Intermediate Phrase (iP) is a prosodic unit higher than the AP, but lower than the IP (Jeon, 2015). Jun (2005b) proposed the iP due to the previous structure not distinguishing the differences in focused phrases, and not capturing prosodic cues used in disambiguating ambiguous structures. The iP provides a way to account for focus, prominence, and to mark syntactic constituents. The iP has a greater pitch range and can contain more than one word, which allows it to realize a focused phrase or a complex syntactic phrase better than an AP (Jeon, 2015; Jun, 2005b). There is a pitch range reset in the iP (Jun, 2005b), which interrupts the downtrend of the pitch across the sentence (Jeon, 2015). There is also a larger disjuncture than what is seen with APs (Jun, 2005b). So, when a constituent is narrowly focused in Korean, it will

create a new iP, with a boundary at the left edge (by raising the pitch), while depressing the pitch of the post-focused words (Jun & Jiang, 2019). In addition to narrow focus, words that are prominent can begin a new iP. In such a case, the prominent word shows a pitch range reset and the iP-final syllable does not have any lengthening. When marking syntactic constituents, the right edge is cued by an iP boundary tone (H-, L-) on the final syllable which is also often lengthened.

To sum up the Intermediate Phrase in Korean, a prominent word (due to narrow focus or other reasons for prominence) and syntactic constituents can form a new iP. Intermediate Phrases interrupt the natural declination of a sentence due to pitch reset, can be formed from more than one word, and have a slightly larger perceived disjuncture than that of an AP. Intermediate Phrases created due to prominence mark the left edge boundary, with no lengthening, whereas iPs due to syntactic groupings mark the right edge boundary, and can have slight lengthening of the final syllable.

2.1.1.3 Intonational Phrase

The Intonational Phrase (IP) is the largest unit in the prosodic hierarchy. The contour of the IP is made up of the tonal contours of the APs it contains (Jun, 1996), plus a boundary tone at the right edge of the unit on the IP-final syllable (Jun, 1996, 2005a). There are nine IP-boundary tones (indicated by the diacritic %) regularly observed in speech data (Jun, 2005a), though up to ten have been claimed to exist (Jeon, 2015; Park, 2003). The nine possibilities are: L%, H%, LH%, HL%, LHL%, HLH%, LHLH%, HLHL%, LHLHL%. The IP-final segment is substantially lengthened (Jun, 2005a), and optionally followed by a pause (Jun, 1996). IP-boundary tones may deliver information on sentence type, information structure, or different pragmatic meanings (Jun, 2005a; Park, 2003). For example, declaratives are typically associated

with a final low (L%), whereas interrogatives are typically associated with a final high (H%) (Jun, 2005a). Jeon (2015) states that although boundary tones can have more than one function, informational use is most salient in the monotonal boundaries, whereas in multitonals the affective and/or structural function is more salient. Jun (1996) explains that IP boundaries are more likely to occur at edges of a constituent high in the syntactic hierarchy. Simply put, an IP boundary is more likely to occur at the end of a clause, rather than at the end of a noun phrase. Much like APs, IPs are also sensitive to weight and speech rate—more syllables result in a higher number of IPs, and a faster speech rate will result in fewer IPs (Jun, 1996).

In summary, the Intonational Phrase is the largest unit in the prosodic hierarchy, marked by boundary tones at the right-edge of the unit, phrase-final lengthening, and optionally, a pause. There are nine boundary tones that are regularly found in speech data. IP-boundaries are most likely to occur at the edge of syntactic constituents high in the syntactic hierarchy, such as clauses. Speech rate and weight can affect the number of IPs in a sentence. IP-boundary tones can function as indicators of sentence type, information structure, and pragmatic meanings.

2.2 Reported speech types

When communicating, our speech is not restricted to our own thoughts, feelings or words. In a conversation, one may need to relay an utterance, essentially reproducing the words that had been spoken by another speaker, or even oneself, at another time. This act of referring to or conveying speech or thoughts can be called a quotation, or reported speech (Buchstaller, 2017). To introduce a reported speech segment, many languages use an introductory verb, also called *verbum dicendi* (pl. *verba dicendi*) (Buchstaller, 2017; Holt, 2009). In order to avoid

confusion, I will use a term from Buchstaller (2017) for the “quoter”, referring to them as the reporting speaker. The “quoted” or “quotee” will be referred to as the original speaker.

Reported speech can be divided further into direct reported speech and indirect reported speech (henceforth direct and indirect speech). Another category, free indirect speech, or quasi-direct speech (Holt, 2009; Klewitz & Couper-Kuhlen, 1999), may be referenced in reported speech literature, but it is not a relevant category for the purpose of the current paper. Buchstaller (2017) states that vantage point is one of the main factors in differentiating direct and indirect speech; as direct speech is conveyed from the perspective of the original source whereas indirect speech is conveyed from the perspective of the reporting speaker. Simply put, indirect speech is akin to paraphrasing the reported speech, whereas direct speech is a recreation; pronouns, tense, and all.

(3) a. $I_{i/*j}$ saw $Mary_j$ today, and she_j said “ $I_{*i/j}$ want to go see a movie tomorrow.” [*direct*]

b. I saw $Mary_i$ today, and she_i said that she_i wants to go see a movie tomorrow. [*indirect*]

The information being conveyed in both examples in (3) is essentially the same: Mary wants to see a movie tomorrow. In both examples, the *verbum dicendi* ‘said’ is present, and the original speaker, Mary, is also mentioned. However, in (3a), we assume that Mary’s speech is being replicated as it was uttered by her, demonstrated by the use of the first person pronoun. The first person singular pronoun in the frame does not refer to the same referent as the first person singular pronoun in the direct speech segment. The pronoun in the direct speech in (3a) refers to Mary, who is the original speaker, not the reporting speaker. In (3b), the reporting speaker uses the third person singular pronoun, ‘she’, to refer to the original speaker, Mary.

Compared to (3a), the indirect speech segment in (b) has been adopted into the perspective of the reporting speaker, not the original as in (a).

Another difference between the direct and indirect speech examples in (3) above is the presence of the complementizer ‘that’. Buchstaller (2017) states that the subordination under the speech verb (or *verbum dicendi*) by the overt complementizer ‘that’ is a formal marking of an indirect quote. The example in (3) illustrates the way direct speech and indirect speech differ in English.

In addition to perspective, or vantage point, the assumption of how closely the reported speech is reproduced from the original utterance differs between direct speech and indirect speech. Holt (2009) explains that the speaker claims to reproduce an authentic version of the utterance in direct speech, but does not purport this kind of reproduction in indirect speech. However, the interpretation of the original utterance by the reporting speaker, and the way in which the reporting speaker interacts with the original, both intentional and unintentional, means that the resulting reported speech may not be a faithful reproduction of the original speaker’s words (Buchstaller, 2017). In short, though direct speech appears to be a verbatim reproduction of the original speaker’s utterance, this may not actually be the case, and instead be an approximate reproduction. Neither the degree of faithfulness nor the truthfulness of the reported speech is of interest for this study.

In sum, reported speech is the act of conveying the words or thoughts of an original speaker by a reporting speaker. The two methods of reporting speech that are of interest to this paper are direct speech (DS) and indirect speech (IDS). The DS segment maintains the perspective of the original speaker; pronouns, tense, and other such elements reflect the speech of the original speaker, not the reporting speaker. Though direct speech can be thought of as a

verbatim reproduction of the original speaker, there is no guarantee of the faithfulness of this reproduction. In contrast, IDS can easily be thought of as a paraphrase; the IDS segment is modified to the perspective of the reporting speaker. Both DS and IDS perform the same function of reporting speech, albeit in different ways.

2.3 Reported speech in Korean

In order to investigate the relationship between structural elements that indicate reported speech and reported speech prosody, it is important to delve into the morphosyntax of quotations in Korean. Korean is an agglutinating language, and thus the morphology and syntax of the language are intertwined (Yoon, 2018). Both direct and indirect speech is used in Korean, and in the simplest terms the quotative morpheme attaches to the root verb of the quotation, followed by a quoting verb such as *malhata* “say” or *hata* “do” (Yeon & Brown, 2019). The structure of reported speech in Korean is illustrated below.

(4) Reported speech in Korean

[Reporting speaker: Original speaker] {reported speech}_{-QM}¹ *quoting verb*

↑

Initial frame

↑

verbum dicendi

(modified from Kim, 2012)

The initial frame shown in (4) provides an example of a reported speech segment that includes both the reporting speaker and the original speaker. The original speaker is syntactically optional (Kim, 2012). Additionally, the initial frame may include a recipient, indicating to whom

the reported speech was directed (ex: I said to the cashier...). A more detailed explanation of the reporting frame is included in Section 2.4.2. The quoting verb will be referred to as the *verbum dicendi*. The quotative morpheme can act as a sentence final particle, dropping the *verbum dicendi* in Korean (Sohn, 2015). However, for the purposes of this paper, I will not be investigating data that excludes the *verbum dicendi*.

The quotative marker as it is used today developed from a construction in Middle Korean that used two verbs of saying. Kwon (1998) provides an analysis on the development of the quotative marker from Middle to Modern Korean, illustrated below:

(5) Development of quotative marker *-ko* (adapted and translated from Kwon, 1998)

Quotative verb overlap (hA-ko² malhAta ‘say-and say’)

- > weakening of the descriptive function of ‘hAko’
- > morphological weakening of ‘hAko’ (hAko → kho → ko)
- > universal use of quotative marker ‘-ko’ grammaticalized

Essentially, the quotative verb *hAko* ‘say’ was used in conjunction with another quoting verb, *malhAta* ‘to say,’ causing weakening, and eventual reduction to the marker seen in Modern Korean. This analysis seems to be generally accepted (Ahn, 2006; Rhee, 2007; Rhee & Koo, 2020, for example). Yeon & Brown (2019) describe the grammar of reported speech as the quotative morpheme *-ko* affixing to the plain style ending of the verb in the indirect reported speech segment. The plain style ending varies depending on the sentence type: *-(n)ta* is used for declaratives, *-nya* for interrogatives, *-la* for imperatives, and *-ca* for hortatives. The plain style ending is a part of a group of affixes called Final Endings, or closing affixes, which turn a verbal

complex into a free form (Yoon, 2018). Verbal roots are bound, and require morphemes to turn them into free forms, or fully realized “words.” This is the function of Final Ending affixes. Rhee (2007) and Ahn (2006) analyze *-tako*, *-lako*, *-nyako*, *-cako* as units together, but where Rhee considers them complementizers, Ahn considers them a part of a quotative paradigm. There are a group of affixes that can attach to free form verbs after the Final Ending, called Sentence-Final Particles (SFP), of which the morpheme *-ko* is used for embedding (Yoon, 2018). As the actual nature and specific details of the morphosyntax mentioned here is beyond the scope of this paper, I will adopt an analysis that combines both Rhee (2007) and Ahn’s (2006) analyses, where I will consider the form a quotative complementizer, as it functions as a complementizer while also encoding the meaning of “quote” within it. I will refer to this unit as the quotative marker.

Due to the morphological variation of the suffixes on the verb influenced by sentence type, and the different intonational patterns associated with them (ex: H% for interrogatives), declaratives will be the focus of this study. An analysis of other structures and the relationship with intonation may prove to be an interesting avenue for future research. Table 1 below provides a breakdown of the possible forms being investigated.

Table 1. Quotative markers

	Non Past	Past
Indirect speech	<i>C-nuntako</i> <i>V-ntako</i>	<i>-tako</i>
Direct speech	<i>lako</i> <i>(hako)</i>	

For indirect speech segments, non-past verb roots that end in a consonant will have the quotative marker *-nuntako* (ex: *pat-nuntako* ‘receive-QUOT.COMP’), those that end in a vowel will have the form *-ntako* (ex: *ka-ntako* ‘go-QUOT.COMP’). For direct speech segments, the fully inflected verb will be followed by the quotative marker *lako*, and sometimes *hako*.

2.3.1 Direct Speech vs. Indirect Speech in Korean

The quotative morpheme is part of a paradigm, and the direct and indirect quotative markers have similar forms. However, the indirect forms can indicate sentence types (ex: declarative, imperative, etc), whereas the direct form does not due to the fact it follows a fully inflected verb. So, the way in which they interact with reported speech is slightly different.

- (6) a. Minsu-nun hakkyo-ey **ka-ntako** haysse-yo.
 minsu-TOP school-LOC go.NONPST-QUOT do.PST-POL
 ‘Minsu said he’s going to school’ [indirect speech]
- b. Minsu-nun “hakkyo-ey ka-yo” **lako** haysse-yo.
 minsu-TOP school-LOC go.NONPST-POL QUOT do.PST-POL
 ‘Minsu said “I’m going to school.” [direct speech]

In the above example, (6a) represents indirect speech, and (6b) direct speech. The topic of the sentence is the same for both and precedes the reported speech. In both sentences, the *verbum dicendi* follows the reported speech. Korean often drops the subject (Kwon et al., 2006), so a deictic shift in IDS may not be a reliable indicator for the kind of reported speech being used. Instead, the quotative marker is what differentiates DS from IDS the most readily in Korean. In (6a), the quotative marker attaches to the root verb directly, whereas in (6b) it appears after the

verb has been fully inflected. Essentially, the IDS verb ‘go’ cannot be separated from the quotative marker and remain grammatical; they are a unit. However, the DS verb is fully inflected and a free form, with the quotative marker floating in the vicinity. Technically, the quotative marker *-lako* should be a bound morpheme, and should not stand alone; however it should not be able to attach to a Final Ending affix, such as *-yo* seen in (6b). I have not found any literature providing a syntactic analysis of this phenomenon.

In short, the main structural difference between DS and IDS in Korean is the quotative marker. In IDS, it forms a “unit” with the IDS verb, and cannot be separated. As for DS, the verbal complex and quotative marker form their own “units”, with the quotative marker following the fully inflected DS verb.

2.4 Previous research on reported speech and prosody

Previous research has found acoustic correlates of quoted speech, although the vast majority of languages investigated have been major Indo-European languages. Prosodic cues associated with quoted segments consistently include higher pitch, and pauses before and often after the reported speech segment (Brazilian Portuguese: Oliveira Jr & Cunha, 2004; Beja: Malibert & Vanhove, 2015; English: Klewitz & Couper-Kuhlen, 1999; German: Kasimir, 2008; Hungarian: Bermúdez, 2023; Modern Hebrew: Malibert & Vanhove, 2015; Polish: Bermúdez, 2023; Spanish: Bermúdez, 2023; Estellés-Arguedas, 2015; Swedish: Bermúdez, 2023). Additionally, lengthening of the segment and higher intensity have also been reported in Spanish, though Bermúdez (2023) proposes that these two cues may serve as a means to convey irony. Studies done on the prosodic correlates of reported speech have investigated both direct and indirect reported speech segments.

2.4.1 Mimicry and Pitch Variation

Prosody can be used as a tool to help convey messages to a listener, such as disambiguating structurally ambiguous sentences or focusing elements of an utterance (Hirschberg, 2002; Wagner & Watson, 2010). When it comes to reported speech, one may imagine changes to the reporting speaker's voice, such as a change in pitch or voice quality, may simply be attributed to the reporting speaker mimicking the original speaker's tone. In her paper of the recontextualization of reported speech in German, Günthner (1999) analyzed the way in which speakers use prosody and voice quality techniques to convey the speech of multiple characters. She makes a distinction between prosody and voice quality—prosody refers to loudness, duration, pitch, and pause, whereas voice quality refers to paralinguistic cues that are used by the speaker to produce different types of voices (ie. whispery, falsetto, breathy, etc) temporarily. Thus we can understand that there are multiple factors that accompany reported speech; prosody may be accompanied by the affectation of speech to mimic some quality of the original speaker, or convey some opinion that the reported speaker has of the original speaker. For instance, a line from one of the dialogues analyzed by Günthner showed a direct speech segment which was produced with a tense, high pitched voice, accompanied by an increase in volume, which gave the utterance a “hysterical and exaggerated subtone” (Günthner, 1999). Though some overlap may occur, it is entirely possible to dissect prosody from paralinguistic elements of reported speech. I am chiefly concerned with the prosodic cues of reported speech.

While investigating the role of prosody in reported speech in Spanish, and whether evidential meaning could be expressed purely through prosody, Bermúdez (2023) consistently found a raise in pitch in the reported speech segment. He further extended the study to include Polish, Swedish, and Hungarian, finding similar results in these languages. Based on his

findings, Bermúdez (2023) stated that a change in voice quality is not just mimicry, as participants in his study raised the pitch of a reported speech segment, regardless of the gender of the original speaker. Bermúdez posits that language embodiment may explain this choice, where hearer-oriented speech acts correlate to higher pitch, and speaker-oriented speech acts correlate to lower pitch. Essentially, he is saying that when a person focuses on themselves, they tend to look down to look at their bodies, and lift their head when told to focus on something outside themselves. “Down” represents self, and “up” represents things external to self. This is then reflected through pitch, possibly explaining the use of a higher pitch with reported speech segments (excluding self-reported cases), though this is purely hypothetical (Bermúdez, 2023).

Estellés-Arguedas (2015) results seem to contradict this idea, as she found both a raise and lowering of pitch over the direct speech segment in her study of the marked prosody of direct speech in Spanish. In addition, Klewitz & Couper-Kuhlen (1999) also found pitch changes in English, either a raised pitch or a lowered pitch over the reported speech. This difference could be accounted for by the differences in methodology— the studies by Estellés-Arguedas and by Klewitz & Couper-Kuhlen looked at a corpus of conversational data, from Spanish and English, respectively, whereas Bermúdez’s speakers participated in a production task. Both the raising and lowering of pitch over reported speech segments may be more easily captured in natural speech data from a corpora of conversations, rather than from a production task. Regardless, changes in pitch seem to be a cue used to indicate reported speech. Oliveira Jr & Cunha (2004) found that in Brazilian Portuguese, of all the cues marking direct speech boundaries, pitch reset was the most significant. Malibert & Vanhove (2015) did an analysis of four Afroasiatic languages, finding that pitch changes were present in two of the languages (Beja and Modern Hebrew) as an

indicator of reported speech. In all, pitch variation consistently appears as a prosodic cue of reported speech.

