Working papers in Applied Linguistics and Linguistics at York 6, 1–33

On the Prosody of Reported Speech in Seoul Korean

Makaila Johnston[1]
York University, Toronto, Canada

Abstract: This paper investigates the use of prosodic cues to indicate reported speech (RS) in Seoul Korean. Previous research on reported speech prosody shows that speakers use a variety of prosodic cues to delineate the boundaries of reported speech. Korean differs from many of the languages in the RS prosody literature in the framing of RS segments, in that the RS segment is embedded in a structure with both an initial and end frame, and a quotative morpheme follows the RS segment. The results show that in direct speech examples, two prosodic boundaries are marked, with the onset of the direct speech being indicated by a strong boundary, and the offset by a weak boundary. Indirect speech, in contrast, only has one prosodic boundary marked. The location of these prosodic boundaries as well as their strength can be predicted by the type of φ-phrase and its relationship to adjacent φ-phrases.

Keywords: prosody; reported speech; Korean; recursivity; syntax-phonology interface

1 Introduction

During a conversation, a speaker may need to relay the words of another. This is commonly known as a quotation, or reported speech (RS). In writing, use of quotation marks helps delineate quotes from the text they are embedded in. For spoken language, structural cues as well as prosodic cues aid in indicating that a quotation, or reported speech, is to follow. Prosody is the suprasegmental features of language—intonational elements such as pitch, duration, intensity, and tonal structure fall into its realm (Féry, 2016). The role of prosody 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 conversations (Hirschberg, 2002). Prosody may also flag reported speech segments. This paper investigates the use of prosody in the indication of reported speech, and its relationship with structural elements, in Seoul Korean (henceforth "Korean").

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

[1] Corresponding Author: makailaj@yorku.ca

Recommended Citation: Johnston, M. (2026). On the prosody of reported speech in Seoul Korean. *Working papers in Applied Linguistics and Linguistics at York*, *6*, 1–33. https://doi.org/10.25071/2564-2855.56

 ISSN 2564-2855

end frame, with the *verbum dicendi* (the 'verb of saying') following the RS segment. The initial frame can be optional; however, as I am interested in both boundaries of reported speech in Korean, I will not be analysing any examples that exclude the initial frame. Korean has a structural morpheme, indicating "quote," following a reported speech segment. How this quotative morpheme may influence intonational elements of the sentence, and vice versa, is of interest for this paper. Consequently, the goal of this paper is to uncover the extent to which 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? 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 (DS) and indirect speech (IDS) 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.

Based on these hypotheses, we can also extrapolate the prosodic parsing of the RS utterances. In IDS examples, the final verb of the segment and quotative marker should form a prosodic unit, likely an Accentual Phrase (AP). In contrast, the final verb of a DS segment and the quotative marker should be separate prosodic units, likely separate APs. DS is a reproduction of an utterance, which leads me to believe that it should form its own Intonational Phrase (IP). Should this be true, it would imply that embedding occurs within the prosodic structure, as the DS IP would be embedding within the matrix IP—mirroring the syntactic structure. 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)? 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 (including how it may be used to analyse reported speech prosody in Korean) is addressed in more detail in Section 6.

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). The goal of this paper is to provide a preliminary description of reported speech prosody in Korean, and, in turn, provide an answer to these questions.

2 Background

2.1 Reported Speech

The structure of reported speech follows the same pattern in a variety of languages. In many of them, the introduction of a reported speech segment employs an introductory verb, also called *verbum dicendi* (Buchstaller, 2017; Holt, 2009). The reported segment then follows. Reported speech can be divided into three types: direct reported speech, indirect reported speech, and quasi-indirect speech (Holt, 2009; Klewitz & Couper-Kuhlen, 1999). The scope of this paper only covers direct speech (DS) and indirect speech (IDS). For clarity, I will use Buchstaller's (2017) term for the quoter, referring to them as the "reporting speaker." The quoted will be referred to as the "original speaker." In the differentiation of reported speech types, the vantage point of the segment is a large factor; DS is conveyed from the perspective of the original source, whereas IDS is conveyed from the perspective of the reporting speaker. Simply put, IDS is akin to paraphrasing the reported speech, whereas DS is a recreation, pronouns, tense, and all. The sentences in (1) demonstrate the differences between DS and IDS in English.

(1) Direct vs. Indirect speech
  (a) I saw Mary today, she said "I want to go see a movie tomorrow"
  (b) I saw Mary today, she said that she wants to go see a movie tomorrow.

The information being conveyed in (1a) and (1b) is essentially the same: Mary wants to see a movie tomorrow. The presence of the *verbum dicendi* "said" can be seen in both examples preceding the RS segment, and the original speaker, Mary, is mentioned. However, in (1a), we assume that Mary's speech is being replicated as it was uttered by her, demonstrated by the first-person pronoun. The first-person pronoun in the frame does not refer to the same referent as the first-person singular pronoun in the DS segment. The pronoun in the DS segment refers to the original speaker, Mary, not the reporting speaker. In contrast, (1b) shows that the reporting speaker uses the third person singular pronoun "she" to refer to the original speaker, Mary. Compared to (1a), the IDS segment in (b) has been adopted into the perspective of the reporting speaker. In addition to perspective, the assumption of how closely the reported speech is reproduced differs between DS and IDS. Holt (2009) explains that the speaker claims to reproduce an authentic version of the utterance in DS, but does not purport this kind of reproduction in IDS. In sum, both DS and IDS perform the same function of reporting speech, albeit in different ways.

2.1.1 Reported Speech in Korean

Reported speech in Korean can be expressed as direct or indirect speech. Korean is an agglutinating language, and thus the morphology and syntax of the language are intertwined (Yoon, 2018). The structure of reported speech in Korean involves a quotative morpheme attaching to or following the root verb of the RS segment, followed by the *verbum dicendi* such as *malhata* 'say' or *hata* 'do' (Yeon & Brown, 2019). This structure is illustrated below.

(2) Reported speech in Korean (Modified from Kim, 2012)
  [Reporting speaker: Original speaker] {reported speech}$_{\text{-QM}}$[2] *quoting verb*
  ↑ Initial frame ↑ *verbum dicendi*

[2] QM refers to "quotative marker."

The initial frame shown in (2) includes both the reporting speaker and original speaker. The original speaker is syntactically optional (Kim, 2012). The initial frame may also include a recipient (e.g., "I said to the cashier…").

The quotative morpheme, *-ko*, developed from a construction involving two verbs of saying which was then reduced to the form seen in Modern Korean (Ahn, 2006; Kwon, 1998; Rhee, 2007; Rhee & Koo, 2020). The quotative morpheme attaches to a group of affixes called Final Endings, or closing affixes, which turn a verbal complex into a free form (Yoon, 2018). The Final Ending varies depending on sentence type. Rhee (2007) and Ahn (2006) analyse *-tako* (declarative), *-lako* (imperative), *-nyako* (interrogative), and *-cako* (hortative) as units together, but where Rhee (2007) considers them complementizers, Ahn (2006) 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 one; it 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's (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 (QM). The quotative markers, *-ko* (IDS) and *lako* (DS), occupy the C head in the syntax. This will be relevant for the Match Theory analysis in Section 6.