2.4.2 Frames, Boundaries, and Pauses

In the orthography of many languages, quotation marks are used in written texts to mark the boundaries of direct speech. In verbal communication, prosodic cues can work as “flags” to prosodically frame reported speech segments (Klewitz & Couper-Kuhlen, 1999). This idea of framing, or marking of boundaries, is investigated in Brazilian Portuguese (Oliveira Jr & Cunha, 2004), English (Klewitz & Couper-Kuhlen, 1999), Russian (Bolden, 2004), and German (Kasimir, 2008). Malibert and Vanhove (2015) also describe framing and boundaries of reported speech in Beja, Zaar, Juba Arabic, and Modern Hebrew. But what is a frame? Bolden (2004) explains that a reporting frame is a device that sets off the reported speech from current speech, and this framing is done using a framing clause or a quotative of the form “*speaker + reporting verb*”. The term reporting verb used here by Bolden is the same as the *verbum dicendi* mentioned prior. The reported speech segment is minimally one unit, which can be a word, phrase, or clause, and the shift from the reporting frame to reported speech can include prosodic cues (Bolden, 2004). The following example provides a breakdown of the frame in English.

(7) [FRAME] {*reported speech*}

Where **FRAME** includes subject, *verbum dicendi*, and optionally ‘that’

→ [MARY SAID] {*I will leave at 3pm*} (direct speech)

→ [MARY SAID THAT] {*she will leave at 3pm*} (indirect speech)

As we can see in (7), the frame in English precedes the reported speech, and will include the subject, *verbum dicendi*, and sometimes the complementizer ‘that.’ Both direct and indirect speech use the frame. This example demonstrates the structure of English, but other languages may not follow the same pattern (Beja (Malibert & Vanhove, 2015) for instance). As seen in (4) in Section 2.3, Bolden’s (2004) description of the reporting frame does not fully align with what is seen in Korean.

Though Bolden (2004) uses the terms “reporting frame” and “quoting frame,” which may be mistakenly understood as being interchangeable. To avoid confusion, the term “frame” will only be used to refer to elements outside of the reported speech segment. Depending on the language, the terms “initial frame” may be used to refer to the frame preceding the reported speech, and “end frame” to refer to the frame following the reported speech.

Pauses are a cue often found in research on reported speech prosody. In a conversation, prosody is used to cue the next speaker’s turn to talk (Bermúdez, 2023; Hirschberg, 2002; Hirschberg et al., 2020), and so pauses employed at reported speech boundaries could cue a “metaphorical shift in speakers” (Bermúdez, 2023). Klewitz & Couper-Kuhlen (1999) found that in English, pauses are used to flag the left-hand boundary of a reported speech segment—essentially breaking the initial frame from the reported speech. In their research into the use of prosody in marking direct speech boundaries in Brazilian Portuguese, Oliveira Jr & Cunha (2004) found that pauses at the end of direct speech are typically longer than pauses found elsewhere, and a low boundary tone is more often used at the end of the direct speech. Bermúdez (2023) found that pauses before the reported speech segment, and often after, were commonly used in Spanish, as well as in Swedish, Polish, and Hungarian, though the pause following the reported speech segment did not appear to be employed in Polish or Hungarian. Pauses set off

the initial frame in Beja, and pauses can set off the end of the reported speech in Zaar and Juba Arabic (Malibert & Vanhove, 2015).

Pauses are a cue often seen in languages at the boundaries of reported speech. Often, they appear to help set off the frame for the reported speech, and seem to mimic the way in which pauses are used in turn-taking in dialogues amongst two or more speakers.

2.4.3 Additional Cues

Pitch changes, pauses, and boundary tones appear to be common cues of reported speech in a variety of languages. Yet, there are cues that may not be as widespread. Klewitz & Couper-Kuhlen (1999) found that a change in volume and changes to the rhythmic pattern often accompanied pitch changes, as cues of direct speech in English. Similarly, Kasimir (2008) found that German speakers employed a shift in the location of pitch accents and change in voice quality (such as the lengthening of vowels), to indicate the reported speech segment in her study on subclausal quotations. These kinds of rhythmic changes may not be useful tools in some languages; for instance, a lengthening of vowels may not be employed in a language with contrastive vowel length. Decrease in speech rate was found in Spanish by Estellés-Arguedas (2015) as well as by Bermúdez (2023). That being said, Bermúdez proposes that the slower speech rate may convey irony, rather than an association with reported speech. Oliveira Jr & Cunha (2004) reported a drop of intensity at the boundary of direct speech segments in Brazilian Portuguese. Additionally, Oliveira Jr & Cunha find that boundary tones play a part in framing, sometimes used in conjunction with pauses. The use of prosodic cues do not seem to be conventionalized, and they are not systematically used to signal reported speech (Klewitz & Couper-Kuhlen, 1999).

2.4.4 Conclusion

Pitch variation and pauses were the two most consistently described cues in research on the prosody of reported speech. In addition, rhythmic changes, lengthening, and changes in volume and intensity were also noted as cues in some languages. However, not all reported speech is marked prosodically (Bolden, 2004; Klewitz & Couper-Kuhlen, 1999). Klewitz & Couper-Kuhlen (1999) state that prosodic cues do not need to be sustained over the entire reported speech segment. This means that unlike quotation marks used in text, prosodic cues seem to be optional devices, and according to Klewitz & Couper-Kuhlen (1999), the non-use may be a stylistic choice. So, prosody may or may not be present to cue reported speech, but when there is no structural indicator (such as a *verbum dicendi*) for reported speech, prosodic cues are often present. In many languages, it appears that cues are present at the onset of the reported speech, “setting off” the frame from the quotation. It appears that the “unquote” is less important, as less cues, or weaker cues, are found at the end boundary of the reported speech segment.

2.5 Research questions

To my knowledge, the prosody of reported speech in Korean has yet to be explored. The goal of this paper is to investigate the ways that prosody is used to indicate reported speech in Korean, and how structural elements will interact with intonational elements in an utterance containing reported speech. How will prosodic cues be employed to indicate quoted speech, and to what extent? The structure of Korean differs from many of the languages researched in the literature, as the reported speech segment is embedded within two frames: an initial frame and an end frame. The initial frame can be optional, however as I am interested in both boundaries of

reported speech in Korean, I will not be analyzing any examples that exclude the initial frame. The end frame minimally includes the *verbum dicendi*, and it is yet unclear if the quotative marker will be included with it. Korean diverges from most of the languages in the literature as the *verbum dicendi* does not appear before the reported speech, but follows it. Furthermore, the end of the reported speech is indicated by a structural marker. To what extent will this structural marker affect prosody? Malibert and Vanhove's (2015) research on Afroasiatic languages is interesting to compare with Korean, as Beja (North Cushitic) has a similar type of framing. In Beja, if the subject or recipient is overtly expressed, it precedes the direct speech, and the *verbum dicendi* follows the direct speech. Malibert and Vanhove found that in Beja, direct speech is not set off from the *verbum dicendi* by intonational boundaries, and it often cliticizes to the direct speech. Though unrelated to Korean, Beja offers an idea as to how the end frame may be treated in IDS examples in Korean.

The presence of pauses is also of interest, as they are commonly found at the onset of reported speech, and often after, in much of the literature. As Korean signals the end of reported speech structurally, will pauses ever be employed? If so, where will they occur? The quotative markers in direct speech and indirect speech vary slightly, where the quotative marker directly affixes to the indirect speech segment, but does not in direct speech. I predict any end frame pauses to occur before the quotative marker in direct speech, but between the quotative marker and *verbum dicendi* in indirect speech. This also leads to a prediction about the frame— I hypothesize that the end frame in direct speech includes the quotative marker and *verbum dicendi*, but only the *verbum dicendi* in indirect speech. This would then suggest that for indirect speech, the quotative marker and the final verb of the segment would form a prosodic unit, likely an AP. In contrast, the final verb in a direct speech segment and the quotative marker would be

separate prosodic units, again, likely APs. Direct speech is reproduction of an utterance, which leads me to believe that it should form its own IP. This would be an IP embedding within another IP, potentially reflecting the recursivity of the syntax. The analysis of Korean prosody has so far assumed that the prosodic structure follows strict layering (Jun, 1996, 1998, 2005a, 2005b for example), disallowing recursivity. What if a recursive analysis can better account for the phenomena observed in direct speech examples (for instance, following the proposal in Selkirk, 2011)? Reported speech provides an interesting opportunity for a recursive analysis of prosodic structures. As reported speech, specifically direct speech, is syntactically an utterance embedded into an utterance, my assumption is that the direct speech utterance has its own IP, as stated in above. In contrast to the Strict Layer Hypothesis (Nespor & Vogel, 1986; Pierrehumbert & Beckman, 1988; Selkirk, 1986), where recursivity is disallowed, Selkirk (2011) proposes Match Theory to account for cases in which a prosodic category dominates another of the same category (i.e. a recursive structure). Match Theory assumes a set of MATCH constraints requiring the correspondence of syntactic and prosodic structures. Other constraints, such as markedness constraints, may outrank the MATCH constraints (as under an Optimality Theoretic framework), which can account for differences between the syntactic and prosodic structures. Match Theory, and how it may be used to analyze reported speech prosody in Korean, is addressed in more detail in Chapter 4. The goal of this paper is to provide a preliminary description of reported speech prosody in Korean, and, in turn, provide and answer to these questions.

Chapter 3: The prosody of reported speech in Seoul Korean

3.1 Methodology

Much of the previous research conducted on reported speech has involved the analysis of corpora of spontaneous speech (for instance Klewitz & Couper-Kuhlen, 1999), though some included reading tasks (as in Bermúdez, 2023). As my goal is to provide a descriptive analysis of the prosody of reported speech in Korean, I opted to conduct three production tasks in order to elicit reported speech from the participants. I designed three tasks, one of which was a reading task completed in pairs, and the other two were production tasks completed by each participant individually. The reading task provided control over the sentences used across speakers, allowing them to be compared more readily. This means that the target structure, with an initial frame, reported speech, and end frame, as well as the type of reported speech were all factors controlled in the reading task, but not the two production tasks. The two production tasks, completed individually, were designed in order to elicit more natural speech, allowing the participants some freedom in choosing what they wanted to say.

The tasks were conducted in Korean, and all materials provided to the participants were in Korean. All examples in this paper have been transliterated from the Korean script, Hangul, based on the Yale system of Romanization. (See Appendix A for a table listing the letters corresponding to the Hangul characters, as well as the corresponding IPA symbols.) The annotation of the data was done following the K-ToBI labelling conventions (Jun, 2000). The Tones and Break Indices (ToBI) system is used to transcribe phonological properties and the degree of juncture between words, and was originally designed for the transcription of English (Jun, 2022). It was then developed for other languages, and K-ToBI, or Korean Tones and Break

Indices, was created (Jun, 2000). (See Appendix B for a description of the tones relevant to the data.)

3.1.1 Participants and Tasks

The six participants are native speakers of Seoul Korean. All the participants currently reside within the Seoul metropolitan area, and did not grow up speaking any other variety of Korean, to ensure that there was no influence of a pitch-accented variety of Korean on their speech. All six speakers are female, with an age range of 26 - 61 and a median age of 29.

Table 2. Participants and ages

Participant	Age
CJR	27
CJW	61
CSJ	36
GBJ	30
KHJ	28
OHJ	26

The tasks were all conducted and recorded through Zoom, with only one pair of participants meeting face-to-face with each other for the reading task. Two participants completed the tasks in two separate sessions, and the other four participants completed the tasks in one session. The audio data was collected from the audio recordings of the Zoom meetings, and analyzed using Praat (Boersman & Weenink, 2024). The materials for the following tasks are included in Appendix C.

3.1.1.1 Task one: Dialogue

The first task, which was completed in pairs, consisted of two dialogues that included instances of both direct and indirect speech. The dialogues mimicked two casual conversations such as one may have with a friend. Participants were given the context of the dialogue, and assigned a role A or role B. The dialogue was provided on the screen, and participants were given an opportunity to read through the dialogue to become familiar with it. When ready, the participants were asked to read the dialogue as naturally as possible. In each dialogue, there were six instances of reported speech, three direct speech segments and three indirect speech segments, and each participant had three lines with reported speech. In dialogue 1, role A had two lines with indirect speech, and one line with direct speech. Role B had one line with indirect speech, and two lines with direct speech. In dialogue 2, role A had one line with indirect speech, and two lines with direct speech. Role B had two lines with indirect speech, and one line with direct speech. In short, between both dialogues, role A and role B both had three lines of direct speech and three lines of indirect speech each.

3.1.1.2 Task two: Q&A

The second task was a question-answer task. Participants were provided with a short paragraph that they were asked to read, and then were asked to answer four questions related to the paragraph. The paragraph did not include any reported speech. Whether the answers to the questions were correct or not was unimportant, but in order to avoid hesitation from the participants, two questions were posed verbally halfway through the text, and the final two were posed at the end. The questions were such that the participants could use reported speech in the answer, though eliciting specifically direct or indirect speech was difficult. Up to four instances of reported speech were possible for each participant. The participants were informed of the

structure of the task. They were then provided the text on screen, and asked to read. At the break in the text, participants were verbally asked the first two questions. Once answered, the participants were provided the second half of the text, and asked to continue reading. At the end of the text, participants were verbally asked the final two questions.

3.1.1.3 Task three: Fill-in-the-blank

The third task was a fill-in-the-blank task. Participants were provided a short paragraph in which certain sentences had a blank space marked by an underline (ie: I said _____). The surrounding sentences provided a context in which the participants had a limited choice of answers. For example, the sentence “I told my friends the hot water wasn’t working at my apartment, and asked them if they were okay with that” was followed by a fill-in-the-blank sentence “My friends said _____.” This allowed participants the freedom to insert their own utterance, while attempting to limit pauses and other similar hesitations from the participants. There was a chance for participants to produce five examples of reported speech with this task. Participants were provided an explanation of the task, and the text was provided on screen. The participants were given a chance to read through the text in order to mentally plan what they would insert into the blanks. This was done in an attempt to avoid disfluencies stemming from participants deciding on their answers instantaneously while reading. They were then asked to read the text, including their additions to the blanks.

3.2 Results

3.2.1 Methodology

In addition to describing the prosody of reported speech in Korean, this study also explored different experimental designs used to elicit reported speech, in order to test how successful they were in producing reported speech examples. The goal of this study was to describe the prosody of a specific structure, and thus collecting data from controlled experiments provided the opportunity to elicit reported speech that included the target structure.

3.2.1.1 Dialogue Task

The reading, or dialogue, task was a way in which instances of both indirect and direct speech could be collected in a controlled manner, and the same utterances could be compared across speakers. One of the potential drawbacks of a reading task is the use of a reading intonation, rather than intonation used in natural speech. For two pairs of participants, the reading task resulted in slightly stilted, unnatural speech, to varying degrees. Participants CJR and CSJ were the most natural, though audibly sounding as though they were reading. Participants CJW and GBJ were the least natural; participant GBJ maintained a monotonous voice and CJW ended each sentence in a raising pitch, diverging from their normal speech. Though the participants were given a chance to read over the dialogues as many times as they needed before beginning the task, some people may not be comfortable in performing a dialogue. Therefore, although the benefit of this task is the potential to analyze the same utterances across different speakers, some of the data proved to be unusable.

One pair of participants deviated from the script during the reading task. Rather than reading the dialogues, participants KHJ and OHJ modified the script in such a way that it flowed

like a natural conversation, including interjections, and making changes to sentences or switching out lexical items or grammar. Their conversation most closely mimicked natural speech, however some of the reported speech segments were unusable as they differed from the specific structure being analyzed for this research.

As the reading task was fully scripted, all the participants produced the target tokens, with the exception of participants KHJ and OHJ. While performing the reading task, participant KHJ modified the reported speech in the script so that the majority of her lines were indirect speech, with a single case of direct speech that did not include a quotative marker and thus excluded from the data. Participant OHJ consistently dropped the marker *-ko* from the speech, which was then followed by *kulayssta* ‘say so/like that.’ An interesting avenue for further research would be exploring the relationship of *-ko* and *kulayssta* ‘say so/like that’, and whether speakers drop the marker *-ko* when *kulayssta* acts as the *verbum dicendi*.

3.2.1.2 Q&A

The second task attempted to elicit reported speech through a question and answer task based off of a text. This task produced the lowest amount of reported speech, where one participant produced no instances of reported speech and the maximum produced was two, by just two participants. The remaining three participants only produced one instance of reported speech from this task. One question was designed to try to elicit direct speech by beginning the question with “In Yejin’s words...” however it did not produce the intended effect. The question that was most successful in eliciting reported speech was: “Is Yejin looking forward to the weekend? What words indicate this?” None of the questions included the quotative marker. Including the target structure in the question may have elicited more reported speech in the

answers as the participants did mirror the structure of most of the questions in their reply. For example, one question posed often elicited the same grammar repeated in the answer.

(8) Q: yeycin ssinun mallo, yeycin ssinun chingkwu senmwullo mwuesul **salyeko**

Yejin HON words Yejin HON friend present what buy

haysseyo

do

“In Yejin’s words, what present is Yejin going to buy her friend?”³

A: saciney kwanhan chaykul **salyeko haysseyo**

picture about book buy do

“(She) is going to buy a book about photography”

The bolded items show the mirrored structure in the question and answer. There is a possibility that participants would have produced more reported speech had the questions also included the reported speech structure. Furthermore, this task often resulted in some hesitation from the participants. Even though the text was divided in order to ask questions throughout, the participants often paused while answering. Had the questions more directly referenced the text, reported speech may have been used more often. The participants may have been more concerned with answering the question properly, rather than answering in full, fluent sentences.

3.2.1.3 Fill-in-the-Blank

The fill-in-the-blank task was the most successful task in the elicitation of reported speech. Out of a possible five instances to produce reported speech, two participants produced four instances of reported speech, and four participants produced five instances of reported speech. A drawback to this method is the limitation in length. In order to prevent disfluencies

due to participants being unsure how to fill the blank space in the sentence, the task is limited to a short paragraph that a participant could feasibly remember after reading it over a few times. A longer paragraph could lead to hesitation and pauses before the blank as the participant plans what to say. For a study such as this, the fill-in-the-blank was extremely useful. It provided a variety of answers from a short text, while also having a somewhat controlled script. How the participants decided to complete the sentence was up to them, but the context provided a limited choice of acceptable answers. For instance, all of the fill-in-the blank sentences specified who the original speaker was. Based on the preceding sentence, the fill-in-the-blank provided some direction. If the preceding sentence stated that the hot water wasn't working, and they contacted the maintenance man, it was clear that the blank in the following sentence should recount what was said to the maintenance man about the hot water issue.

3.2.2 Direct and Indirect speech comparisons

The number of indirect speech examples produced by the participants was almost double that of the direct speech examples. Furthermore, the types of prosodic cues used with direct speech was more varied than those used with indirect speech. Table 3 shows the number of sentences produced by each speaker, how many instances of direct or indirect speech was produced, and the number of sentences analyzed.