Due to the morphological variation of the suffixes on the verb influenced by sentence type, and the different intonational patterns associated with them (e.g., H% for interrogatives), declaratives will be the focus of this study. Table 1 provides a breakdown of the possible forms being investigated.

For indirect speech segments, the quotative marker will affix to the verbal root (e.g., *patnunta-ko* 'receive-QUOT', *kanta-ko* 'go-QUOT'). For direct speech segments, the fully inflected verb will be followed by the quotative marker *lako*, and sometimes *hako*. The ways in which the quotative marker interacts with DS and IDS differ structurally. The quotative marker in IDS examples cannot separate from the verbal root of the IDS segment—they are bound. However, the verbal complex of DS segments is fully inflected and a free form, with the quotative marker floating in the vicinity. Technically, the QM *lako* should be a bound morpheme and unable to stand alone, but it also cannot attach to a Final Ending affix. I have not found any literature providing a syntactic analysis of this phenomenon.

2.2 Korean Prosody

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),

Table 1

Quotative Markers

Reported speech type	Quotative marker
Indirect speech	*-ko*
Direct speech	*lako* *(hako)*

the Intermediate Phrase (iP), and the Intonational Phrase (IP) (Jeon, 2015). 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.

The Accentual Phrase 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 for a four-syllable word is THLH, where T is determined by the laryngeal features of the initial segment—aspirated consonants, fortis consonants, /s/, or /h/ result in a H tone, whereas a L tone occurs elsewhere (Jeon, 2015; Jun, 1998, 2005a, 2005b). A four-syllable AP will realize all the tones of the contour, but in other APs, 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. For APs shorter than four syllables, medial tones may be undershot, and which tones are chosen to be undershot vary (Jun, 2005a). IP-final APs will have their final tone preempted by the IP boundary tone(s) (Jun, 1996). In total, Korean has 14 surface contours for an AP (Jun, 2005a).

The Intermediate Phrase (iP) is a prosodic unit higher than the AP, but lower than the IP (Jeon, 2015). 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 narrowly focused constituents, words that are prominent can also 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.

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 10 have been claimed to exist (Jeon, 2015; Park, 2003). The nine possibilities are: L%, H%, LH%, HL%, LHL%, HLH%, LHLH%, HLHL%, and 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).

2.3 Reported Speech and Prosody

Pitch changes, pauses, and boundary tones appear consistently as prosodic cues of reported speech in a variety of languages. However, there are cues that may not be as widespread. Klewitz and 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. Decrease in speech rate was found in Spanish by Estellés-Arguedas (2015) as well as by Bermúdez (2023), though Bermúdez (2023) proposes that the slower speech rate may convey irony, rather than having an association with reported speech. Oliveira and Cunha (2004) reported a drop in intensity at the boundary of direct speech segments in Brazilian Portuguese. Additionally, Oliveira and Cunha (2004) find that boundary tones play a part in framing, sometimes used in conjunction with pauses. The use of prosodic cues does not seem to be conventionalized, and they are not systematically used to signal reported speech (Klewitz & Couper-Kuhlen, 1999).

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 (e.g., Brazilian Portuguese: Oliveira & Cunha, 2004; Beja: Yatziv-Malibert & Vanhove, 2015; English: Klewitz & Couper-Kuhlen, 1999; German: Kasimir, 2008; Hungarian: Bermúdez, 2023; Modern Hebrew: Yatziv-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.

It is important to differentiate between linguistic and paralinguistic elements of a sentence's melody. One may imagine that a change of pitch or voice quality is a result of mimicry. Günthner (1999) analysed the way in which speakers use prosody and voice quality techniques to convey the speech of multiple characters, and makes a distinction between prosody and voice quality. Prosody refers to loudness, duration, pitch, and pause, whereas voice quality is a paralinguistic element used by the speaker to produce different types of voices (i.e., whispery, falsetto, breathy, etc.) temporarily. 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 rise 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 21 participants in his study raised the pitch of a reported speech segment, regardless of the gender of the original speaker. Either a raising or lowering of pitch was found by Estellés-Arguedas (2015) in the study of marked prosody of direct speech in Spanish, and in Klewitz and Couper-Kuhlen's (1999) study of reported speech in English. Oliveira and Cunha (2004) found that in Brazilian Portuguese, pitch reset was the most significant cue in marking DS boundaries. Yatziv-Malibert and 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.

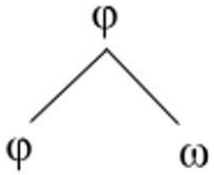
In the orthography of many languages, quotation marks are used in written texts to mark the boundaries of direct speech. In verbal communication, prosodic cues can work as "flags" (Klewitz & Couper-Kuhlen, 1999) to prosodically frame reported speech segments. This idea of framing, or marking of boundaries, is investigated in Brazilian Portuguese (Oliveira & Cunha, 2004), English (Klewitz & Couper-Kuhlen, 1999), Russian (Bolden, 2004), and German (Kasimir, 2008). Yatziv-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 that 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 (2004) 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). I will be using the terms "initial frame" 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, p. 140). Klewitz and 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 and 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 both before and after the reported speech segment were commonly used in Spanish and Swedish. Pauses before the reported speech segment were also observed in Polish and Hungarian, though they did not appear to employ pauses at the end of the RS segment. Pauses set off the initial frame in Beja, and pauses can set off the end of the reported speech in Zaar and Juba Arabic (Yatziv-Malibert & Vanhove, 2015).

2.4 Match Theory

Match Theory assumes the Optimality-Theoretic framework of a constraint-based ranking system; it maps syntactic trees onto prosodic trees, assuming a correlation of CPs, XPs, and X^0s of syntax with intonational phrases (ι), phonological phrases (φ), and prosodic words (ω) of the prosodic hierarchy (Ishihara & Kalivoda, 2022). The correspondence between these two structures comes from MATCH constraints, and misalignments 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 (Ishihara & Kalivoda, 2022). However, the work by Jun (2005a) and others assumes that the intonational phonology of Korean is organized based on the Strict Layer Hypothesis (Nespor & Vogel, 1986; Pierrehumbert & Beckman, 1988; Selkirk, 1986). 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:

(3)



Strict layering bans recursivity, so a prosodic unit cannot dominate the same category type (Selkirk, 2011). This violation is exemplified in (3). In Section 6, I suggest that a recursive analysis may better explain the phenomena found in my RS data, based on Match Theory (proposed by Selkirk, 2011). I propose constraints as well as a suggested ranking in order to predict why and where prosodic cues have occurred in the data.

3 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. 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). (See Appendix B for a description of the tones relevant to the data.)

3.1 Participants and Procedure

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. 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 analysed using Praat (Boersma & Weenink, 2024).

3.1.1 Tasks

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. Between the two dialogues, role A and role B each had three lines of each reported speech type.

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. In order to avoid hesitation from the participants, two questions were posed 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.