Table 3. Number of sentences produced and analyzed

Participant	# of sentences produced		# of sentences analyzed	
	DS	IDS	DS	IDS
CJR	5	5	5	4
CJW	5	7	2	2
CSJ	5	7	3	5
GBJ	4	8	2	5
KHJ	2	9	2	3
OHJ	2	9	2	3
	23	45	16	22

As this paper is interested in both the onset and offset of reported speech, sentences that omitted the initial frame were excluded. Occasionally, participants omitted part of the quotative, *-ko*; most notably participant OHJ, who dropped *-ko* in the majority of her indirect speech examples. In some sentences, participants used the imperative or hortative in their reported speech segments, which were also excluded. Participant CJW had many sentences excluded due to long pauses caused by hesitation while performing the fill-in-the-blank task.

Prosodic cues co-occurred often in the data. However, not all prosodic cues consistently occurred together at the same boundaries. In order to organize the results concisely, I will define boundaries as either strong or weak. A strong boundary is the co-occurrence of a boundary tone plus another cue, though an H tone boundary requires a pause to be considered a strong boundary. This is due to the typical contour of an AP typically ending on a H tone; as pauses are not predicted to occur at an AP boundary it is the only cue that differentiates the H tone of an AP from some other potential phrase. A weak boundary tone is predicted to have one prosodic cue or less. This means a weak boundary may be marked by a boundary tone other than H, a pause, or

lengthening, but not the co-occurrence of prosodic cues. A table illustrating the co-occurrence of prosodic cues is available in Appendix D. Sections 3.2.3 and 3.2.4 will look more closely at the co-occurrence of prosodic cues at prosodic boundaries for each type of reported speech, and Section 3.2.5 provides a breakdown of the number of prosodic cues occurring in direct and indirect speech.

3.2.3 Direct speech

The results show that the participants used prosodic cues to mark the boundaries of the direct speech segments. There is evidence that both the beginning and end of the direct speech segment were indicated using these cues, though to varying degrees. The onset of direct speech was clearly and consistently set off from the initial frame, whereas fewer cues were used to flag the end of the segment. Two patterns emerged in terms of the final tone in the initial frame—some participants had a falling tone (either LHL or HL), others had a high or rising tone (H or LH). The falling tones and the rising tone are consistent with IP boundary tones and the diacritic % will be used for these instances. Some boundary tones are not marked by a diacritic, as it is unclear what kind of boundary it is, such as the H tone in Figure 5. In addition to the boundary tones, pauses and lengthening were also used to set off the direct speech from the initial frame.

The rising-falling or falling boundary tones were almost always followed by a pause, with the exception of two examples where it was unclear whether the silence was an intentional break or simply the beginning of the articulation of the stop consonant that followed. The rising or high boundary tones were optionally followed by a pause before the direct speech segment. Participant CSJ employed the H tone in all her direct speech examples and had an audible break following the boundary tone. Another participant, OHJ, had one example of the frame ending in a rising tone followed by an audible pause, in addition to lengthening the final segment of the

frame, illustrated in Figure 3. For clarity, I will only be marking the boundary tones, or relevant contours, in the following figures.

Figure 3 indicates an example of direct speech from participant OHJ, where the initial frame ends in a rising tone LH%.

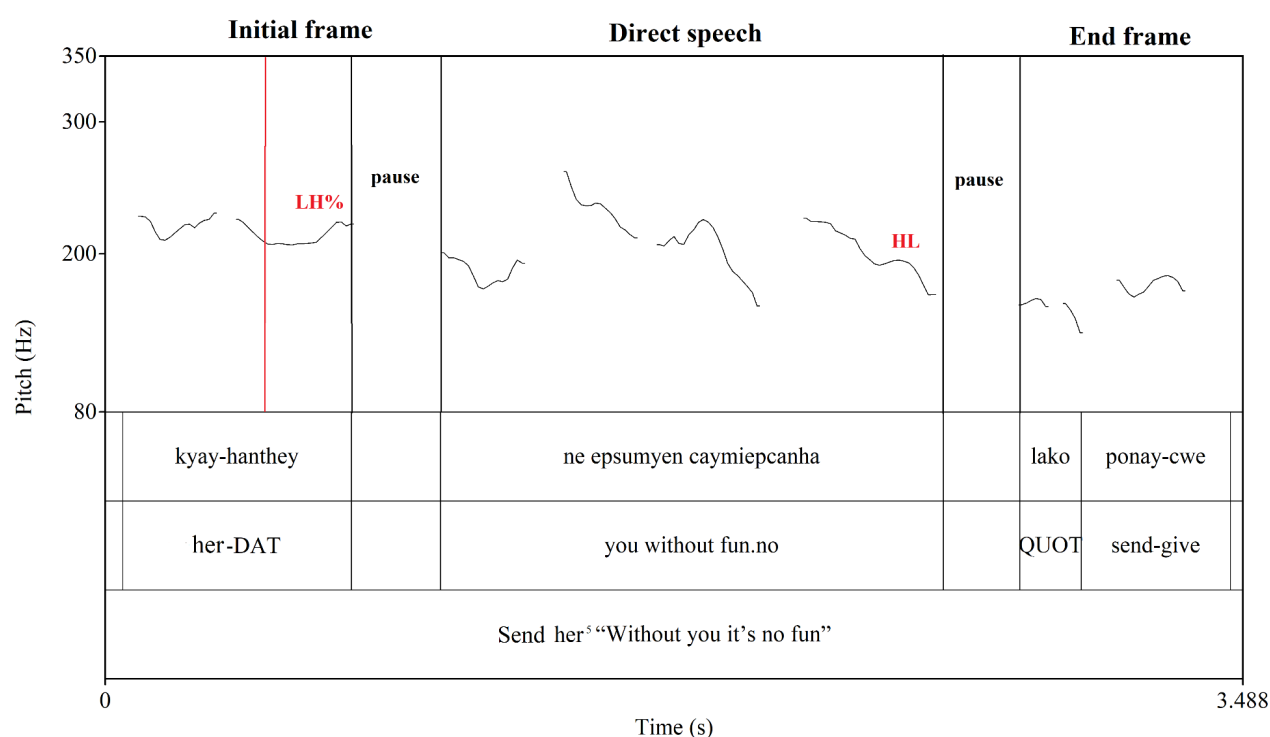


Figure 3. OHJ direct speech⁵

The red line in Figure 3 shows the start of the final syllable in the initial frame. The final syllable has a rising tone, where it starts low and raises within the syllable, near the end. In addition, the segment is audibly lengthened and followed by a pause. The pattern of LH%, along with the pause and lengthening, is consistent with an IP boundary. The rising tone could be a continuation rise (Féry, 2016). As for the end boundary of Figure 3, the end of the direct speech segment is marked by the boundary tone HL, followed by a pause. There was no lengthening on the final syllable of direct speech.

Initial frames ending in a falling tone were typically accompanied by audible pauses. In addition to the falling boundary tone and the pauses, the final segment of the initial frame was often lengthened as well. Figure 4 is an example of direct speech, with an initial frame boundary marked by a rising-falling tone.

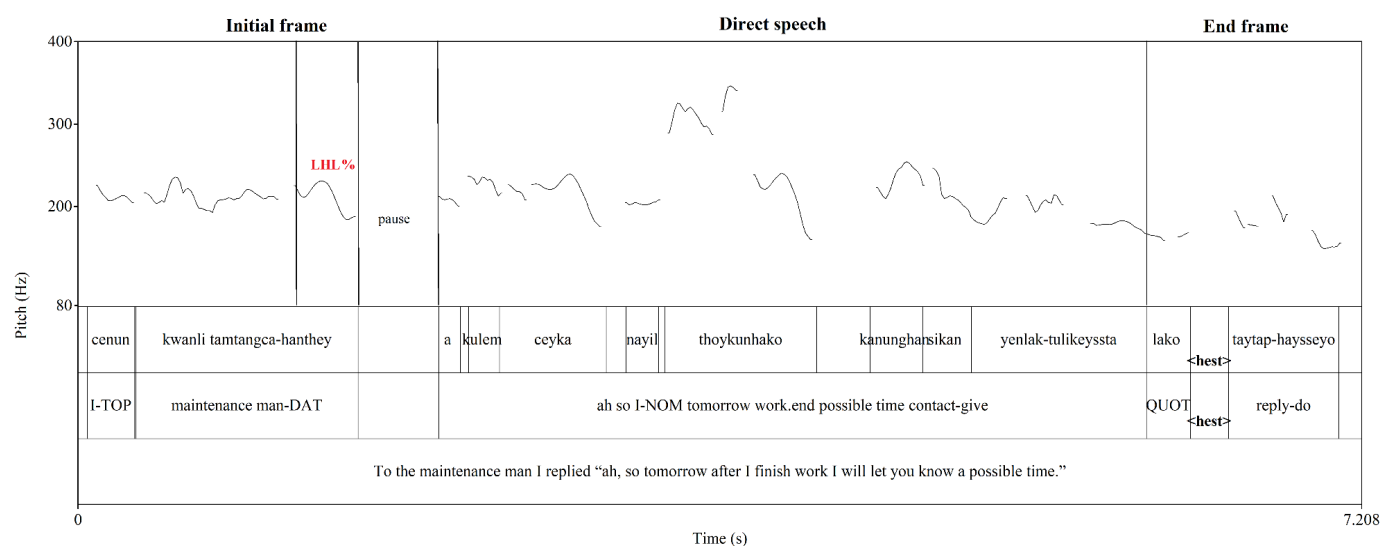


Figure 4. KHI5 direct speech

The pitch track in Figure 4 provides an example from participant KHI, who consistently used a rising-falling (LHL) boundary tone at the edge of the initial frame followed by a pause. She also tended to audibly lengthen the final segment of the initial frame, though this may not be clear in Figure 4 as it is a long pitch track. Participants varied with the amount and type of prosodic cues they employed, where participants such as OHJ and KHI not only used boundary tones, but also lengthening and pauses. It is important to note that in Figure 4, the break between the quotative and the *verbum dicendi* is not a deliberate pause, but a hesitation. The direct speech boundary will be discussed next.

The boundary between the direct speech segment and the end frame was much weaker than that of the initial frame; only one or two cues were present for most speakers. Pauses were scarcely used, but lengthening and a low boundary tone were often present. The most consistent indication of the end boundary was lengthening of the final syllable of the direct speech segment.

Figure 5 illustrates a direct speech example produced by participant CJR, with the initial frame ending with a H tone, lengthening of the final syllable of the direct speech segment, and no pauses at either boundary.

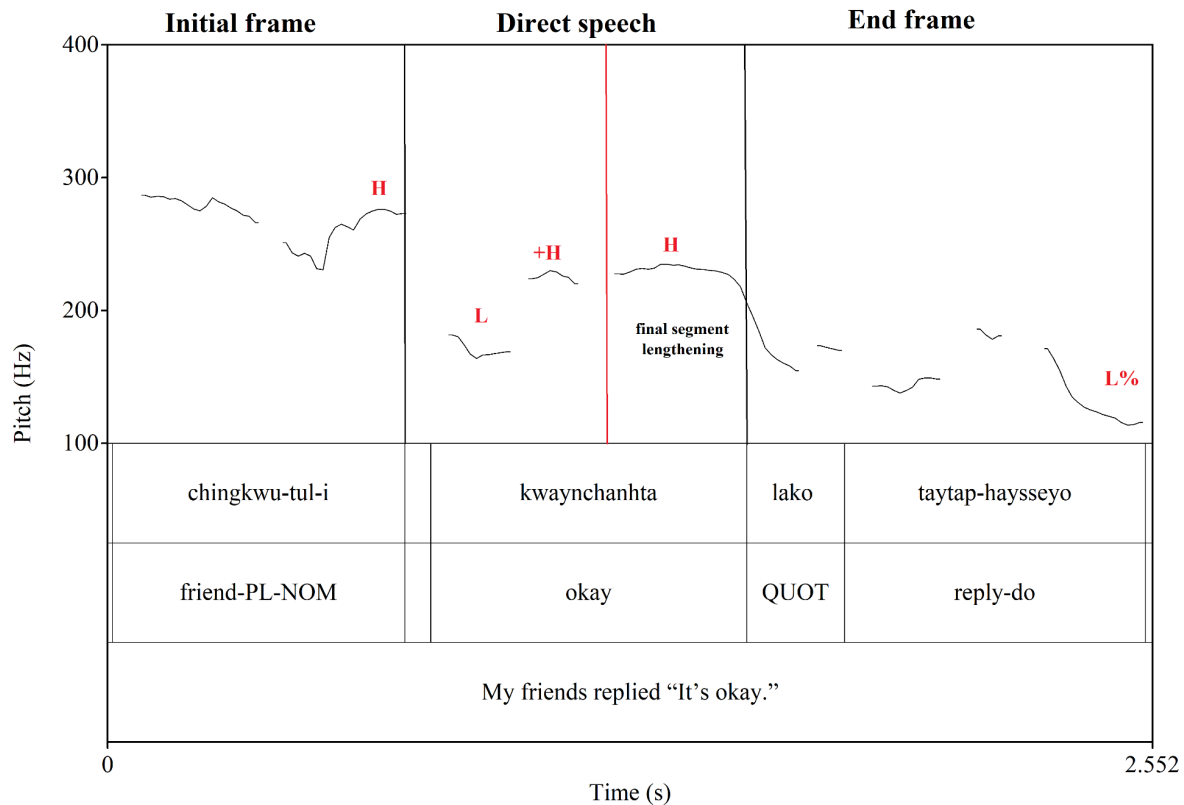


Figure 5. CJR3 direct speech

The red line indicates the beginning of the final segment in the one-word direct speech example of Figure 5. The contour of the direct speech segment is LHH. Though a H tone marks the right edge of an AP, the lengthening seen in the final syllable of Figure 5 suggests that it is not; APs are never lengthened.

The fact that the end of a quoted segment is indicated structurally by the quotative marker may have influenced how strongly speakers felt the need to mark the “unquote” prosodically. The end of the direct speech segment often ended low, as seen in Figure 3, however some examples ended high, as in Figure 5. In all cases, the *verbum dicendi* was pronounced with the lowest pitch of the utterance, at the low end of the participant’s pitch range, sometimes resulting in a creaky voice, particularly with participant CSJ as shown in Figure 6. Occasionally, the *verbum dicendi* seemed to be produced at a faster speech rate than the rest of the sentence.

Figure 6 is a direct speech example produced by participant CJR, highlighting the *verbum dicendi*.

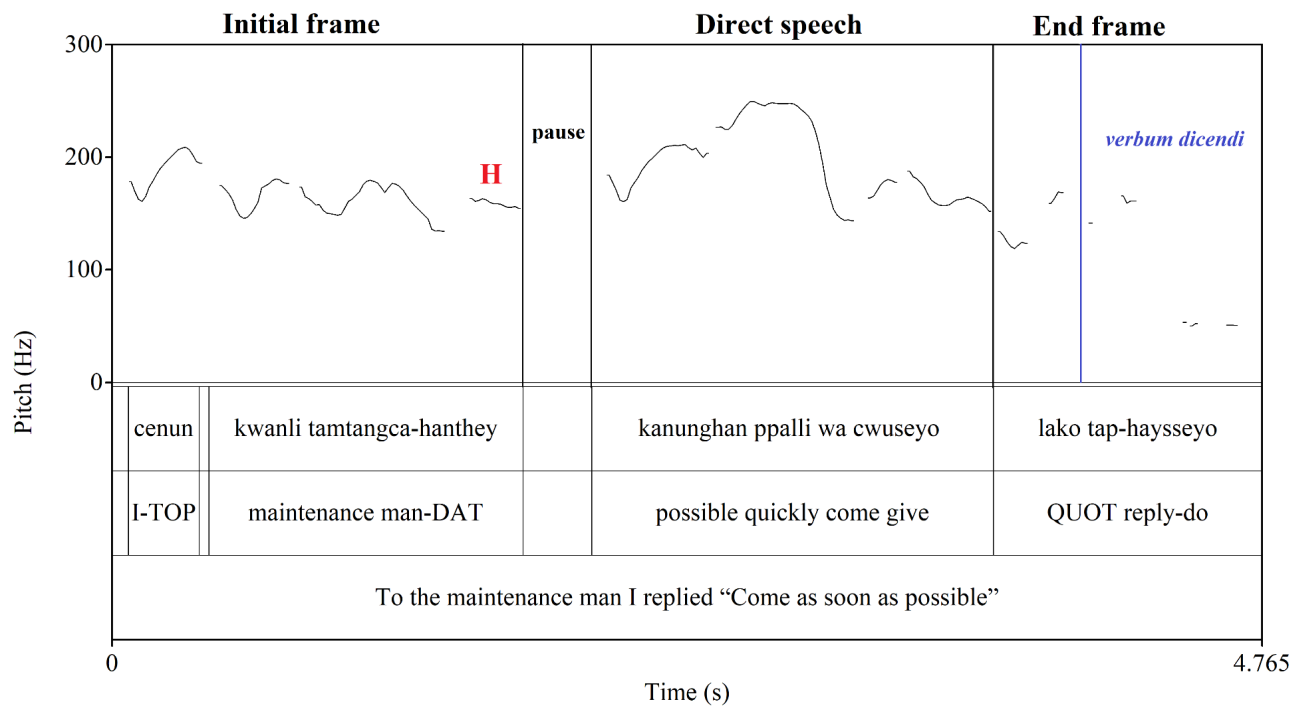


Figure 6. CSJ5 direct speech

Figure 6 above shows that the *verbum dicendi* is produced with a creaky voice, with the last two syllables at the bottom range of the speaker's pitch. The quotative was clearly pronounced and the creaky voice progressed over the *verbum dicendi*.

The direct speech segment also appeared to undergo pitch reset. It is not the most reliable prosodic cue as Korean intonation is sensitive to segment type, and the presence of fortis and aspirated consonants (in addition to /s/ and /h/) triggers a H tone (Jeon, 2015). For many examples, the first prosodic unit of the direct speech segment (aligning with the first H tone of the initial AP contour) had the highest pitch of the sentence. Whether pitch reset occurred or not was more difficult to ascertain when the direct speech included one of the segments that triggered the H tone. This can be seen in Figure 6, where the highest pitch of the utterance is within the direct speech. This is due to the fortis consonant in the word *ppalli* 'quickly' (represented by <pp>). Therefore, pitch reset cannot be relied upon as the sole indication of a reported speech segment.

Figure 7 is a direct speech example produced by participant CJR and illustrates possible pitch reset of the direct speech segment, comparing it to the high pitch of the aspirated consonant present in the initial frame.

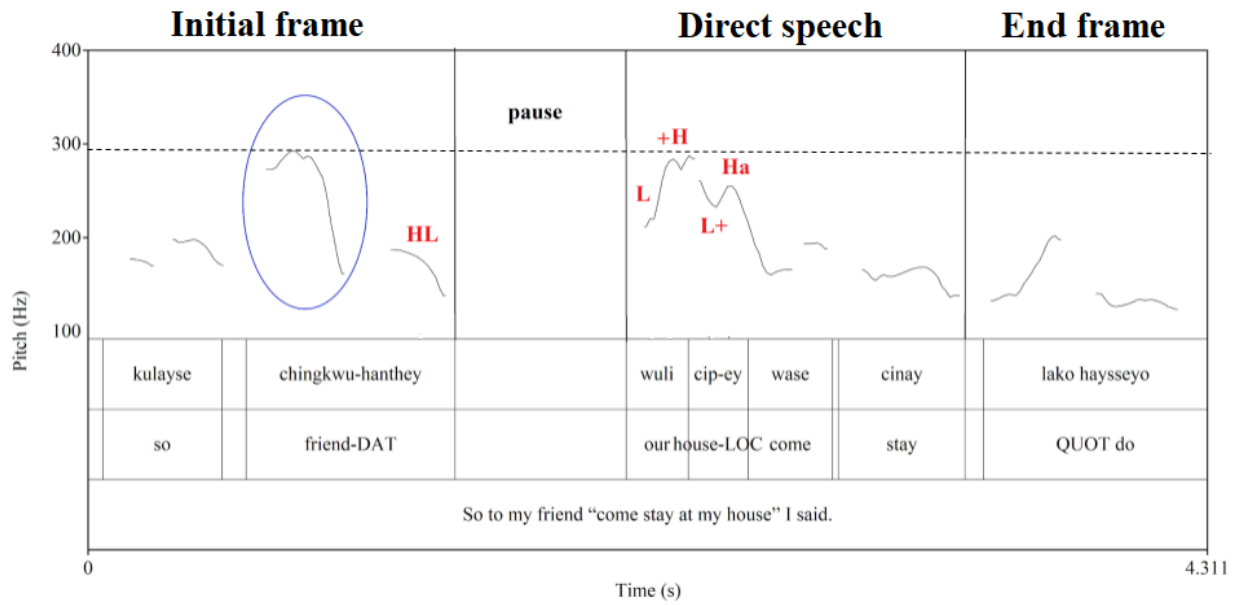


Figure 7. CJR1 direct speech

Due to the subject matter of the tasks, the word *chingkwi* ‘friend’ was often present in the initial frame. The first segment of this word is aspirated (represented by <ch>). There are cases where the pitch of the first high tone in the first intonational contour of the direct speech is comparable to the pitch of the aspirated consonant in the initial frame. The circled pitch in Figure 7 is the highest of the utterance. The tones of the first AP of the direct speech section are also illustrated. The dotted line provides a visual reference to compare the f0 heights of the H tone of the aspirated consonant and the H tone in the direct speech AP. If pitch reset did not occur, the H tone of the direct speech AP would be expected to be lower– it does not have a fortis, aspirated, /s/, or /h/ segment. Thus, we can conclude that pitch reset can and does occur with direct speech in Korean, although unreliably so.

The number of times that different prosodic cues occurred at a strong or weak prosodic boundary was also analyzed. Chart 1 provides a comparison of the number of times a prosodic

cue at the end boundary of the initial frame was present as a strong boundary, versus a weak boundary, for direct speech examples.

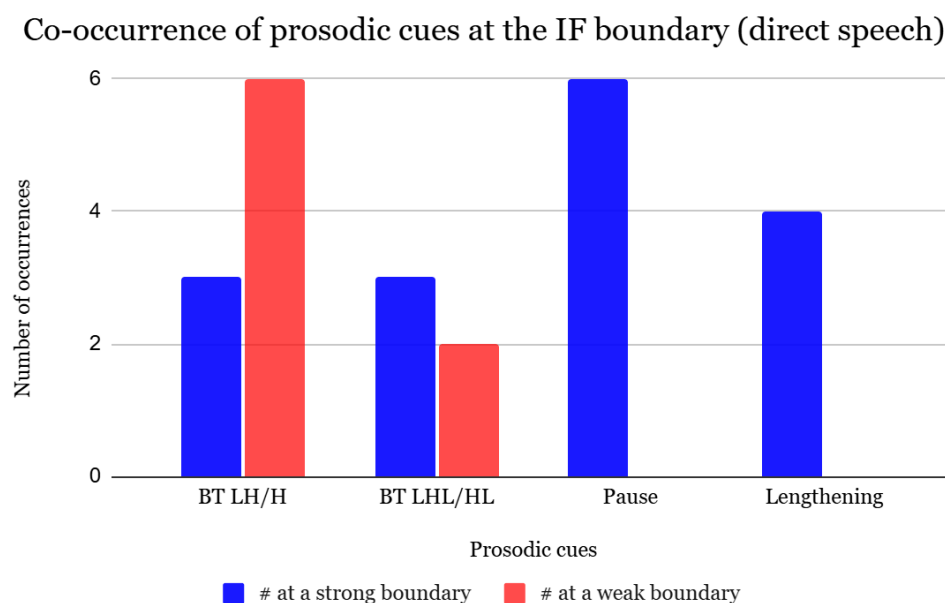


Chart 1. Co-occurrence of prosodic cues at the initial frame (IF) boundary of direct speech examples

As demonstrated by Chart 1, pauses and lengthening never occurred at a weak boundary at the initial frame in direct speech examples. However, the prosodic boundary at the end of the direct speech segment shows slightly different results. Chart 2 illustrates the occurrence of prosodic cues at either strong or weak boundaries at the end of the direct speech segment.

Co-occurrence of prosodic cues at the direct speech boundary

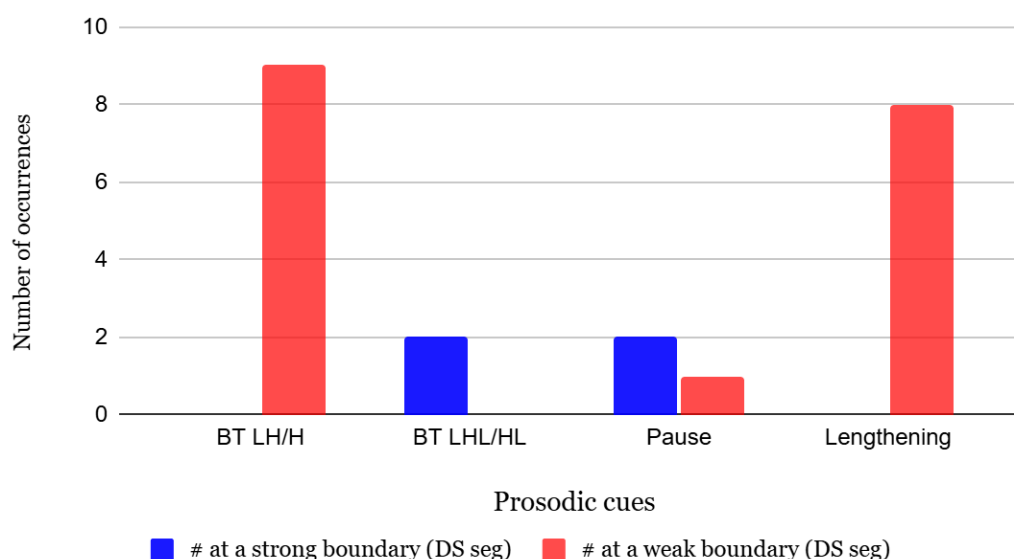


Chart 2. Co-occurrence of prosodic cues at the end boundary of direct speech segments

Unlike the initial frame boundary, lengthening at the direct speech segment boundary almost always occurred at a weak boundary. At the initial frame boundary, lengthening most often occurred when the boundary tone was LHL; lengthening was likely a product of the tonal contour. In contrast, lengthening at the end of the direct speech segment was often accompanied by a high tone. Furthermore, a rising or high boundary tone is most often found at the end of the direct speech segment.

To conclude, two boundaries are marked prosodically in sentences containing direct speech: at the onset, and at the end, of the direct speech section. Lengthening is present at both boundaries, with pauses more often used at the onset of direct speech, as a strong boundary. Weak prosodic boundaries often occurred at the end boundary of the direct speech. Pitch reset was also observed in the direct speech segment, though it remains unreliable due to the influence of segment type on F0 values. The utterances ended with the *verbum dicendi* which was

sometimes produced with a creaky voice. The *verbum dicendi* ended with a low boundary tone, consistent with the low IP boundary associated with the end of declarative sentences.

3.2.4 Indirect speech

In comparison with the direct speech examples discussed in Section 3.2.3, there was a relative lack of prosodic cues in sentences that included indirect speech. The boundary between the end of the initial frame and the onset of the indirect speech was the only place that prosodic cues occurred, though their use was inconsistent and often mainly included the use of boundary tones, rather than other cues such as pauses or lengthening.

Almost all of the initial frames had a boundary tone ending high (LH or H), as illustrated in Table 4, provided in the following Section 3.2.5. Three different speakers all produced one example of a falling tone at the edge of the initial frame, but unlike direct speech, it was unclear if the pauses that occasionally accompanied the falling tone in indirect speech were intentional prosodic cues. In one of the cases of the falling tone, the break that followed was a breath, not a pause. The following figures show an example of a sentence with the falling tone at the end of the initial frame (Figure 8), and a sentence with the H tone at the end of the initial frame (Figure 9), produced by the same speaker. Figure 8 below shows participant CSJ's use of a falling tone at the end of the initial frame.

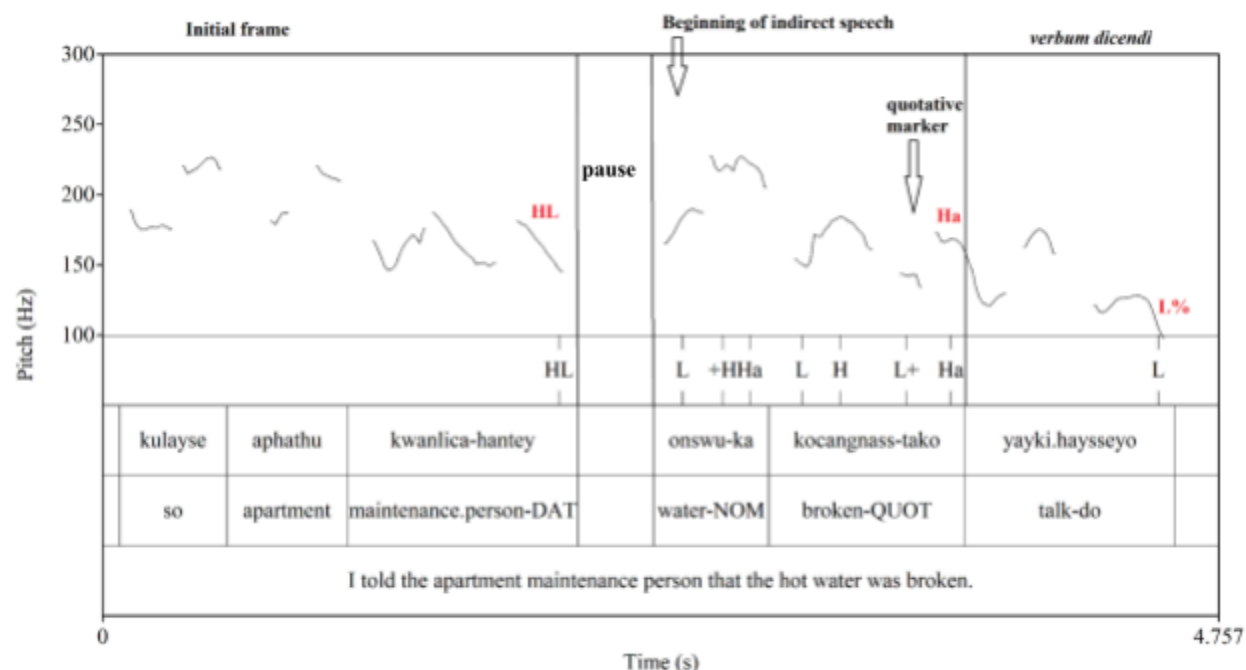


Figure 8. CSJ2 indirect speech with a falling tone at the initial frame boundary

When a disjuncture occurred between the initial frame and the onset of indirect speech, it was often very short, and for some participants such as GBJ, there was an audible intake of breath. However, when comparing Figure 8 with Figure 9, the lack of pause in Figure 9 indicates that there may be some similarities with the patterns in direct speech. Falling tones at the boundary of initial frames, as in Figure 8, may be accompanied by short pauses. Figure 9 shows participant CSJ's use of a high tone at the end of the initial frame.

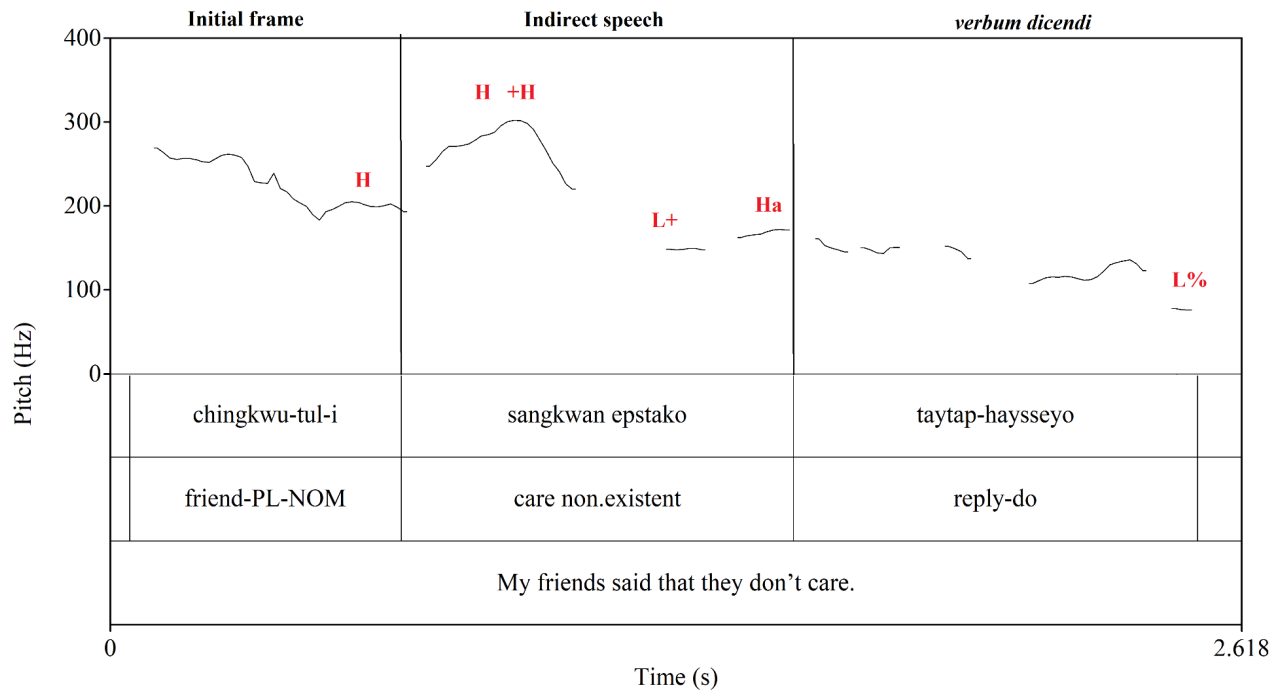


Figure 9. CSJ3 indirect speech with a high tone at the initial frame boundary

As mentioned above, it appears that falling tones at the initial frame boundary more often predict pauses than a high tone at the initial frame boundary, as shown in Figure 9.

Most of the sentences containing indirect speech did not appear to undergo pitch reset in the indirect speech segment, exhibiting a downtrend across the entire sentence. As mentioned earlier, pitch reset is an unreliable cue due to the variation caused by specific segments on pitch—indirect speech that appeared to interrupt the downtrend of the sentence always contained a segment that caused a raise in pitch. This is made clear in the aforementioned Figure 9 as well as Figure 10 below. Figure 10 shows an indirect speech example produced by speaker GBJ that includes a segment (/s/) that raises the initial tone of an AP contour (THLH) to H, which makes it appear as though the indirect speech segment undergoes pitch reset.

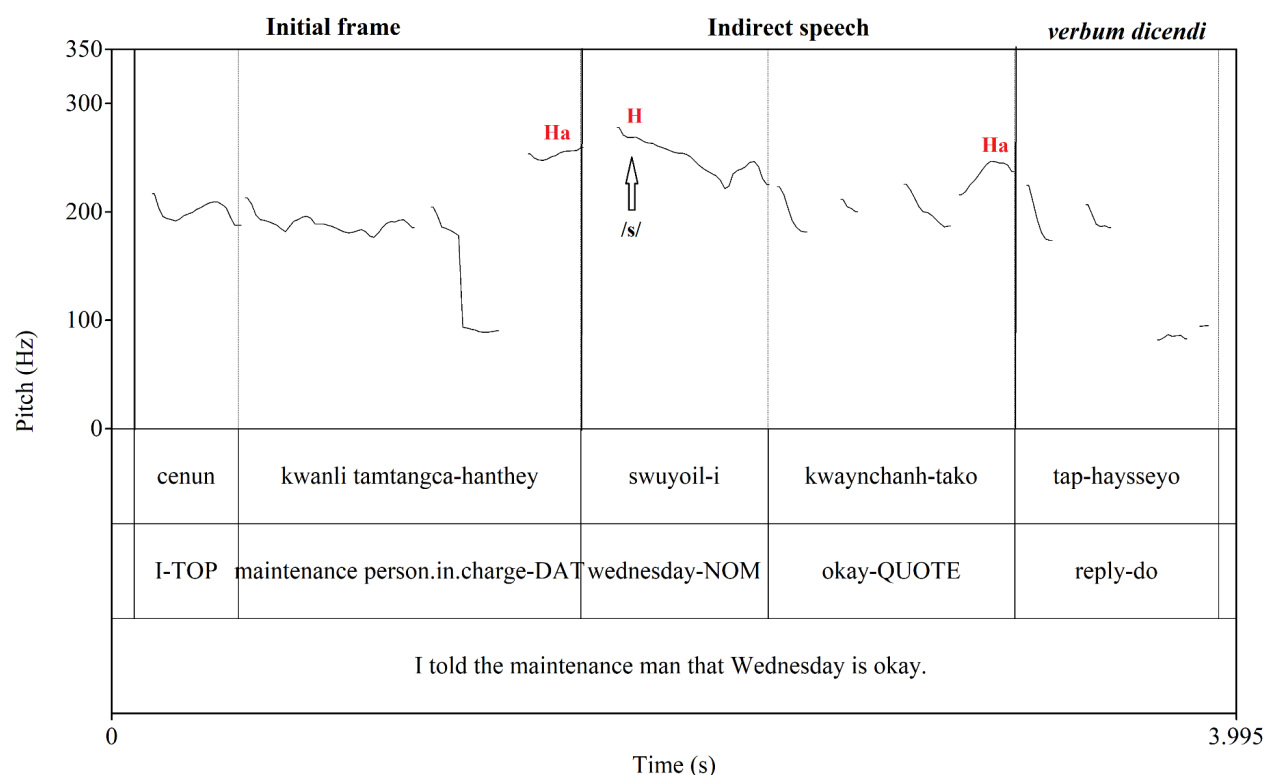


Figure 10. GB5 indirect speech

Figure 9 has the segment /s/ in the first syllable of the first word of the indirect speech segment, and an aspirated consonant (<ch>) in the initial frame, so the sentence itself starts with an H tone. Compared to the pitch in the initial frame, the first pitch of the indirect speech section is high. The indirect speech segment *sangkwan epstako* forms a five syllable AP, and with the /s/ as the initial segment the contour is HHLH. Figure 10 is another example with an initial /s/ segment, though with no aspirated or fortis segment raising the pitch of the initial frame.