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 (e.g., "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 spontaneously while reading. They were then asked to read the text, including their additions to the blanks.

4 Results

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 were more varied than those used with indirect speech. Table 2 shows the number of sentences produced by each speaker, how many instances of direct or indirect speech were produced, and the number of sentences analysed. Though there was a variation in age, the oldest speaker's production of reported speech was consistent with the younger speakers' utterances; no generational differences in the production of reported speech were observed in this study. 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 S6, 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 S2 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 with another cue, though a H tone boundary requires a pause to be considered a strong boundary. This is due to the 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

Table 2

Number of Sentences Produced and Analysed by Participant by Reported Speech Type

Participant	# of sentence produced		# of sentences analysed	
	DS	IDS	DS	IDS
S1	5	5	5	4
S2	5	7	2	2
S3	5	7	3	5
S4	4	8	2	5
S5	2	9	2	3
S6	2	9	2	3
Total	23	45	16	22

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. Section 4.1 discusses the results of direct speech examples, and Section 4.2 discusses the results of indirect speech examples.

4.1 Direct Speech Results

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.

The boundary tone of the initial frame patterned in two ways: a rising-falling/falling tone (LHL or HL) and a rising/high tone (LH or H). Pauses and lengthening were also employed to set off the initial frame from the DS segment. The rising/high tones were optionally followed by a pause, whereas the rising-falling/falling tones were almost always followed by a pause. Participants varied with regards to the amount and types of prosodic cues they employed. For instance, participants S3 and S5 consistently used the same boundary tones (H and LHL respectively), and participants S6 and S5 also employed lengthening and pauses at the initial frame boundary.

An example from participant S6 is illustrated in Figure 1. The final syllable of the initial frame has a rising tone (LH), where it starts low and rises within the syllable, near the end. The segment is also audibly lengthened and followed by a pause. The rising tone along with the pause and lengthening are prosodic cues consistent with an IP boundary. The end boundary of the DS segment is marked by the boundary tone HL, followed by a pause. No lengthening occurred on the final syllable of the DS segment.

Figure 1

S6 Direct Speech

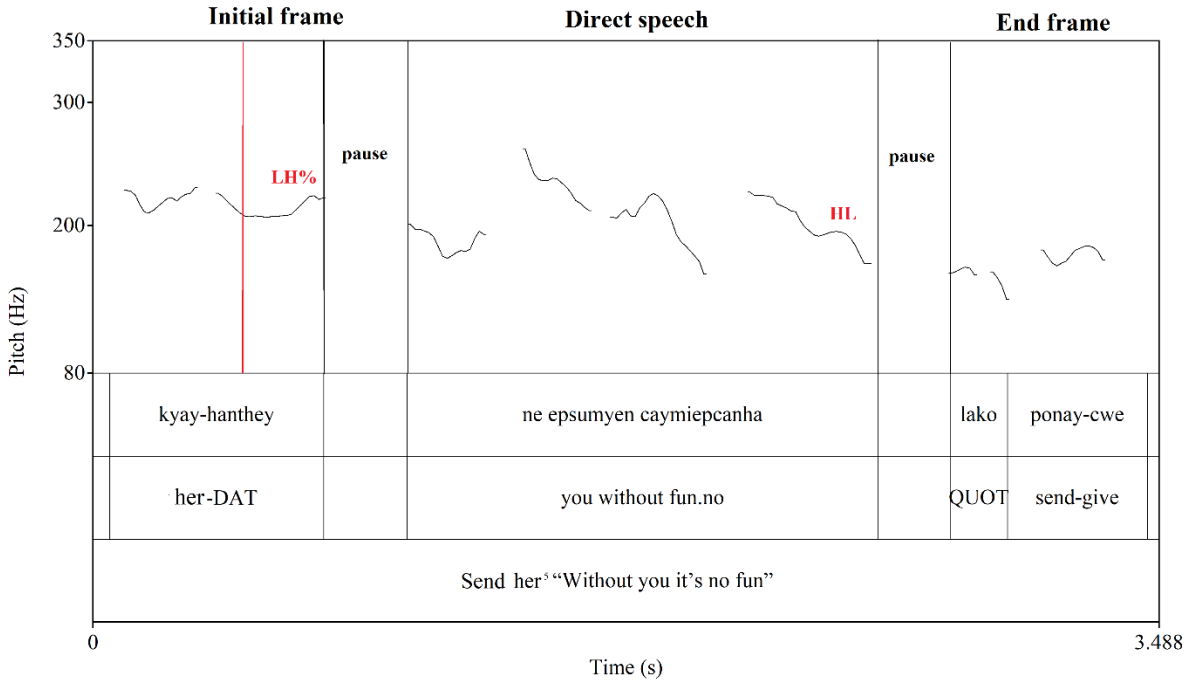
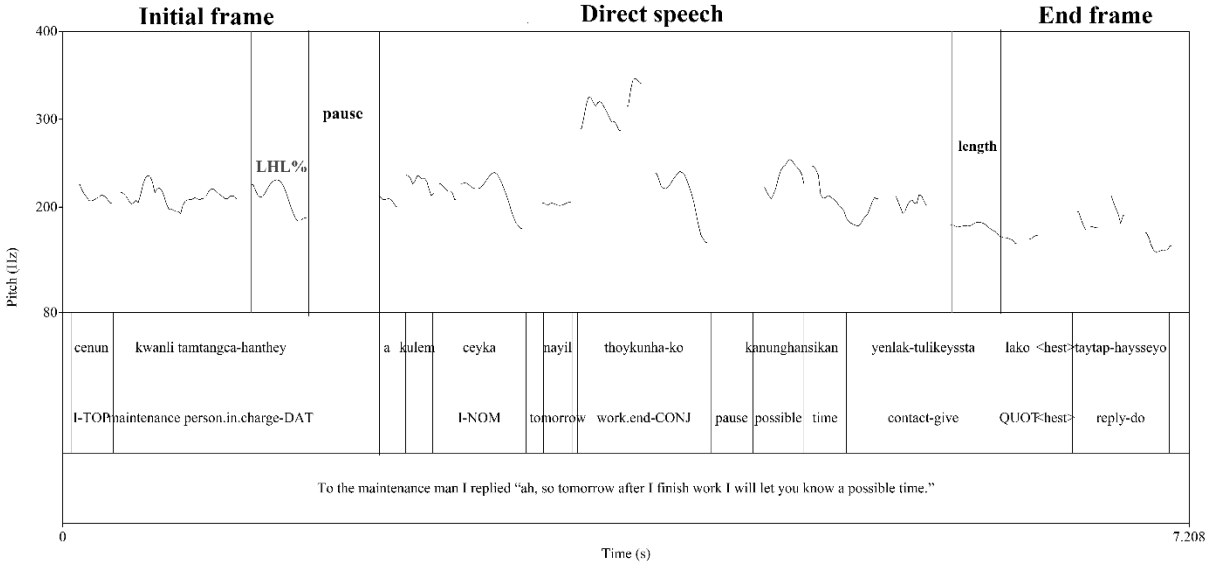


Figure 2 is an example of direct speech with an initial frame boundary marked by a rising-falling tone. The pitch track in Figure 2 provides an example from participant S5, who consistently

Figure 2

S5 Direct Speech



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.