The occurrence of prosodic cues at a strong or weak boundary in indirect speech was also analyzed. Chart 3 provides a comparison of the number of times prosodic cues at the initial frame occur with a strong or weak boundary.

Co-occurrence of prosodic cues at the IF boundary (indirect speech)

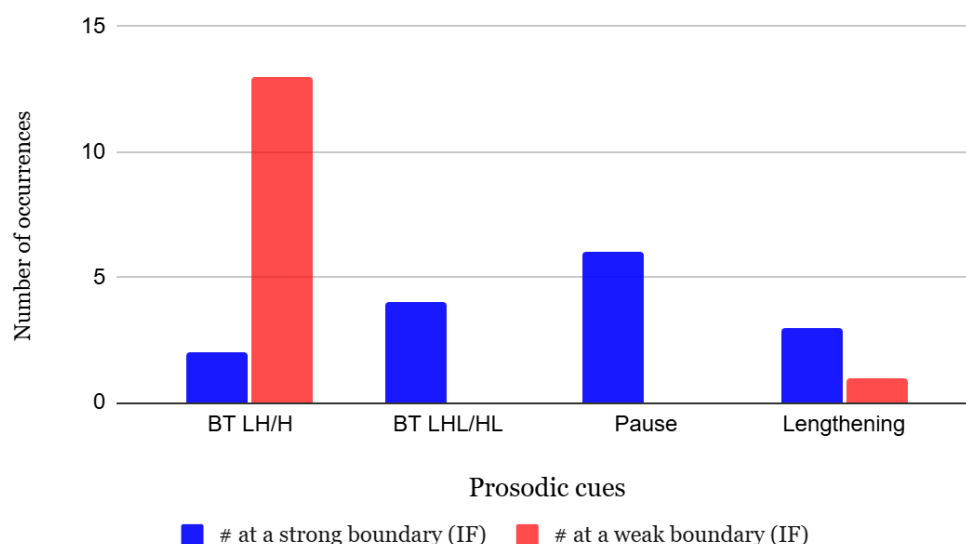


Chart 3. Co-occurrence of prosodic cues at the initial frame boundary in indirect speech examples

The rising-falling and falling boundary tones never appear at a weak boundary in the initial frame of indirect speech. The rising or high boundary tones are the most common, and more often appear with a weak boundary. A chart is not provided showing the co-occurrence of cues at the indirect speech boundary. This is due to the fact that only one prosodic boundary occurred at the end of the indirect speech segment, with the boundary tone LHL followed by a pause produced by participant KHJ. In all other cases, no prosodic cues were found at the end of the indirect speech segment, suggesting that no boundary occurs at that position.

The results show that only one prosodic boundary, occurring between the initial frame and indirect speech section, is consistently and clearly marked in indirect speech examples. The end boundary observed in direct speech examples was not observed in indirect speech examples. Further, the observed prosodic boundaries in IDS examples, those at the initial frame, were often

weak. The *verbum dicendi* is the lowest pitch of the sentence, consistent with the end of a declarative.

3.2.5 Occurrence of cues in Direct and Indirect speech

In terms of number of prosodic cues present in both types of reported speech, Table 4 provides the type of prosodic cue observed, which boundary it occurred at, and how many instances were present in the sentences analyzed. Tables for each individual participant are provided in Appendix D.

Table 4. Number of prosodic cues occurring in direct speech and indirect speech examples

Prosodic Cues		Direct speech		Indirect speech	
		Initial frame	DS seg	Initial frame	IDS seg
Boundary tones	<i>rising-falling/falling:</i>	5	2	2	1
	<i>rising/high:</i>	11	4	19	—
Pauses		6	2	4	0
Lengthening		1	8	1	0
Pitch reset		3		0	

The columns of “DS/IDS seg” refers to the final syllable of the direct speech or indirect speech segment for lengthening and boundary tones, and immediately after the final syllable for pauses.

The boundary tones here were concerned with four patterns: LHL, HL, LH, and H. A low boundary tone (L) was not marked in Table 4, as it occurred only once (at the end of the direct speech segment). Furthermore, the contour of the IDS segment often included the *verbum dicendi*; the lack of cues at the end boundary of the indirect speech segment was discussed in

Section 3.2.4. As such, the boundary tones in the “IDS seg” column (referring to the edge of the indirect speech segment) is left blank, rather than given a value of zero. As pitch reset concerned the entire utterance, it was not divided into the two boundary points. Additionally, the total number of pitch reset only included utterances in which it clearly occurred.

3.3 Summary and Discussion

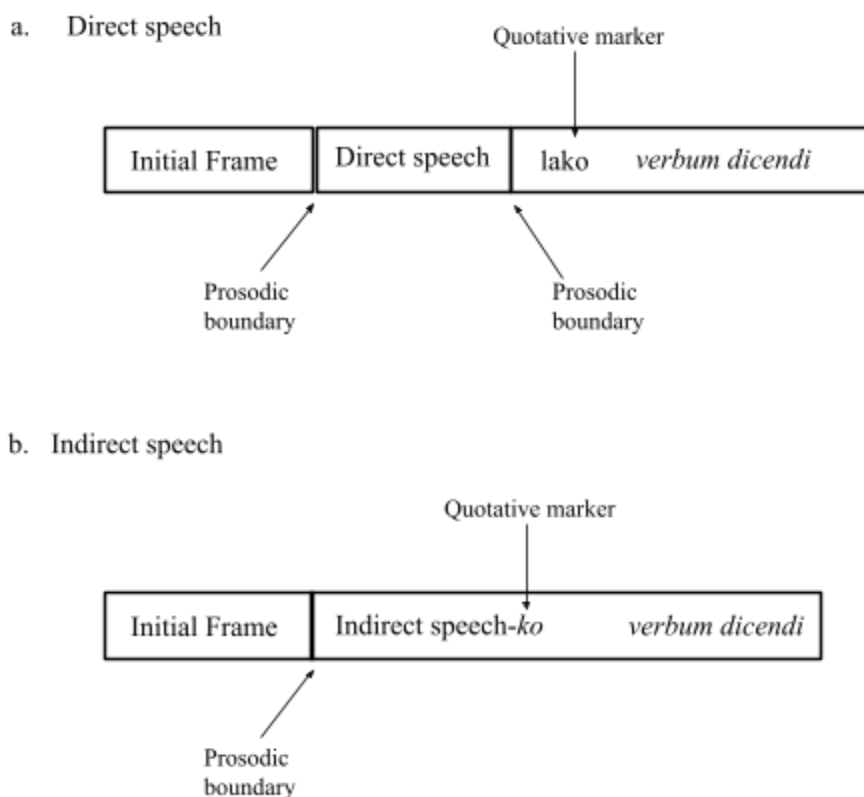
Based on the results, it appears that prosodic cues are used in the delineation of some reported speech boundaries. Participants used prosodic cues, to varying degrees, at the onset of direct speech, and often at the end as well. The initial frame appeared to have two types of boundary tones: rising-falling/falling, or rising/high. A high or rising tone was the most common boundary tone observed at the edge of the initial frame in both direct and indirect speech examples. One participant, CJR, tended to use a falling boundary tone at the edge of the initial frame, but used a high boundary tone when the initial frame consisted of one word. Whether or not there is a correlation between the number of words (or weight) in the initial frame and the boundary tone would be of interest for future study. Though not the most reliable cue, pitch reset did occur in some examples of direct speech, with three clear instances. The DS-final segment was often lengthened as an indication of the weak end boundary. In contrast, indirect speech seemed to employ very minimal, or no cues at the initial boundary, and never at the end of the indirect speech segment. For instance, participant CJR often lengthened the final segment of a direct speech section, but never did so for indirect speech. The lack of a prosodic boundary at the end of the IDS segment seems to be the most consistent prosodic difference between the two types of reported speech.

Syntactically, the embedded direct speech segment is a clause, and so I hypothesized that the edge of a direct speech segment may align with the edge of an IP. If so, would the IP-final prosody be present in the embedded speech segment? It appears that this is not quite the case. Though some speakers seem to produce a low or falling pitch on the final syllable of the direct speech segment, it cannot be definitively analyzed as the low boundary tone that is associated with most IPs. In all of the sentences, the quotative marker begins with a L tone (consistent with the tonal contour of an AP expected with this kind of segment). Thus, the falling or low tone seen at the end of the direct speech could be a natural fall to meet the following low target. Furthermore, I could expect an IP-final L% to be at the lowest range of an utterance, but this is not the case, as the lowest pitch always occurred on the utterance final syllable. Other cues associated with IP boundaries occurred at the ends of the initial frame and direct speech segment, such as pauses and lengthening. This is an interesting discovery, as only ips or IPs are associated with lengthening, based on previous research on Korean prosody, so the units with these cues cannot be analyzed as APs. Does the direct speech trigger an Intermediate Phrase, or can it be explained another way? In Chapter 4 I will address this question, providing an explanation using Match Theory.

The difference between the use of cues with direct speech in contrast with indirect speech may come from the fact that in the two types of reported speech, the quotative marker behaves differently. The earlier analysis of the quotative marker in Section 2.3 assumed that a complete breakdown of the morphosyntax was unnecessary. However, to explain why direct and indirect speech are treated differently, the quotative morpheme must be extracted from what I have been calling the quotative marker. Recalling Section 2.3, *-ko* is a quotative clitic developed from Middle Korean (Kwon, 1998) that is part of a group of affixes called Sentence Final Particles

(Yoon, 2018). SFPs can attach to the group of affixes called Final Endings (Yoon, 2018), of which the *-(nu)nta/ta* morpheme, seen on the verb of indirect speech examples, is a part of. However, for direct speech, the verbal complex is fully inflected and therefore the quotative clitic *-ko* cannot attach. As *-ko* cannot stand on its own, I assume that it must be strengthened by including the morpheme *la*. Thus, the verbal complex and the quotative are two separate entities in direct speech. In contrast, the verbal complex includes the cliticized quotative in indirect speech examples. The schematic in (9) below demonstrates the difference in organization of the quotative marker in both reported speech types.

(9)



Where the direct speech is framed by two prosodic boundaries, as reflected in (9a), the results show that the second prosodic boundary is not realized in indirect speech, as reflected in (9b).

The Match Theory analysis in Chapter 4 addresses these differences.

Support for the claim that prosody is used as an indication of reported speech is further provided in the following example from the data, which shows how prosody can disambiguate ambiguous pairs of direct and indirect speech. As mentioned in Section 2.3 on reported speech in Korean, the quotative marker is part of a paradigm. The imperative quotative marker has the same form as the quotative marker used for direct speech: *-lako*. Consider the sentence from participant CSJ:

- (10) Kulayse chinkwu-hanthey wuli cip-eyse ca lako haysseyo
 So friend-DAT our house-LOC sleep QUOT do

- a) kulayse chinkwu-hanthey **wuli cip-eyse ca** lako haysseyo
 so friend-DAT our house-LOC sleep QUOT do
 ‘So, to my friends “stay at my house” (I) said.’

- a’) kulayse chinkwuhanthey wuli cip-eyse ca-lako haysseyo
 so friend-DAT our house-LOC sleep-IMP.QUOT do
 ‘So, I told my friends to stay at my house.’

- b) kulayse chinkwu-hanthey wuli cip-eyse ca-nyako haysseyo
 so friend-DAT our house-LOC sleep-INT.QUOT do
 ‘So, to my friends “are you staying at my house?” I said.’

The sentence (10a) is present in the data as a direct speech example, and example (a’) is the predicted form of an imperative indirect speech segment. The sentences in (a) and (a’) have

the same sequence of morphemes following the verb *ca* ‘sleep.’ However, the interpretation is different. The indirect speech in sentence (a’) is understood as an imperative due to the *-la* morpheme, it can only be understood as a command. Compare this with (b), where the quotative marker is paired with the interrogative *-nya*, so the indirect speech segment is understood as a question. The sentence in (a), in contrast, is a sentence with direct speech, and unlike the *-la* of the indirect speech’s imperative, the *-la* of the direct speech quotative is not an imperative. When spoken, there should be ambiguity with the sentences (a) and (a’). However, prosody may disambiguate them. The sentence in (a) was produced by participant CSJ, with audible lengthening of the final segment of the direct speech; in this case, the verb. As the results from the experiments show that the end frame is never marked in indirect speech, I would predict that no lengthening would occur on the verb in (a’). Thus, the lengthening cues a boundary that is only flagged in direct speech, allowing a listener to interpret (a) as a direct quotation, rather than the imperative report seen in (a’). This sentence provides a unique example, as the verb did not have the politeness suffix *-yo*, and ended in a vowel, which resulted in a minimal pair with the imperative. This phenomenon of disambiguating direct speech and indirect speech through prosody could be interesting to explore further.

In sum, Korean uses prosody to mark the boundaries of reported speech, where both the onset of DS and IDS, and the offset of DS are flagged through a variety of prosodic cues. The quotative marker may influence the need for prosodic cues at the end boundary, especially when it cliticizes to the IDS verb.

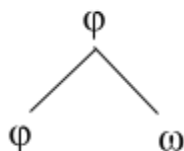
Chapter 4: Match Theory Analysis

Sections 3.2 and 3.3 assumed the theories developed based on strict layering and the resulting prosodic structure in Korean. However, reported speech provides a unique opportunity to observe the relationship between syntactic embedding and recursivity in prosodic domains. The following is my analysis of Korean prosody using Match Theory (Selkirk, 2011), a framework in which recursive prosodic structures are permitted, in an attempt to explain and predict *why* and *where* boundaries occur in the reported speech data.

4.1 Direct speech

In work by Jun (2005a) and others, the intonational phonology of Korean is assumed to be organized under the assumptions of the Strict Layer Hypothesis. The Strict Layer Hypothesis is the idea that the prosodic structure is arranged such that constituents dominate constituents of a lower category in the prosodic hierarchy (Nespor & Vogel, 1986; Pierrehumbert & Beckman, 1988; Selkirk, 1986, 2011). Thus, Intonational Phrases will dominate Phonological Phrases, which in turn dominate Prosodic Words. An example of a strict layering violation would be a Phonological Phrase dominating another Phonological Phrase:

(11)



Strict layering bans recursivity, so a prosodic unit cannot dominate the same category type (Selkirk, 2011). This violation is exemplified in (11). Korean has been analyzed as having a

three-layer (higher-level) prosodic structure: the Prosodic Word, Accentual Phrase, and Intonational Phrase (Jun, 2005a), but to account for certain phenomena, the Intermediate Phrase was later proposed (Jun, 2005b).

However, I suggest that Match Theory, proposed by Selkirk (2011) which allows recursivity in the prosodic structure, can explain the phenomena found in my reported speech data. Match Theory maps syntactic trees onto prosodic trees, assuming a correlation between CPs, XPs, and X^0 s of syntax with intonational phrases (ι), phonological phrases (ϕ), and prosodic words (ω) of the prosodic hierarchy (Ishihara & Kalivoda, 2022). The correspondence between syntactic structure and prosodic structure comes from MATCH constraints, and misalignments in the correspondence between syntax and prosody are due to other constraints outranking the MATCH constraints. Essentially, the prosodic structure will correlate to the syntax, unless there is a constraint that outranks the MATCH constraint. Match Theory assumes the Optimality Theoretic framework of a constraint-based ranking system (Ishihara & Kalivoda, 2022).

On the basis of the Korean intonational phonology framework, I propose that the IP correlates to an intonational phrase (ι), associated with a clause (CP) in the syntax, and that both APs and ips correlate to phonological phrases (ϕ), associated with a phrase (XP) in the syntax. Since this analysis does not deal with the level of the Prosodic Word, only two MATCH constraints are relevant, Match Clause and Match Phrase. I adopt Elfner's (2015) definition of Match Phrase, also applying it to Match Clause.

(12) Match Constraints

a) Match Clause (modified from Elfner, 2015)

MATCH(CP, ι): “For every syntactic clause (CP) in the syntactic representation that exhaustively dominates a set of one or more terminal nodes α , there must be a prosodic domain (ι) in the phonological representation that exhaustively dominates all and only the phonological exponents of the terminal nodes in α .”

b) Match Phrase (Elfner, 2015)

MATCH(XP, φ): “For every syntactic phrase (XP) in the syntactic representation that exhaustively dominates a set of one or more terminal nodes α , there must be a prosodic domain (φ) in the phonological representation that exhaustively dominates all and only the phonological exponents of the terminal nodes in α .”

Thus, Match Phrase and Match Clause are only concerned with the overt phonological elements in the syntax (ignoring non-overt elements such as traces) (Elfner, 2015).

I propose that the characteristics of IPs, ips, and APs discussed in this paper (based on Jun, 2005a, 2005b) will correspond to ι -phrases and φ -phrases. Recall that the Intonational Phrase is associated with nine possible boundary tones at the right edge, lengthening of the final syllable, and followed by a pause. The Intermediate Phrase is associated with a left- or right-edge boundary tone, optional lengthening, and an optional pause. The Accentual Phrase is the smallest unit with an intonational contour, and is most often associated with a H boundary tone, no lengthening, and no pauses. So, φ -phrases should have the characteristics of APs and ips, and ι -phrases should have the characteristics of IPs. First, consider this example of direct speech from participant CJR:

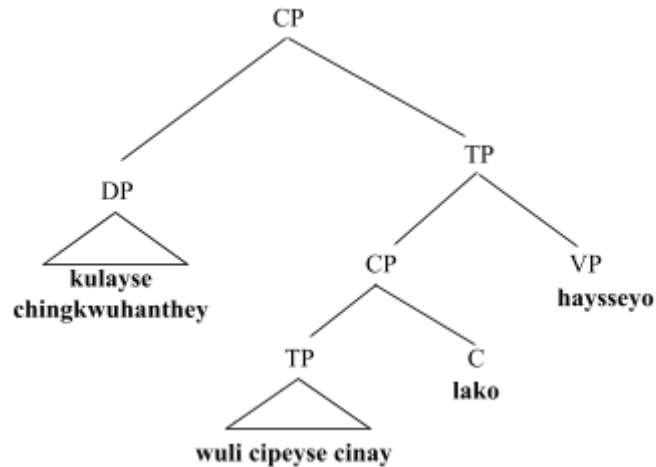
(13) CJR direct speech ⁴

HL
H

kulayse chinkwu-hanthey wuli cipecy wa-se cinay lako haysseyo
 so friend-DAT our house-LOC come-CONJ stay QUOT.COMP do-PST-POL
 “So to my friend ‘Come stay at my house’ (I) said.”

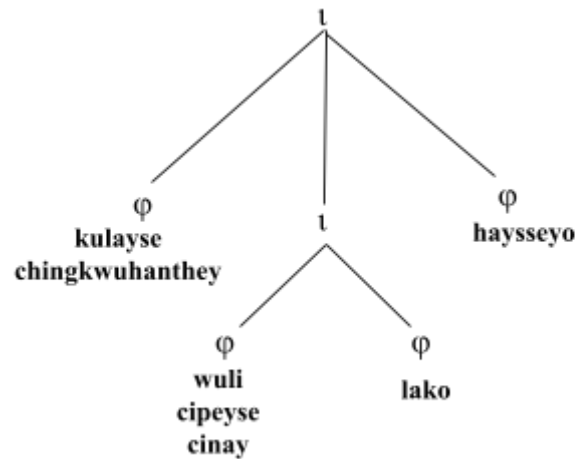
The syntactic constituents can be broken down like so:

a.



Therefore, assuming the correlation between the syntactic constituents and prosodic units, we predict that the prosodic structure would be the following:

b.



There are a few issues with the structure provided in (13b). The HL contour at the right edge of the first phonological phrase of the utterance would appear to be consistent with an IP contour, not that of a ϕ (or AP) contour. There is also a pause after the first unit, which should not occur with a ϕ -phrase. Further, the right edge of the embedded l -phrase should be followed by a pause, as that is an element of an intonational phrase in Korean, but it is not. It is, however, lengthened. Based on (13b), we would predict the strongest boundaries would occur at the right edge of the highest unit, the intonational phrase, but this is not the case for the embedded l . There are no cues to indicate a boundary between *lako* of the embedded IP and the following phonological phrase. Instead, a boundary occurs within the embedded IP, between the end of the direct speech segment and the quotative marker.

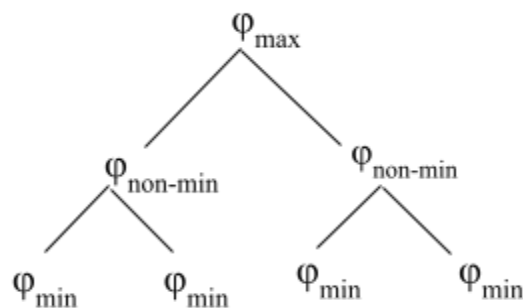
A further potential issue with the structure shown in (13b) is the violation of a Match Theory constraint called STRONGSTART.

(14) STRONGSTART: “A prosodic constituent optimally begins with a leftmost daughter constituent which is not lower in the prosodic hierarchy than the constituent that immediately follows.” (Selkirk, 2011)

So, this constraint prevents a phrase from starting with a prosodic unit that is lower on the hierarchy than the following unit. The structure seen in (13b) shows a φ preceding an ι , which is a higher unit, thus violating STRONGSTART.

Another issue with (13b) is the amalgamation of APs and ips as one unit, φ . Not all φ -phrases behave the same in Korean, as we see with the need for an AP and ip in the strict layer structure. Unlike strict layering, there is no need to add an additional phrase, simply a distinction between minimal and non-minimal projections of φ . In their Match Theory reanalysis of the ip in Korean, Baek & Yun (2018) find that the edges of $\varphi_{\text{non-min}}$ were associated with strong prosodic boundaries, and the co-occurrence of these edges affected the boundary between relative clauses and noun phrases which influenced the reading of the phrase. The distinction of φ -phrases into φ_{max} , $\varphi_{\text{non-min}}$, and φ_{min} has been used to explain phenomena in languages such as Irish (Elfner, 2015) and Japanese (Ito & Mester, 2012), for example. Illustrated in (14) is a possible structure of embedded φ .

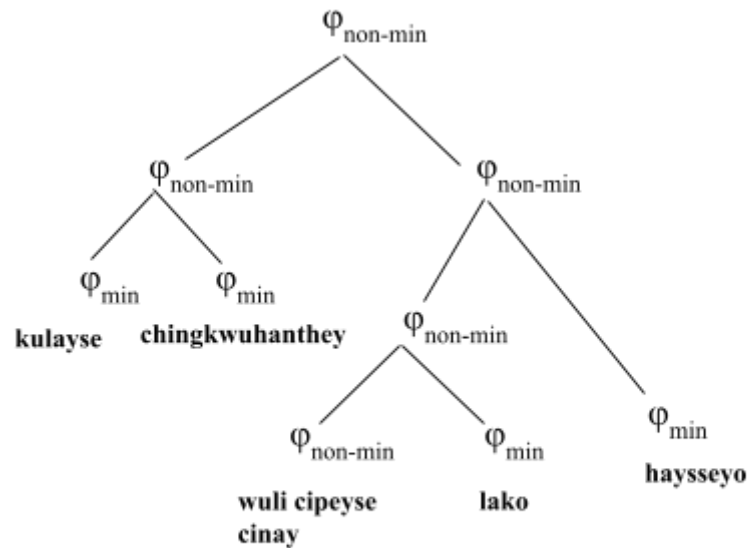
(15)



The minimal and non-minimal projects allow for more precision; we can now specify that the smallest unit with an intonational contour is φ_{min} .

There are a few possible ways to resolve the issues that arise from the structure seen in (13b). First, the recursivity of ϕ -phrases and how they differ has been explained above, providing a more in depth description of the structures being observed. To solve the issue with the STRONGSTART constraint, we could assume that the first unit is not a ϕ , but actually a ι -phrase, which also explains the prosodic cues observed there. Thus, the sentence could be divided into two or more ι -phrases: one formed from the initial frame, and another from the direct speech segment, and the third from the *verbum dicendi*. However, this solution does not explain the lack of cues at the right edge of the direct speech ι -phrase. Therefore, I propose an alternative analysis that includes $\phi_{\text{non-min}}$ and ϕ_{min} , provided in (16).

(16) Alternative structure of CJR direct speech:



In the above example, we have both ϕ_{min} and $\phi_{\text{non-min}}$, and the previously embedded ι -phrase is parsed as $\phi_{\text{non-min}}$. Selkirk (2011) notes that in Xitsonga, a preposed syntactic phrase is parsed as an ι -domain, not the predicted ϕ -phrase, which could be explained by the STRONGSTART

constraint outranking the MATCHCLAUSE constraint, and thus promoting the preposed phrase. Perhaps a similar constraint ranking occurs in Korean, where the initial, lower constituent is not promoted as in Xitsonga, but the following higher constituent is demoted. Thus, the embedded ι -phrase is parsed as a ϕ -phrase as to not violate STRONGSTART. Alternatively, the constraint EQUALSISTERS proposed by Myrberg (2013) also accounts for the embedded clause being parsed as a ϕ -phrase. EQUALSISTERS states that sister nodes in the prosodic structure are of the same prosodic category (Myrberg, 2013). A crucial difference between STRONGSTART and EQUALSISTERS is that STRONGSTART cares about the leftmost constituent. This makes no difference for the current data, therefore I will assume that STRONGSTART is the relevant constraint for this phenomena.

Further support for the analysis in (16) comes from Baek & Yun (2018). They found that co-occurring $\phi_{\text{non-min}}$ edges predicted a stronger boundary than that of a single $\phi_{\text{non-min}}$ edge. Recall the definitions of strong and weak boundaries in this paper, stated in Section 3.2.2. Strong boundaries are marked prosodically with a boundary tone plus another prosodic cue, or an H tone minimally followed by a pause. Weak boundaries have one or no prosodic cues. In Figure 11 below, we can see where the different prosodic cues occur for this utterance.

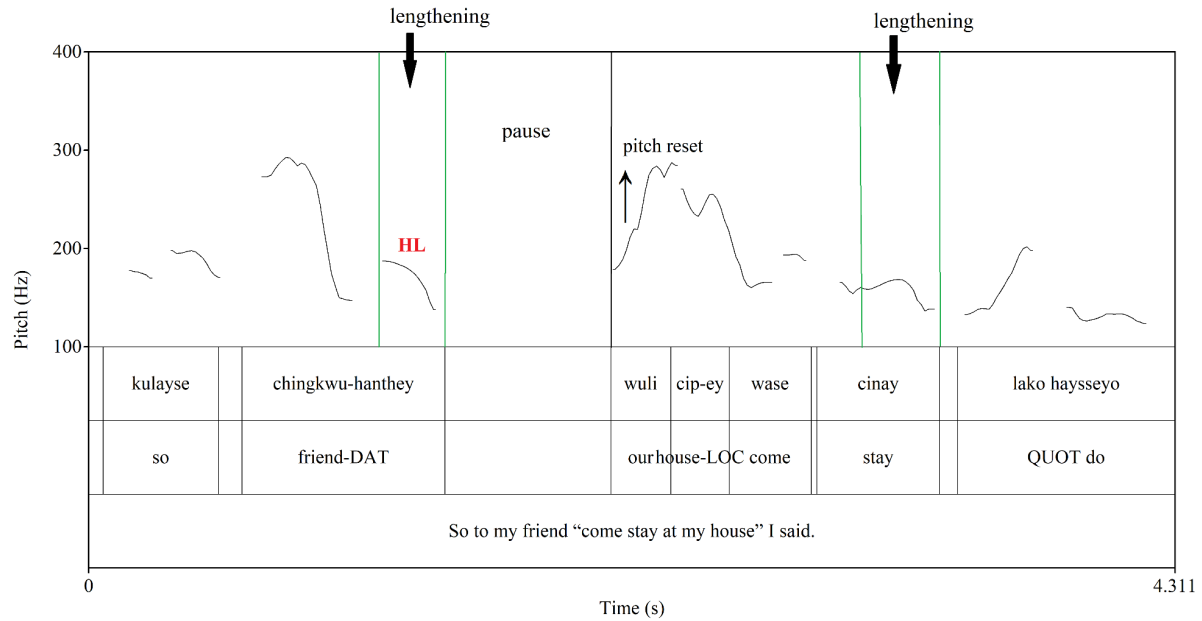
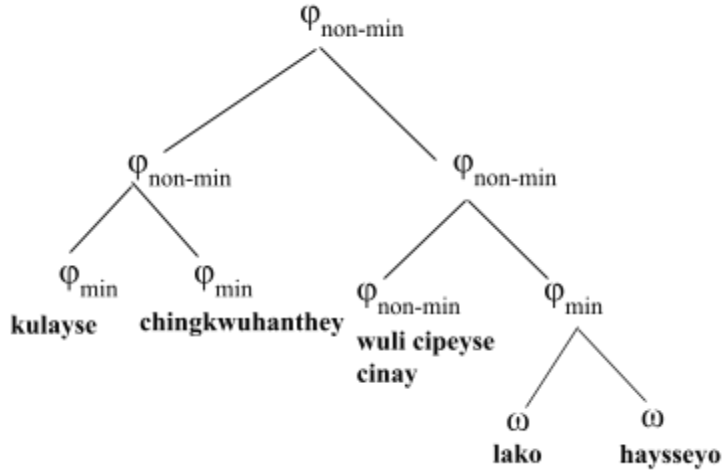


Figure 11. CJR boundary cues

Comparing the structure in (16) to the cues in Figure 11, we can see that a falling boundary tone, pause, and pitch reset are all present where two $\phi_{\text{non-min}}$ edges co-occur, at the end of the initial frame and at the onset of direct speech. The structure in (16), however, raises an issue with the end frame. There is an intervening $\phi_{\text{non-min}}$ (that containing the $\phi_{\text{non-min}}$ of the direct speech and the ϕ_{min} of the quotative marker), thus predicting a weak boundary. However, no prosodic cue ever occurs between the quotative marker and the *verbum dicendi*, which suggests that no boundary exists in this position. Thus, a revised structure is provided in (17) below:

(17) Revised structure of CJR direct speech



Further, the data shows that the *verbum dicendi* is pronounced together with the quotative marker. Thus, the *verbum dicendi* must be a prosodic word (ω), reflected in the revised structure (17) above. Recalling the constraint EQUALSISTERS, I propose that the *verbum dicendi* is parsed together with the quotative marker as a $\phi_{\min}$, due to this constraint. Thus, no boundary is predicted between the quotative marker and *verbum dicendi*.

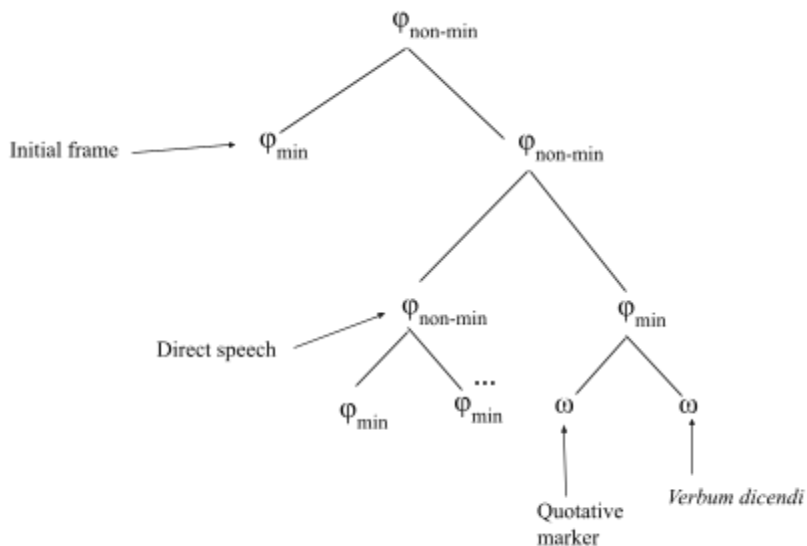
So, this analysis has established that the type of ϕ -phrase will have an effect on the strength of boundaries. Strong boundaries are predicted to occur with adjacent $\phi_{\text{non-min}}$ phrases, and weaker boundaries are predicted to occur when there is a single $\phi_{\text{non-min}}$ edge. The adjacency of the $\phi_{\text{non-min}}$ edges of the initial frame and direct speech trigger a strong boundary. The edge of the direct speech $\phi_{\text{non-min}}$ is adjacent to the $\phi_{\min}$ edge of the quotative marker, thus causing a weak boundary. However, this only accounts for initial frames that contain multiple $\phi_{\min}$ within a $\phi_{\text{non-min}}$. Does this analysis hold when the initial frame is $\phi_{\min}$? Based on the findings of Baek & Yun (2018), the right edge is more active in Korean, and in the case of a $\phi_{\min}$ initial frame, the right edge of the $\phi_{\min}$ is adjacent to the left edge of a $\phi_{\text{non-min}}$. This suggest that only the $\phi_{\min}$

boundary tone (ie, H) will be present at the right edge of the frame. By chance, three participants produced the same sentence that had a single-word initial frame, shown in (18).

- (18) chingkwu-tul-i kwaynchanhta lako taytap-haysseyo
 friend-PL-NOM okay QUOT reply-do
 “My friends replied ‘It’s okay.’”

The prosody for all three participants was the same for the initial frame: a H tone at the end of the first word, with no other cues. All three participants treated the end boundary differently, by lengthening direct speech final segment, a falling boundary tone, or no cues at all. This is consistent with the predictions of a $\phi_{\min}$ edge at the end of the initial frame. This is illustrated in the schematic below.

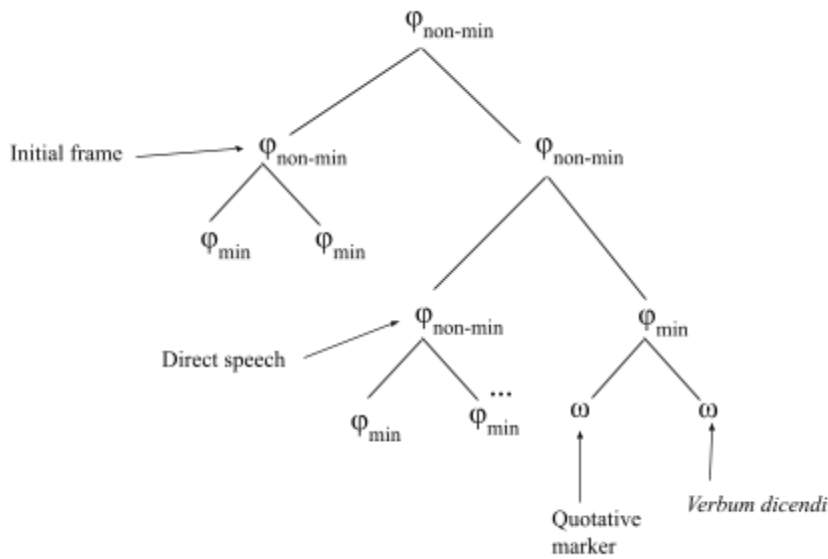
(19) Direct speech structure with $\phi_{\min}$ initial frame



The structure in (19) demonstrates that no boundary is predicted to occur at the edge of an initial frame with a $\phi_{\min}$ edge. This is consistent with the utterances produced by the three participants

who used the phrase described in (18), and consistent with the analysis of ϕ -phrase edges in Korean that I have proposed so far. The schematic in (20) shows the structure of when the initial frame is $\phi_{\text{non-min}}$.