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. This can be observed in Figure 2.

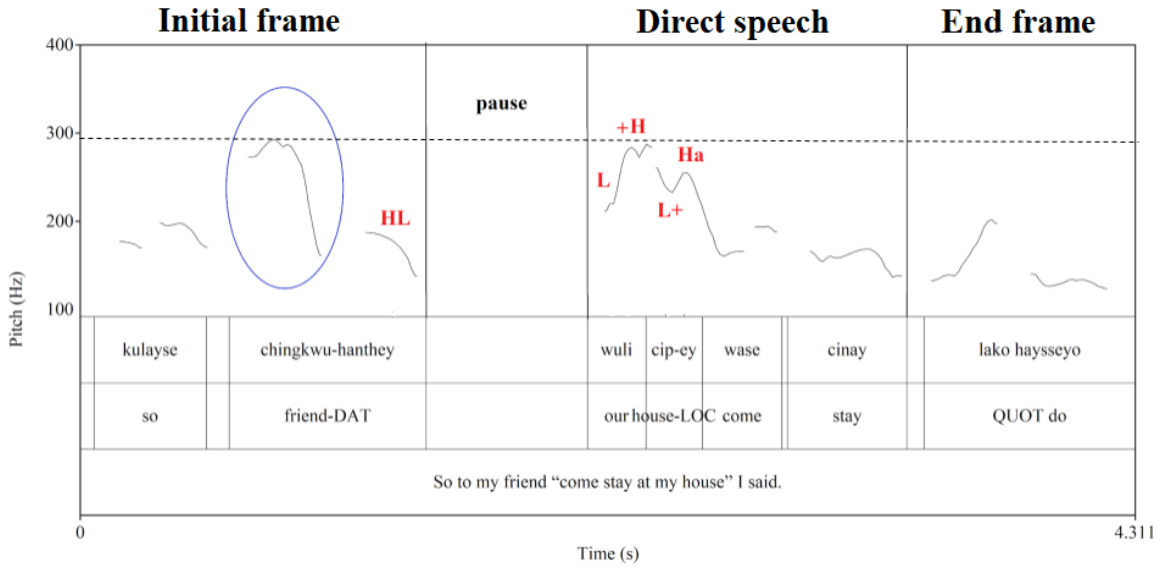
The fact that the end of a quoted segment is indicated structurally by the quotative marker may have influenced how strongly speakers felt the need to mark the "unquote" prosodically. The direct speech segment often ended low, as seen in Figure 1; however, some examples ended high. In all cases, the *verbum dicendi* was pronounced with the lowest pitch of the utterance, at the low end of the participant's pitch range. This sometimes resulted in a creaky voice, particularly with participant S3. This is consistent with typical declarative sentence endings in Korean. Occasionally, the *verbum dicendi* seemed to be produced at a faster speech rate than the rest of the sentence.

The direct speech segment may have undergone pitch reset. 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 this reason, a rise in pitch cannot be relied upon as the sole indicator of a reported speech segment. In 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.

Figure 3 is a direct speech example produced by participant S1 and illustrates possible pitch reset of the direct speech segment, comparing it to the high pitch of the aspirated consonant present

Figure 3

S1 Direct Speech



in the initial frame. The subject matter of the tasks meant that the word *chinkwu* 'friend' was often present in the initial frame; the first segment (<ch>, /tɕʰ/) is aspirated and therefore triggers a H tone. This H tone was often comparable to the pitch of the first H of the first intonational contour (AP) of the DS segment. The circled pitch in Figure 3 is the highest of the utterance. 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 may occur as a prosodic cue for direct speech in Korean, although unreliably so.

Prosodic cue types and their frequency were analysed at both the initial frame and end frame boundaries. At the initial frame boundary, lengthening most often occurred when the boundary tone was LHL; lengthening was likely a product of the tonal contour. However, the prosodic boundary at the end of the direct speech segment shows slightly different results. Lengthening at the direct speech segment boundary almost always occurred at a weak boundary, and was often accompanied by a rising/high tone.

To conclude, two boundaries are marked prosodically in sentences containing direct speech: the onset and the offset of the direct speech section. Lengthening is present at both boundaries, with pauses more often used with a strong boundary at the onset of direct speech. Weak prosodic boundaries often occurred at the offset of the direct speech. Pitch reset may have been observed in the direct speech segment, though it remains unreliable due to the influence of segment type on F0 values. The *verbum dicendi* ended with a low boundary tone, consistent with the low IP boundary associated with the end of declarative sentences, and was sometimes produced with a creaky voice.

4.2 Indirect Speech Results

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 only included the use of boundary tones, rather than other cues such as pauses or lengthening. Indirect speech segments did not appear to undergo pitch reset, exhibiting a downtrend across the entire sentence.

Almost all of the initial frames had a boundary tone ending high (LH or H). Three different speakers all produced one example of a falling tone at the edge of the initial frame, but unlike in direct speech, the pauses that occasionally accompanied the falling tone in indirect speech were inconsistent. 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 4), and a sentence with the H tone at the end of the initial frame (Figure 5), produced by the same speaker. The following Figure 4 shows participant S3's use of a falling tone at the end of the initial frame.

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 S4, there was an audible intake of breath. However, Figures 4 and 5 may demonstrate some similarities with the patterns seen in direct speech examples. Falling tones at the initial frame boundary often co-occurred with pauses in direct speech, whereas high boundary tones co-occurred with pauses less often. Figure 4 shows a falling boundary tone and pause, and Figure 5 shows a high tone, with no pause, at the end of the initial frame.