(20) Direct speech structure with $\phi_{\text{non-min}}$ initial frame



The structure of (20) reflects the utterance in (17), where the initial frame is a $\phi_{\text{non-min}}$, with a strong boundary between the initial frame and direct speech due to co-occurring $\phi_{\text{non-min}}$ edges. The weak boundary between the direct speech and the end frame of the structures (19) and (20) are the same.

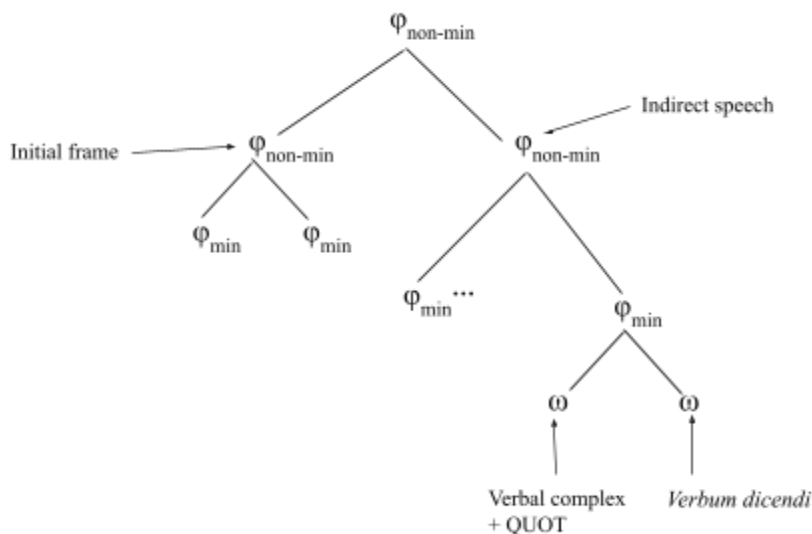
In sum, Match Theory can account for the prosody of direct speech in Korean. The constraint STRONGSTART demotes the embedded clause, that should be an ι -phrase, to a ϕ -phrase—specifically $\phi_{\text{non-min}}$. The phrase ϕ_{min} is the smallest unit with an intonational contour, equivalent to the AP in the strict layering theory of the Korean prosodic hierarchy. The type of ϕ -phrase can predict not only the strength of the prosodic cues used, but also where the

boundaries occur. The adjacency of two $\phi_{\text{non-min}}$ results in a strong boundary occurring between them, whereas a single $\phi_{\text{non-min}}$ edge predicts a weak boundary.

4.2 Indirect speech

So far, my analysis has addressed direct speech in Korean. However, this analysis can also be applied to indirect speech, even considering the difference in the use of cues. The initial frame is assumed to function as proposed for direct speech— $\phi_{\text{non-min}}$ edges predict the boundary strength. However, the lack of a prosodic boundary after the IDS segment diverges from the prosodic structure proposed for DS.

(21) Indirect speech report with $\phi_{\text{non-min}}$ initial frame



Much like the direct speech examples, the *verbum dicendi* must be absorbed into a ϕ -phrase in order to not violate EQUALSISTERS, but unlike direct speech, it is included in the ϕ -phrase of the verbal complex in indirect speech. This explains why no boundary cues were observed at the end of the indirect speech segment— simply, they are a unit so no boundary is predicted to occur.

charge', and thus has over four syllables, resulting in the interpolation of the second tone over the medial syllables, a low tone on the penultimate syllable, and a high boundary tone. I would predict a slower speech rate would result in multi-word initial frames to form a $\phi_{\text{non-min}}$, but in such a case a strong boundary would occur, consistent with the predictions made in this analysis. Figure 13 below demonstrates that some initial frames that predict a strong boundary do, in fact, occur as expected.

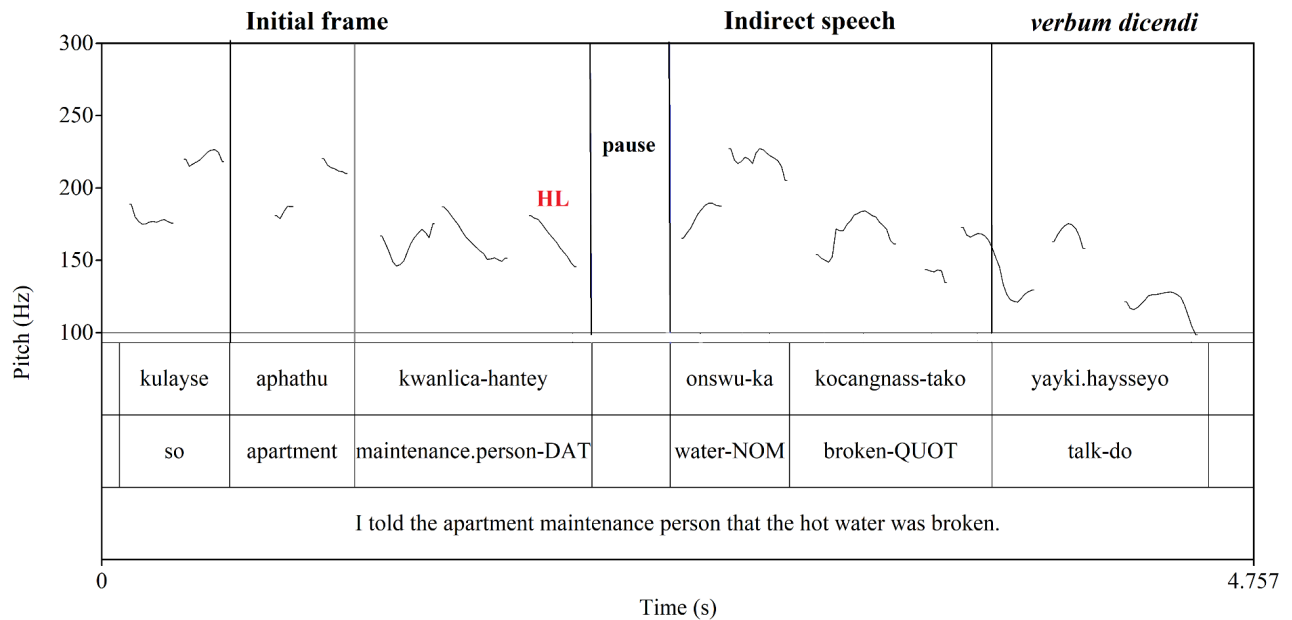


Figure 13. CSJ2 $\phi_{\text{non-min}}$ initial frame

So, although prosodic cues are used to a lesser degree in indirect speech than in direct speech, the prediction of how and where the boundaries occur can be explained through Match constraints, specifically MATCHPHRASE, for both types of reported speech. The STRONGSTART constraint applies in indirect speech as well, and therefore is consistent across both reported speech types.

4.3 Conclusion

My analysis proposed that a recursive prosodic structure predicts where and what kind of boundaries occur in an utterance that includes reported speech. Specifically, the edges of ϕ -phrases determine the strength of the boundary. Adjacent $\phi_{\text{non-min}}$ edges trigger a strong boundary (i.e. more prosodic cues), whereas a singular $\phi_{\text{non-min}}$ edge triggers a weak boundary (i.e. less prosodic cues). This accounts for the observations made regarding the differences in direct speech boundaries and indirect speech boundaries. Notably, the end frame is often indicated by prosodic cues in direct speech, but is never indicated in indirect speech. The weak end boundary of direct speech is predicted by a single $\phi_{\text{non-min}}$ edge of the direct speech segment. No such boundary occurs in indirect speech due to the assimilation of the *verbum dicendi*, a prosodic word, into the ϕ_{min} of the verbal complex (with the cliticized quotative marker) due to the constraint EQUALSISTERS.

I proposed that the constraint STRONGSTART outranking the MATCH constraints accounts for the mismatch between the embedded reported speech clause and an ι -phrase. The initial frame is parsed as a ϕ -phrase, so having an ι -phrase follow it (as predicted by MATCHCLAUSE) would violate STRONGSTART. Since the embedded reported speech clause is not parsed as an ι -phrase, but rather a ϕ -phrase, I proposed that a high ranking STRONGSTART constraint demotes the embedded ι -phrase, and it is thus parsed as a ϕ -phrase. Another mismatch between the syntactic structure and the prosodic structure is due to the *verbum dicendi* being absorbed into the ϕ_{min} of the quotative marker (direct speech) or verbal complex (indirect speech) in order to avoid a violation of EQUALSISTERS.

Chapter 5: Conclusion

The goal of this paper was to describe the ways in which prosody is used with reported speech in Korean, and the interaction between structural and intonational elements in the reported speech utterance. This paper also explored different experimental designs, with a reading task and two different production tasks, in order to test how best to elicit reported speech. Though a potential drawback of the dialogue task was reading intonation, it provided control over the types of reported speech produced, as well as their frequency. With the exception of the two participants part of the one pair that deviated from the script, the remaining four participants produced all the reported speech targets in the two dialogues of the reading task. Of the two production tasks, the question-answer task was the least successful. It was difficult to elicit specific types of reported speech, and often participants did not employ the reported speech structure to answer the questions. The fill-in-the-blanks task was the most successful of all three, with almost all possible reported speech sentences being produced. An analysis of spontaneous speech data would be an interesting avenue for further research on the prosody of reported speech in Korean.

The results showed that reported speech prosody is used in Korean, though prosodic cues are used considerably more often in direct speech compared to indirect speech. The most common prosodic cues employed by the participants were boundary tones, pauses, and lengthening, with pitch reset present in direct speech examples only. The location of the boundaries were consistent across speakers, where the onset of reported speech was most often set off, but the offset only occurring in direct speech examples. The “unquote” boundary would consistently occur between the verbal complex of the direct speech section and the quotative marker, suggesting that the quotative marker and the *verbum dicendi* form the end frame. The

findings are consistent with the literature on reported speech prosody, and variation across speakers did occur. The Strict Layer Hypothesis was assumed up to the Match Theory analysis in Chapter 4, but some inconsistencies with the boundaries, such as lengthening or pauses seemingly appearing at the edge of Accentual Phrases, lead to the proposal of a recursive analysis of reported speech prosody.

In Chapter 4, I proposed that a Match Theory analysis of Korean reported speech prosody can account for, and predict the location of, boundaries in a reported speech utterance. A reanalysis of the strict layering of Korean in order to allow for recursion in the prosodic structure lead me to propose that the smallest unit with an intonational contour is not an Accentual Phrase, but rather a $\phi_{\min}$, and that the strength of prosodic boundaries differ depending on whether the edges of ϕ -phrases are a minimal or non-minimal. I have argued that non-minimal ϕ -phrases will trigger boundaries at their right edges, and that co-occurrence of such edges will cause stronger boundaries to occur.

While the analysis proposed in this paper is supported by the findings of my speech data, more research is needed in order to determine whether it can be extended to the prosody of Korean in general. More specifically, whether this analysis works to explain phenomena such as focus, which is attributed to a $\phi_{\text{non-min}}$ in the Match Theory framework, and the Intermediate Phrase in the strict layering approach. There is limited research on recursivity in the Korean prosodic structure, which would be worth exploring further.

Notes

1. QM refers to “quotative marker”
2. The ‘A’ in *hAko* represents the arae-a (written in Hangul as *o*, or sometimes *ㅜ*), not used in Modern Korean (Lee & Ramsey, 2011).
3. The interlinear gloss in this example is not exhaustive, as it is simply meant to help illustrate the mirroring of the structures.
4. Neither the internal structure of the embedded reported speech clause, nor the level of Prosodic Word is relevant for this analysis, and therefore not reflected in this example.
5. The third person does not exist in Korean and is instead formed through compounds, but for ease of understanding *kyayhanthey* (lit. ‘that.child-DAT’) is glossed as “her.”

Appendix A

The following tables provide the Hangul, Yale, and IPA symbols. The IPA symbols are not exhaustive and may not represent the surface realizations of the sounds.

Table 1. Consonants

Hangul (consonants)	Yale	IPA
ㅁ	m	/m/
ㄴ	n	/n/
ㅇ	ng	/ŋ/
ㅂ	p	/b/, /p/
ㅃ	pp	/p̚/
ㅅ	ph	/p ^h /
ㅈ	t	/t/
ㅉ	tt	/t̚/
ㅊ	th	/t ^h /
ㅋ	k	/k/
ㆁ	kk	/k̚/
ㆁ	kh	/k ^h /
ㄷ	c	/dʒ/
ㄸ	cc	/dʒ̚/

ㅈ	ch	/tʃ/
ㅅ	s	/s/
ㅆ	ss	/s̺/
ㅎ	h	/h/
ㄹ	l	/l̥, /r/

Table 2. Vowels

Hangul (vowels)	Yale	IPA
ㅏ	a	/a/
ㅓ	ay	/ɛ/
ㅗ	ya	/ja/
ㅜ	yay	/jɛ/
ㅜㅏ	wa	/wa/
ㅜㅓ	way	/wɛ/
ㅜㅗ	e	/ə/
ㅜㅓ	ey	/e/
ㅜㅗ	ye	/jə/
ㅜㅓ	yey	/je/
ㅜㅗ	we	/wə/
ㅜㅓ	wey	/we/

ㅏ	o	/o/
ㅑ	oy	/ø/
ㅓ	yo	/jo/
ㅕ	(w)u	/u/
ㅗ	wi	/y/
ㅛ	yu	/ju/
ㅡ	u	/ʊ/
ㅟ	uy	/ui/
ㅣ	i	/i/

Appendix B

The following list provides descriptions for the boundary tones observed in the data, and is not exhaustive.

IP boundary tones (Jun, 2000)

L%: Level ending or gently falling boundary tone spread over the IP-final AP.

LH%: A rising boundary tone that rises sharply from a valley within the final syllable.

LHL%: a rising-falling boundary tone that rises within the IP-final syllable.

**HL%: A falling boundary tone that rises to a peak before the last syllable, then falls during the last syllable. The HL pattern seen in the data has a rise and sharp fall within the final syllable, thus not marked with the IP boundary tone diacritic, %.*

Appendix C

The following are the original Korean materials along with a translated English version of the three tasks.

Reading task

Dialogue 1

[친구들과 주말 여행을 가기로 해서, 숲속에 있는 펜션으로 가는 중이다. 몇몇 친구들은 이미 도착했고, 같이 차 타고 가는 친구랑 길을 잃은 것 같다...]

[*You are on your way to a cabin in the woods for a weekend trip with you friends. Some friends have already arrived, but you and your friend driving together appear to be lost...*]

가: 오래 운전을 하고 있어. 우리 이미 도착했어야 하지 않아?

We've been driving for a long time. Shouldn't we have arrived?

나: 예진이는 한 시간밖에 안 걸렸다고 했어. 우리 어디서 길을 잘못 든 거 아냐?

Yejin said it didn't take more than an hour. Did we take the wrong road somewhere?

가: 지도를 보면 이 길이 맞는데.. 펜션 주인이 찾기 어렵다고 했어

Looking at the map, this is the right way... the cabin owner said that finding the cabin is hard.

나: 그래? 그럼 지도를 믿어야지.

Really? So we should trust the map.

가: 근데 기억나? 저번에 예진이랑 저녁 먹으러 가고 있었는데, 갑자기 전화 와서 울면서 ‘나 어딘지 모르겠고 지도도 고장난 것 같아서 나 좀 찾아와줘!’ 라고 했잖아

But do you remember? Last time when we were going to have dinner with Yejin, suddenly she called and while crying said “I don't know where I am and my map is not working, come find me!”

나: 기억나! 진짜 웃겼어. ... 지금 애한테 메시지를 보내볼게. 혹시 길 알려줄 수 있을지도 몰라. 이 길이 아닌 것 같아...

I remember! So funny... I'm going to send her a message now. Maybe she can tell us the way. I don't think this is the right way...

가: 응 좋아. 펜션 주인한테도 연락해봐.

Okay. Send a message to the cabin owner, too.

나: 예진이는 그냥 지도를 따라갔다고 했어

Yejin said she just followed the map.

가: 우리도 그러고 있잖아. 예진이한테 ‘우리도 지도를 따라가고 있는데 끊임없이 계속 운전만 하고 있어’ 라고 대답해

That’s what we’re doing. Reply to Yejin “We’re following the map too but we just keep driving.”

나: 응 알겠어, 보낼게.

Alright, I’ll send it.

가: 배고파. 과자 더 많이 샀어야 했어

I’m hungry. We should’ve bought more snacks.

나: 그니까... 이렇게 오래 걸릴 줄 몰랐지

Right... we didn’t know we’d be driving so long.

가: 그러게. 애들이 우리 도착하기 전에 밥 좀 해놔으면 좋겠다

Right. I hope they make food before we arrive.

나: 그치... 아, 주인이 “보라색 꽃 옆에 있는 은행나무가 보이면 좌회전 하시고, 5분쯤 가면 펜션이 나와요”라고 답장 왔어.

Yeah... ah, the cabin owner replied “When you see the ginkgo tree with purple flowers beside it, turn left and after 5 minutes the cabin will appear.”

가: 20분 전에쯤 그런 나무 봤지?

Didn’t we see that tree about 20 minutes ago?

나: 봤어... 저기서 유턴 하자.

Yeah... let’s make a U-turn there.

Dialogue 2

[모임에서 친구 두 명이 이야기를 나눈다.]

[Two friends are talking at a get-together.]

가: 예진이 올 줄 알았는데 아직 못 봤어. 이런 일을 좋아하지?

I thought Yejin was coming but I haven’t seen her yet. Doesn’t she like this kind of thing?

나: 아니 근데... 지호가 예진이랑 서준이 헤어졌다고 했어. 서준이도 여기에 초대받았지? 그래서 예진이가 안 온걸까?

No but... Jiho said that Yejin and Seojun broke up. Seojun was also invited, right? So maybe Yejin won’t come?

가: 뭐?! 언제 헤어졌어? 서준이는 둘이 결혼할 거라고 했어.

What? When did they break up? Seojun said they were going to get married.

나: 언제 그랬어?

When did he say that?

가: 몇주 전인 것 같아.

A few weeks ago, I think.

나:헐.. 근데 지난번에 예진이 짜증난 것 같더라. 예진이가 “개는 아무것도 안하고 매일 게임만 해”라고 했어

Wow.. but last time I think Yejin was annoyed. She said “He doesn’t do anything and only plays game everyday.”