Figure 4

S3 Indirect Speech with a Falling Tone at the Initial Frame Boundary

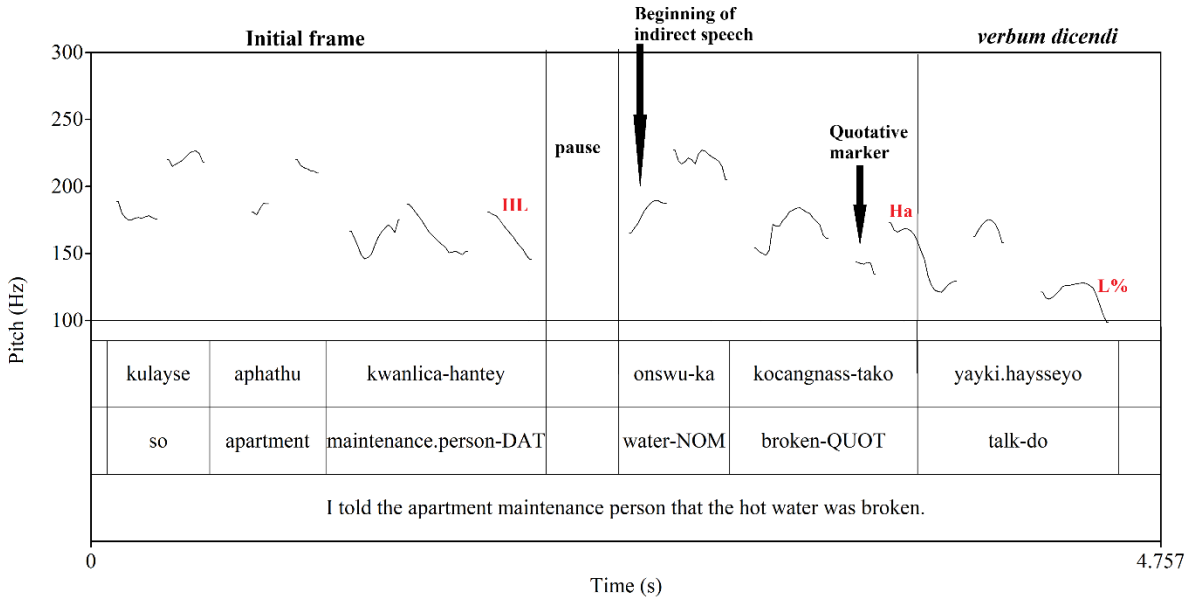
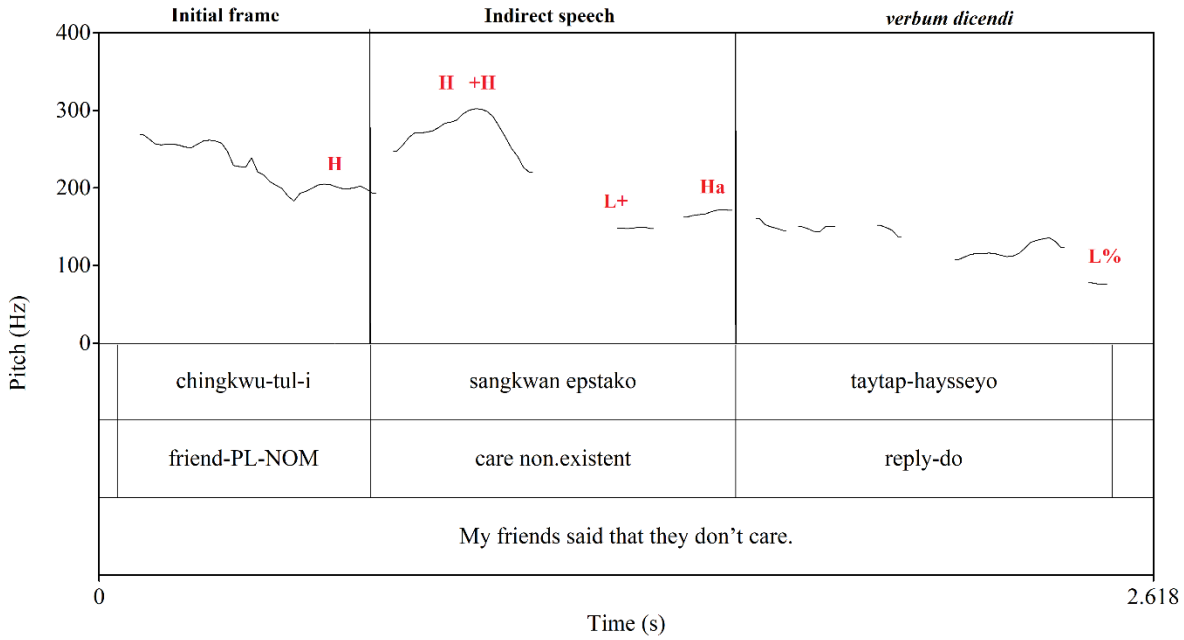


Figure 5

S3 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 5.

The occurrence of prosodic cues at a strong or weak boundary in indirect speech was also analysed. The rising/high boundary tones are the most common and most often appear at a weak boundary. The rising-falling/falling boundary tones never appear at a weak initial frame boundary. In one indirect speech example, participant S5 produced a LHL boundary followed by a pause at the end of the indirect speech segment. As the end boundary is not flagged by any other speaker, nor produced more than once by participant S5, I assume this example to be an outlier. As such, I will assume that the end frame boundary does not occur 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. With the singular exception of one utterance produced by participant S5, the end boundary 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 typical end of a declarative sentence.

4.3 Occurrence of Cues in Direct and Indirect Speech

In terms of number of prosodic cues present in both types of reported speech, Table 3 provides each type of prosodic cue observed, which boundary it occurred at, and how many instances were present in the sentences analysed.

The columns "DS/IDS seg" refer 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 relevant here are: LHL, HL, LH, and H. A low boundary tone (L) was not marked in Table 3, 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 4.2. 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. Pitch reset was an unreliable cue, and the total number included here was for utterances in which it appeared to occur. Pitch reset is further discussed in Section 7.

Table 3

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	LHL/HL	5	2	6	1
	LH/H	11	8	16	—
Pauses		6	2	6	0
Lengthening		4	8	4	0
Pitch reset		3		0	

5 Summary and Discussion

The results show that prosodic cues appear to be used to delineate some reported speech boundaries. These prosodic cues were used to varying degrees by participants to indicate the onset of direct speech, and often the offset as well. DS examples had both strong and weak initial frame boundaries, and a weak end boundary often indicated by lengthening. IDS examples, on the other hand, had minimal or no prosodic cues at the initial frame boundary, and almost never marked the end boundary prosodically. For instance, participant S1 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.

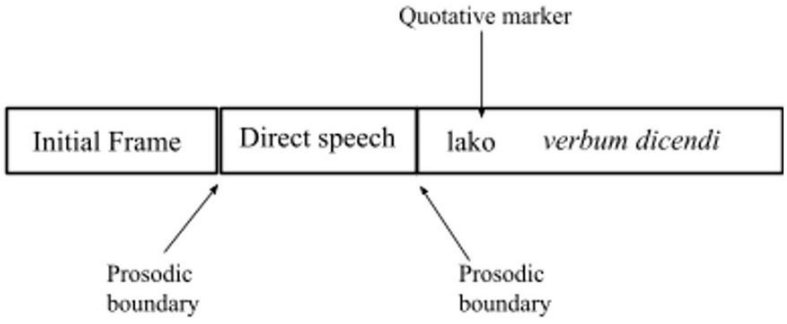
Two types of boundary tones were observed at the edge of the initial frame: rising-falling/falling (LHL/HL) or rising/high (LH/H). The rising/high pattern was the most common boundary tone observed at the edge of the initial frame in both DS and IDS examples. One participant, S1, tended to use a falling 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. Pitch reset seemed to occur in some DS examples, but due to its unreliable nature, it cannot be used as the sole cue for reported speech.

I hypothesized that the edge of a direct speech segment may align with the edge of an IP as the DS segment is syntactically an embedded clause. If so, does the DS segment reflect the expected prosody at the edge of the IP? In short, no. Though some speakers produce a low or falling pitch on the final syllable of the direct speech segment, it cannot be definitively analysed as the low boundary tone that is associated with most IPs. In all of the sentences, the quotative marker that follows the DS segment 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 DS segment could be a natural fall to meet the following low target. Furthermore, an IP-final L% could be expected to be at the lowest range of an utterance; this was never observed as the lowest pitch always occurred on the utterance-final syllable. Other cues associated with IP boundaries, such as pauses and lengthening, occurred at the ends of the initial frame and direct speech segment. This is an interesting discovery, as only iPs or IPs are associated with lengthening, based on previous research on Korean prosody. Thus, these units cannot be analysed as APs. Does the direct speech trigger an Intermediate Phrase, or can it be explained another way? In Section 6, I will address this question, providing an explanation using Match Theory.

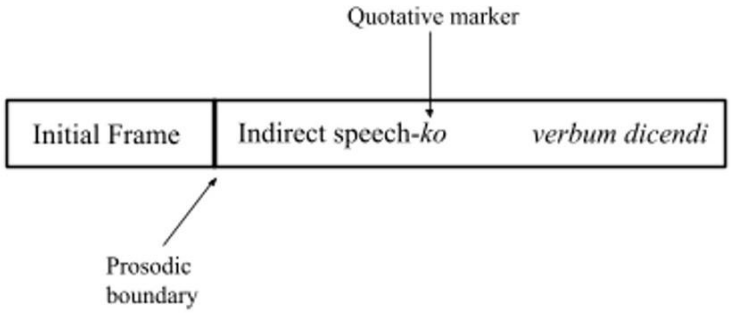
The behaviour of the quotative marker in DS versus IDS may explain the difference in the use of prosodic cues observed in these two types of reported speech. Recalling Section 2.1.1, *-ko* is a quotative clitic developed from Middle Korean (Kwon, 1998) that is part of a group of affixes called Sentence-Final Particles (SFP; Yoon, 2018). SFPs can attach to the group of affixes called Final Endings (Yoon, 2018), which the *-(nu)nta/ta* morpheme is a part of; both affix to verbal complexes. In direct speech, the quotative clitic *-ko* cannot attach to the verbal complex as it is already fully inflected. As *-ko* cannot stand on its own, I assume that it must be strengthened by including the morpheme *la,* in order to fulfil a morphosyntactic requirement for bound roots. Thus, the verbal complex and the strengthened quotative are two separate entities in direct speech. In contrast, the quotative cliticizes to the Final Endings in indirect speech, and is considered a part of the verbal complex. The schematic in (4) demonstrates the difference in organization of the quotative marker between both reported speech types.

(4) Quotative marker in DS vs. IDS

(a) Direct speech



(b) Indirect speech



Where the direct speech is framed by two prosodic boundaries, as reflected in (4a), the results show that the second prosodic boundary is not realized in indirect speech, as reflected in (4b). A more in-depth analysis of these differences is provided in Section 6.

Further support for the claim that prosody is used to indicate RS is provided in the following example from the data, which demonstrates how prosody could disambiguate ambiguous pairs of direct and indirect speech. The quotative marker used for direct speech, *lako*, has the same form as the imperative quotative in IDS, -*lako*. Consider the sentence from participant S3:

(5) Kulayse chinkwu-hanthey wuli cip-eyse ca lako haysse-yo
So friend-DAT our house-LOC sleep QUOT do.PST-POL

(a) kulayse chinkwu-hanthey **wuli** **cip-eyse** **ca** lako haysse-yo
so friend-DAT our house-LOC sleep QUOT do.PST-POL
So, to my friends "Stay at my house" I said.

(b) kulayse chinkwu-hanthey wuli cip-eyse ca-la-ko haysse-yo
so friend-DAT our house-LOC sleep-IMP-QUOT do.PST-POL
So I told my friends to stay at my house.

(c) kulayse chinkwu-hanthey wuli cip-eyse ca-nya-ko haysse-yo
so friend-DAT our house-LOC sleep-INT-QUOT do.PST-POL
So, to my friends "Are you staying at my house?" I said.

The sentence (5a) is present in the data as a direct speech example, and Example (b) is the predicted form of an imperative indirect speech segment. This sentence provides a unique example, as the verb of the direct speech segment, *ca* 'sleep', ended in a vowel and did not have the politeness suffix *-yo*, which resulted in a minimal pair with the imperative. The sentences in (a) and (b) have the same sequence of morphemes following the verb *ca* 'sleep'. However, the interpretation of each sentence is different. The indirect speech in sentence (b) is understood as an imperative due to the *-la* morpheme; it can only be understood as a command. Compare this with (c), 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 direct speech, and unlike the *-la* of the indirect speech's imperative, the *la* of the direct speech quotative is simply there to strengthen the quotative morpheme *-ko*. The forms of (a) and (b) appear ambiguous, but prosody seems to disambiguate them. The sentence in (a) was produced by participant S3, 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 not marked in indirect speech, I would predict that no lengthening would occur on the verb in (b). 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 (b). 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.

6 Match Theory Analysis

The preceding sections assumed theories which were based on strict layering, along with the resulting prosodic structure in Korean (e.g., Jun, 1996, 1998, 2005a, 2005b). 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.

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, MATCHCLAUSE and MATCHPHRASE. I adopt Elfner's (2015) definition of MATCHPHRASE, also applying it to MATCHCLAUSE.

(6) MATCH constraints

(a) MATCHCLAUSE (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) MATCHPHRASE (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, MATCHPHRASE and MATCHCLAUSE 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 a following 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 pause. Thus, φ-phrases should have the characteristics of APs and iPs, and ι-phrases should have the characteristics of IPs. I will first address DS data in Section 6.1, then follow up with IDS data in Section 6.2.

6.1 Direct Speech

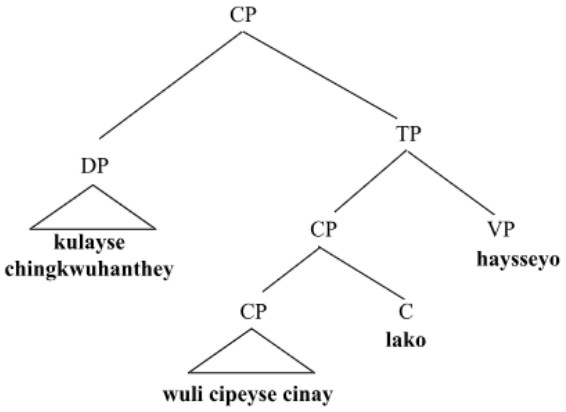
First, consider this example of direct speech from participant S1:

(7) S1 direct speech

	HL			**H**				
kulayse	chinkwu-hanthey	wuli	cip-ey	wa-se	cinay	lako	haysse-yo	
so	friend-DAT	our	house-LOC	come-CONJ	stay	QUOT	do.PST-POL	

So to my friend "Come stay at my house" I said.