가: 그래? 그런 사람인 줄 몰랐어. 근데 오늘 아직 안 보이는 걸 보니까 안 온 것 같긴 하다.

Really? I didn’t know he was that kind of person. But I haven’t seen him here yet so I think he might not come.

나: 나도 아직 못 봤어. 근데 이따 도착할 수도 있지.

I haven’t seen him here either. But he could arrive later.

가: 그래도 지금 없으니까 예진이 와야지! 걱정 안해도 되잖아. 메세지 보내볼까?

Regardless he’s not here now so Yejin should come! She won’t have to worry. Should I sent her a message?

나: 좋아!

Sounds good!

가: 아... “오늘동료들이랑 놀기로 해서 못 가지만 다들 밤 재밌게 보내!”라고 대답했어

Ah... She said “Today I’m out with my co-workers so I can’t go, but you guys enjoy your night!”

나: 아쉽다. 개한테 “너 없으면 재미없잖아’ 라고 보내줘

That’s too bad. Send “Without you it’s no fun.”

가: 우리 셋이 같이 논 지도 꽤 됐잖아. 조만간 저녁이나 같이 먹자

It’s been quite a long time since the three of us hung out. Let’s have dinner together!

나: 나 다음 주말에 시간 돼!

I have time next week!

가: 완전 좋다. 내 룸메이트도 고향 내려간다고 해서 집에 아무도 없어. 우리 집으로 와!

That's great! My roommate said she's going to her hometown so there is no one home. Come to my place!

Question and Answer Task

[예진 씨의 주말에 대한 이야기를 읽어보세요.]

Please read about Yejin's weekend.

이번 주말에 친구의 생일 파티를 위해 고향에 가요. 미루고 있던 선물을 사기 위해 쇼핑물에 가야 해요. 오늘은 금요일이지만 야근할 것 같아서 내일 선물을 사러 갔다가 기차를 탈 거예요. 기차는 오후 5시에 출발하기 때문에 선물을 살 시간이 충분할 것 같아요. 선물로는 친구가 좋아하는 사진에 관한 책이 있으니 제가 사줄 거예요. 파티는 오후 3시에 시작하기 때문에 아침에 선물을 포장할 시간이 있을 것 같아요.

1. 어디로 가고, 몇 시에 출발한다고 했어요?
2. 예진 씨의 말로, 예진 씨는 친구 선물로 무엇을 사려고 했어요?

고향에 가기 때문에 부모님 댁에도 방문할 예정이에요. 부모님을 마지막으로 뵈는 것은 지난 어버이날이었기 때문에, 부모님을 뵈는 것이 오랜만이에요. 아버지는 제가 집에 간다는 걸 알고 계시지만, 어머니께는 깜짝 선물이에요. 선물을 사러 나가는 동안 부모님 댁에 가져갈 간식도 고를 생각이에요. 이번 주말은 꽤 바쁘겠지만 정말 재밌었을 것 같아요.

1. 예진 씨는 주말에 뭘 한다고 했어요?
2. 예진 씨가 주말은 기대하고 있나요? 무슨 말을 통해 알 수 있나요?

This weekend I'm going to my hometown for my friend's birthday. I have to go to the mall to buy the birthday present I put off buying. Today is Friday but I think I have to work overtime so I have to buy the present tomorrow and then catch the train. Because the train leaves at 5pm, I think I'll have enough time to buy the present. As for the present, I am going to buy a photography book that my friend likes. Because the party starts at 3pm I think I will wrap it in the morning.

1. *Where is she going and what time did she say she was leaving?*
2. *In Yejin's words, what present is Yejin buying for her friend?*

Because I'm going to my hometown, I will visit my parent's house. The last time I saw my parent's was Parent's Day, so it's been a long time. My dad knows I am coming home but it is a surprise for my mom. While I'm buying my friend's birthday present, I'm thinking of buying snacks to bring to my parent's house. This weekend will be quite busy but I think it will be fun.

1. *What did Yejin say she was doing this weekend?*
2. *Is Yejin looking forward to this weekend? Through what words do you know this?*

Fill-in-the-Blank

이번 주는 많이 바쁠 거예요. 저는 아파트 관리자, 친구들과 메시지를 주고받고 있어요. 친구들은 4일 동안 제가 사는 도시에 방문할 거예요. 호텔 말고 우리 아파트에서 지내면, 돈을 아낄 수 있어요. 그래서 친구한테 “_____”라고 했어요. 근데 친구들이 도착하기 전에 아파트 관리를 해야해요. 며칠 전부터 온수 나오지 않았어요. 그래서,

아파트 관리자에게 _____라고 얘기했어요. 이틀 후에 제 친구들이 도착할 예정이라서, 우리 아파트 온수가 나오지 않는데, 괜찮냐고 물어봤어요. 친구들이 “_____”라고 대답했어요. 아파트 관리자는 자신의 내일 근무 시간을 메시지로 보냈어요. 관리자는 _____라고 했어요. 그 시간 중에서 결정만 하면 돼요. 저는 관리 담당자에게 _____라고 답했어요.

This weekend will be very busy. I have been sending messages to the maintenance man and my friends. For four days, my friends are visiting the city I live in. Instead of staying in a hotel, they can save money by staying at my apartment. So, to my friends I said “_____.” But before my friends arrive, I need maintenance for my apartment. For the last few days, I have not had hot water. So, to the apartment maintenance man I said _____. My friends arrive in two days, so I asked my friends if it was okay that my hot water wasn’t working. My friends replied “_____.” The maintenance man sent me his work hours for tomorrow. The maintenance man said _____. From those hours, I just need to pick a time. To the maintenance man, I replied _____.

Appendix D

Occurrence of prosodic cues divided by speaker

Table 1. CJR

Prosodic Cues		Direct speech		Indirect speech	
		Initial frame	DS seg	Initial frame	IDS seg
Boundary tones	<i>rising-falling/falling:</i>	3	1	0	0
	<i>rising/high:</i>	2	2	4	0
Pauses		1	0	0	0
Lengthening		1	4	0	0
Pitch reset		1			

Table 2. CJW

Prosodic Cues		Direct speech		Indirect speech	
		Initial frame	DS seg	Initial frame	IDS seg
Boundary tones	<i>rising-falling/falling:</i>	0	0	0	0
	<i>rising/high:</i>	2	0	2	0
Pauses		0	1	0	0
Lengthening		0	0	0	0
Pitch reset					

Table 3. CSJ

Prosodic Cues		Direct speech		Indirect speech	
		Initial frame	DS seg	Initial frame	IDS seg
Boundary tones	<i>rising-falling/falling:</i>	0	0	0	1
	<i>rising/high:</i>	3	1	4	0
Pauses		2	0	2	0
Lengthening		0	2	0	0
Pitch reset		1			

Table 4. GBJ

Prosodic Cues		Direct speech		Indirect speech	
		Initial frame	DS seg	Initial frame	IDS seg
Boundary tones	<i>rising-falling/falling:</i>	0	0	0	0
	<i>rising/high:</i>	2	1	5	0
Pauses		0	0	0	0
Lengthening		0	0	0	0
Pitch reset					

Table 5. KHJ

Prosodic Cues		Direct speech		Indirect speech	
		Initial frame	DS seg	Initial frame	IDS seg
Boundary tones	<i>rising-falling/falling:</i>	2	0	2	0
	<i>rising/high:</i>	0	0	1	0
Pauses		2	0	2	0
Lengthening		0	1	0	0
Pitch reset					

Table 6. OHJ

Prosodic Cues		Direct speech		Indirect speech	
		Initial frame	DS seg	Initial frame	IDS seg
Boundary tones	<i>rising-falling/falling:</i>	0	1	0	0
	<i>rising/high:</i>	2	0	2	0
Pauses		1	1	0	0
Lengthening		0	1	1	0
Pitch reset					

Co-occurrence of cues:*Table 7. Direct speech*

Prosodic Cues		Boundary tones (IF)		Boundary tones (DS segment)		Pause		Lengthening		Pitch reset
		LH/H	LHL/HL	LH/H	LHL/HL	IF	DS seg.	IF	DS seg.	
Boundary tones (IF)	LH/H					3		1		2
	LHL/HL					3		3		1
Boundary tones (DS seg)	LH/H						0		7*	
	LHL/HL						2		1	
Pause	IF	3	3					3		1
	DS seg			0	0				0	
Length.	IF	1	3			3				
	DS seg			7	1					
Pitch reset		2				1				

*Lengthening only occurring with an H tone (and no pause) was considered a weak boundary cue

Table 8. Indirect speech

Prosodic Cues		Boundary tones (IF)		Boundary tones (IDS segment)		Pause		Lengthening		Pitch reset
		LH/H	LHL/HL	LH/H	LHL/HL	IF	IDS seg.	IF	IDS seg.	
Boundary tones (IF)	LH/H					2		2		
	LHL/HL					4		2		
Boundary tones (IDS seg)	LH/H						0		0	
	LHL/HL						0		0	
Pause	IF	2	4					3		
	IDS seg			0	0				0	
Length.	IF	2	2			3	0			
	IDS seg			0	0		0			
Pitch reset										

Table 9. Single occurrence of prosodic cues

Cues/ RS type	Boundary tones (IF)		Boundary tones (DS/IDS seg)		Pauses		Lengthening	
	LH/H	LHL/HL	LH/H	LHL/HL	IF	DS/ID S seg	IF	Ds/ID S seg
DS	6	2	1			1		
IDS	13							

References

- Ahn, J.-H. (2006). Grammatical Discussion on Quotation Verbs focused on Modern Korean. *URIMALGEUL: The Korean Language and Literature*, 37, 143–169.
- Baek, H., & Yun, J. (2018). Prosodic Disambiguation of Syntactically Ambiguous Phrases in Korean. *MIT Working Papers in Linguistics*, 88, 89–100.
- Bermúdez, F. (2023). Using prosody to express evidentiality. The case of the quotative. *Journal of Pragmatics*, 214, 127–143. <https://doi.org/10.1016/j.pragma.2023.06.009>
- Bolden, G. (2004). The quote and beyond: Defining boundaries of reported speech in conversational Russian. *Journal of Pragmatics*, 36(6), 1071–1118.
<https://doi.org/10.1016/j.pragma.2003.10.015>
- Boersman, P. & Weenink, D. *Praat: doing phonetics by computer*. (Version 6.4.19).
www.praat.org
- Brown, L., & Yeon, J. (2015). Varieties of Contemporary Korean. In L. Brown & J. Yeon (Eds.), *The Handbook of Korean Linguistics* (pp. 459–476). John Wiley & Sons, Ltd.
<https://doi.org/10.1002/9781118371008.ch26>
- Buchstaller, I. (2017). Reported Speech. In A. Barron, Y. Gu, & G. Steen (Eds.), *The Routledge Handbook of Pragmatics* (1st ed., pp. 399–416). Routledge.
- Elfner, E. (2015). Recursion in prosodic phrasing: Evidence from Connemara Irish. *Natural Language & Linguistic Theory*, 33(4), 1169–1208.
<https://doi.org/10.1007/s11049-014-9281-5>
- Estellés-Arguedas, M. (2015). Expressing evidentiality through prosody? Prosodic voicing in reported speech in Spanish colloquial conversations. *Journal of Pragmatics*, 85, 138–154.
<https://doi.org/10.1016/j.pragma.2015.04.012>

Féry, C. (2016). *Intonation and Prosodic Structure*. Cambridge University Press.

<https://doi.org/10.1017/9781139022064>

Günthner, S. (1999). Polyphony and the ‘layering of voices’ in reported dialogues: An analysis of the use of prosodic devices in everyday reported speech. *Journal of Pragmatics*, 31(5), 685–708. [https://doi.org/10.1016/S0378-2166\(98\)00093-9](https://doi.org/10.1016/S0378-2166(98)00093-9)

Hirschberg, J. (2002). Communication and prosody: Functional aspects of prosody. *Speech Communication*, 36(1–2), 31–43. [https://doi.org/10.1016/S0167-6393\(01\)00024-3](https://doi.org/10.1016/S0167-6393(01)00024-3)

Hirschberg, J., Beňuš, Š., Gravano, A., & Levitan, R. (2020). Prosody in Discourse and Speaker State. In C. Gussenhoven & A. Chen (Eds.), *The Oxford Handbook of Language Prosody* (p. 468-476). Oxford University Press.

<https://doi.org/10.1093/oxfordhb/9780198832232.013.31>

Holt, E. (2009). Reported speech. In *The Pragmatics of Interaction* (pp. 190–205). John Benjamins Publishing Company.

<https://www.degruyterbrill.com/document/doi/10.1075/hoph.4.11hol/html>

Ishihara, S., & Kalivoda, N. (2022). Match Theory: An overview. *Language and Linguistics Compass*, 16(1), e12446. <https://doi.org/10.1111/lnc3.12446>

Ito, J., & Mester, A. (2012). Recursive prosodic phrasing in Japanese. In T. Borowsky, S. Kawahara, T. Shinya, & M. Sugahara (Eds.), *Prosody matters: Essays in honor of Elisabeth Selkirk*, 280-303.

Jeon, H.-S. (2015). Prosody. In L. Brown & J. Yeon (Eds.), *The Handbook of Korean Linguistics* (pp. 41–58). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118371008.ch3>

Jun, S.-A. (1996). *The phonetics and phonology of Korean prosody : intonational phonology and prosodic structure*. Garland Pub.

- Jun, S.-A. (1998). The Accentual Phrase in the Korean prosodic hierarchy. *Phonology*, 15(2), 189–226. <https://doi.org/10.1017/S0952675798003571>
- Jun, S.-A. (2000). K-ToBI (Korean ToBI) Labelling Conventions. *UCLA Working Papers in Phonetics*, 99. <https://linguistics.ucla.edu/people/jun/ktobi/k-tobi.html>
- Jun, S.-A. (2005a). Korean Intonational Phonology and Prosodic Transcription. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 201–229). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199249633.003.0008>
- Jun, S.-A. (2005b). Intonational Phonology of Seoul Korean Revisited. *UCLA Working Papers in Phonetics*, 104, 14–25.
- Jun, S.-A. (2022). The ToBI Transcription System: Conventions, Strengths, and Challenges. In J. Barnes & S. Shattuck-Hufnagel (Eds.), *Prosodic Theory and Practice* (pp. 151–181). The MIT Press. <https://doi.org/10.7551/mitpress/10413.003.0007>
- Jun, S.-A., & Jiang, X. (2019). Differences in prosodic phrasing in marking syntax vs. focus: Data from Yanbian Korean. *The Linguistic Review*, 36(1), 117–150. <https://doi.org/10.1515/tlr-2018-2009>
- Kim, S. (2012). *Interaction, grammar, and stance in reported speech* [Doctoral dissertation, UCLA]. <https://escholarship.org/uc/item/0rh43329>
- Klewitz, G., & Couper-Kuhlen, E. (1999). Quote – unquote? The role of prosody in the contextualization of reported speech sequences. *Pragmatics*, 9(4), 459–485. <https://doi.org/10.1075/prag.9.4.03kle>
- Kwon, J. -I. (1998). Historical change of the quotation sentence constructions in Korean. *Journal of the Linguistic Society of Korea*, 22(0), 59–79.

- Kwon, N., Polinsky, M., & Kluender, R. (2006). Subject Preference in Korean. *Proceedings of the 25th West Coast Conference on Formal Linguistics*.
- Lee, K.-M., & Ramsey, S. R. (2011). Late Middle Korean. In *A History of the Korean Language* (pp. 100–240). Cambridge University Press.
<https://doi.org/10.1017/CBO9780511974045.007>
- Malibert, I.-I., & Vanhove, M. (2015). Quotative constructions and prosody in some Afroasiatic languages: towards a typology. In A. Mettouchi, M. Vanhove, & D. Caubet (Eds.) *Corpus-based Studies of Lesser-described Languages: The CorpAfroAs corpus of spoken AfroAsiatic languages*. (Vol. 68, pp. 117-169). John Benjamins Publishing Company.
<https://doi.org/10.1075/scl.68>
- Myrberg, S. (2013). Sisterhood in prosodic branching. *Phonology*, 30(1), 73–124.
<https://doi.org/10.1017/S0952675713000043>
- Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Foris Publications.
- Oliveira Jr., M. & Cunha, D. (2004). Prosody as marker of direct reported speech boundary. *Speech Prosody 2004*, 263–266. <https://doi.org/10.21437/speechprosody.2004-61>
- Park, M. J. (2003). *The meaning of Korean prosodic boundary tones* [Doctoral dissertation, University of California, Los Angeles].
<https://www.proquest.com/docview/305339124/abstract/2AA14FE6EEE042D8PQ/1>
- Pierrehumbert, J. B., & Beckman, M. E. (1988). *Japanese tone structure*. MIT Press.
- Rhee, S. (2007). Through a borrowed mouth: Reported speech and subjectification in Korean. *LACUS Forum*, 34, 201–211.
- Rhee, S., & Koo, H. J. (2020). From quotation to surprise: The case in Korean. *Journal of Pragmatics*, 155, 83–100. <https://doi.org/10.1016/j.pragma.2019.10.006>

- Sandler, W., Lillo-Martin, D., Dachkovsky, S., & Müller de Quadros, R. (2020). Sign Language Prosody. In C. Gussenhoven & A. Chen (Eds.), *The Oxford Handbook of Language Prosody* (pp. 104-122). Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780198832232.013.44>
- Selkirk, E. (1986). On Derived Domains in Sentence Phonology. *Phonology Yearbook*, 3, 371–405. <https://doi.org/10.1017/S0952675700000695>
- Selkirk, E. (2011). The Syntax-Phonology Interface. In J. Goldsmith, J. Riggle, & A. Yu (Eds.), *The Handbook of Phonological Theory* (pp. 435–484). John Wiley & Sons, Ltd.
<https://doi.org/10.1002/9781444343069.ch14>
- Sohn, S.-O. (2015). The emergence of utterance-final particles in Korean. In S. Hancil, A. Haselow, & M. Post (Eds.), *Final Particles* (Vol. 284, pp. 181–196). DE GRUYTER.
<https://doi.org/10.1515/9783110375572-008>
- Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, 25 (7–9), 905–945.
<https://doi.org/10.1080/01690961003589492>
- Yeon, J., & Brown, L. (2019). Quotations. In *Korean* (2nd ed., pp. 446-468). Routledge.
<https://doi.org/10.4324/9781315160351-10>
- Yoon, J. H. S. (2018). Korean Syntax. In J. H. S. Yoon, *Oxford Research Encyclopedia of Linguistics*. Oxford University Press.
<https://doi.org/10.1093/acrefore/9780199384655.013.171>