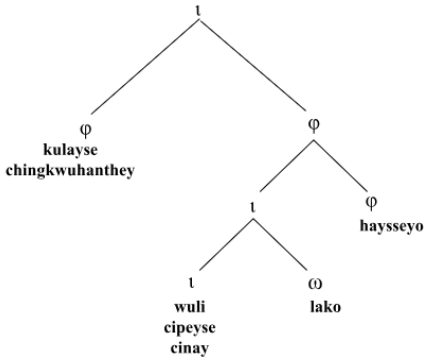
(a) Syntax of S1 direct speech[3]



[3] Note: the example sentence in the syntactic and prosodic trees from (7a) onwards does not exactly match the direct speech sample in Example (7); however, as this study is concerned with prosodic boundaries at the level of the phonological phrase, the breakdown of morphemes is not relevant to the present analysis, which remains unaffected by these minor discrepancies.

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

(8) Prosodic structure of S1 direct speech



There are a few issues with the structure provided in (8). 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 ι-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 (8), 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 ι. 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.

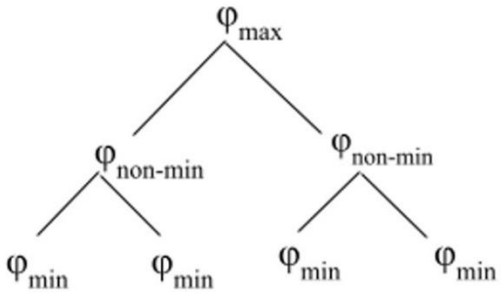
A further potential issue with the structure shown in (8) is the violation of a Match Theory constraint called STRONGSTART.

(9) 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 (8) shows a φ preceding an ι, which is a higher unit, thus violating STRONGSTART.

Another issue with (8) is the amalgamation of APs and iPs as one unit, φ. Not all φ-phrases behave the same in Korean, as we see from the need for an AP and iP in the strict layer structure. Unlike in strict layering, there is no need to add an additional phrase, but simply a distinction between minimal and non-minimal projections of φ. In their Match Theory reanalysis of the iP in Korean, Baek and 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 boundaries 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 (10) is a possible structure of embedded φ.

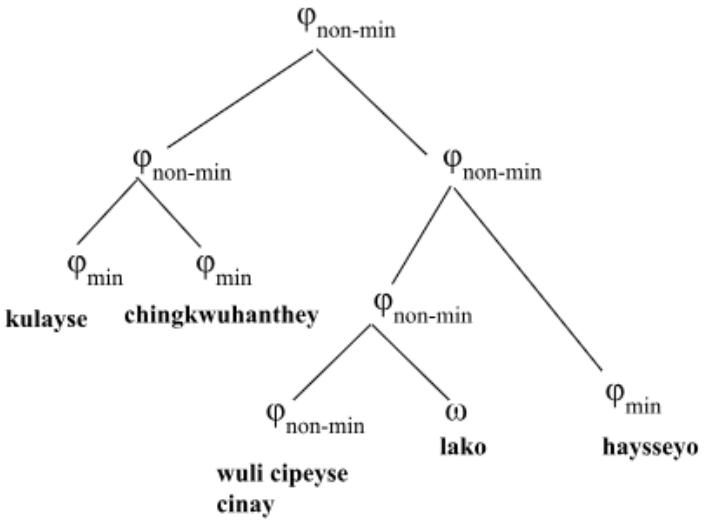
(10)



The minimal and non-minimal projections 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 (8). 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 an 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, 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 $\varphi_{non\text{-}min}$ and φ_{min}, provided in (11).

(11) Alternative structure for S1 direct speech:



In the above example, we have both φ_{min} and $\varphi_{non\text{-}min,}$ and the previously embedded ι-phrase is parsed as $\varphi_{non\text{-}min}$. 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 so as to not violate STRONGSTART. Alternatively, the constraint

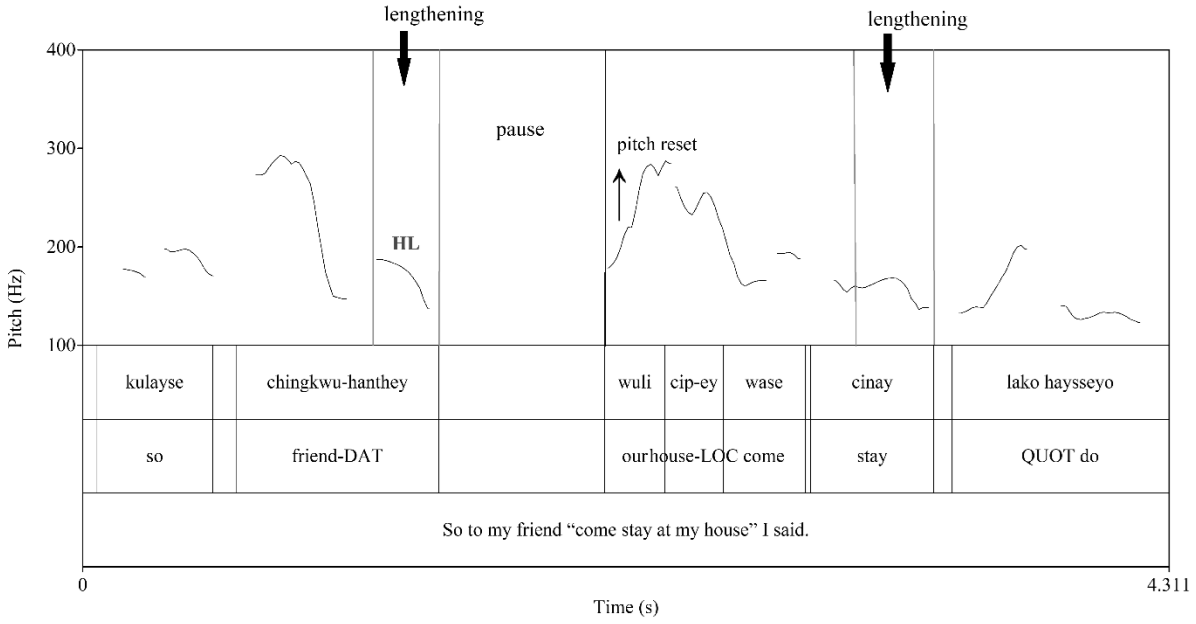
EQUALSISTERS proposed by Myrberg (2013) also accounts for the embedded clause being parsed as a φ-phrase. EQUALSISTERS states that sister nodes in the prosodic structure are of the same prosodic category (Myrberg, 2013). A crucial difference between STRONGSTART and EQUALSISTERS is that STRONGSTART cares about the leftmost constituent. This makes no difference for the current data; therefore, I will assume that STRONGSTART is the relevant constraint for this phenomenon.

Further support for the analysis in (11) comes from Baek and Yun (2018). They found that co-occurring $\varphi_{non\text{-}min}$ edges predicted a stronger boundary than that of a single $\varphi_{non\text{-}min}$ edge. Recall the definitions of strong and weak boundaries in this paper, stated in Section 4. Strong boundaries are marked prosodically with a boundary tone plus another prosodic cue, or a H tone minimally followed by a pause. Weak boundaries have one or no prosodic cues. In Figure 6, we can see where the different prosodic cues occur for this utterance.

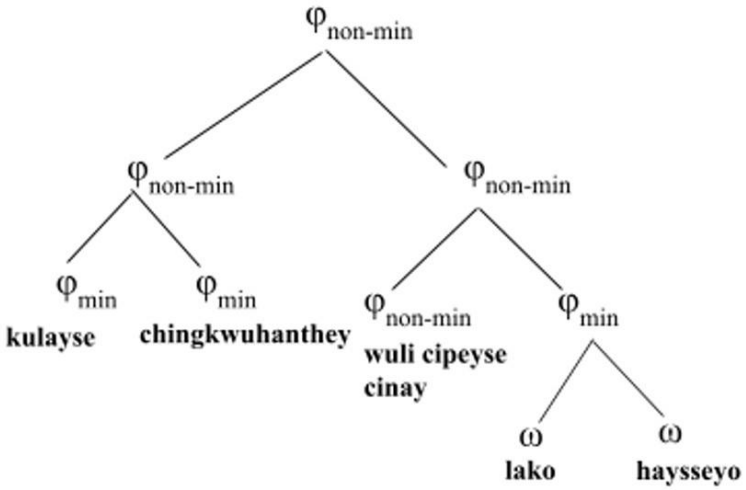
Comparing the structure in (11) to the cues in Figure 6, we can see that a falling boundary tone, pause, and pitch reset are all present where two $\varphi_{non\text{-}min}$ edges co-occur, at the end of the initial frame and at the onset of direct speech. The structure in (11), however, raises an issue with the end frame. There is an intervening $\varphi_{non\text{-}min}$ (containing the $\varphi_{non\text{-}min}$ of the direct speech and the ω of the quotative marker), thus predicting a weak boundary. However, no prosodic cue occurs between the quotative marker and the *verbum dicendi* (with only a single exception by participant S5, which seemed to be anomalous even within the participant's own utterances), which suggests that no boundary exists in this position. Therefore, a revised structure is provided in (12):

Figure 6

S1 Boundary Cues



(12) Revised structure of S1 direct speech



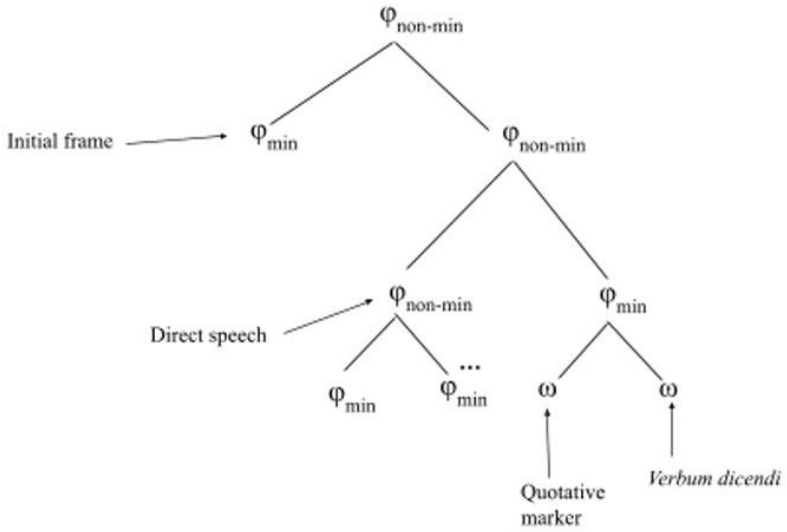
Further, the data shows that the *verbum dicendi* is pronounced together with the quotative marker. Thus, the *verbum dicendi* must be a prosodic word (ω), reflected in the revised structure (12) above. Recalling the constraint EQUALSISTERS, I propose that the *verbum dicendi* is parsed together with the quotative marker as a φ_{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 $\varphi_{non\text{-}min}$ phrases, and weaker boundaries are predicted to occur when there is a single $\varphi_{non\text{-}min}$ edge. The adjacency of the $\varphi_{non\text{-}min}$ edges of the initial frame and direct speech triggers a strong boundary. The edge of the direct speech $\varphi_{non\text{-}min}$ is adjacent to the φ_{min} edge of the quotative marker, thus causing a weak boundary. However, this only accounts for initial frames that contain multiple φ_{min} within a $\varphi_{non\text{-}min}$. Does this analysis hold when the initial frame is φ_{min}? Based on the findings of Baek and Yun (2018), the right edge is more active in Korean, and in the case of a φ_{min} initial frame, the right edge of the φ_{min} is adjacent to the left edge of a $\varphi_{non\text{-}min}$. This suggests that only the φ_{min} boundary tone (i.e., 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 (13).

(13) chinkwu-tul-i kwaynchanhta lako taytap-haysse-yo
friend-PL-NOM okay QUOT reply-do.PST-POL
My friends replied "It's okay."

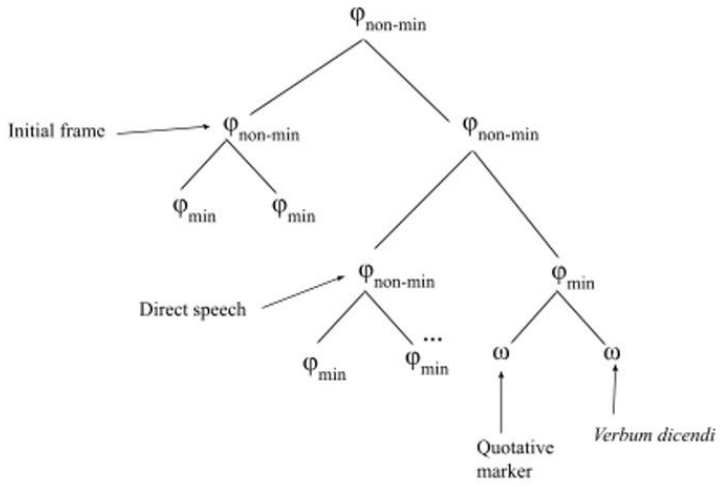
The prosody for all three participants was the same for the initial frame: a H tone at the end of the first word, with no other cues. All three participants treated the end boundary differently: lengthening of the direct speech final segment, a falling boundary tone, or no cues at all, respectively.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

(14) Direct speech structure with φ_{min} initial frame



The structure in (14) demonstrates that no boundary is predicted to occur at the edge of an initial frame with a φ_{min} edge. This is consistent with the utterances produced by the three participants who used the phrase described in (13), and consistent with the analysis of φ-phrase edges in Korean that I have proposed so far. The schematic in (15) shows the structure where the initial frame is $\varphi_{non\text{-}min}$.

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



The structure of (15) reflects the utterance in (12), where the initial frame is a $\varphi_{non\text{-}min}$, with a strong boundary between the initial frame and direct speech due to co-occurring $\varphi_{non\text{-}min}$ edges. The weak boundary between the direct speech and the end frame is found in both structures (14) and (15).

In sum, Match Theory can account for the prosody of direct speech in Korean. The constraint STRONGSTART demotes the embedded clause, which should be an ι-phrase, to a φ-phrase—specifically $\varphi_{non\text{-}min}$. 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 $\varphi_{non\text{-}min}$ results in a strong boundary occurring between them, whereas a single $\varphi_{non\text{-}min}$ edge predicts a weak boundary.

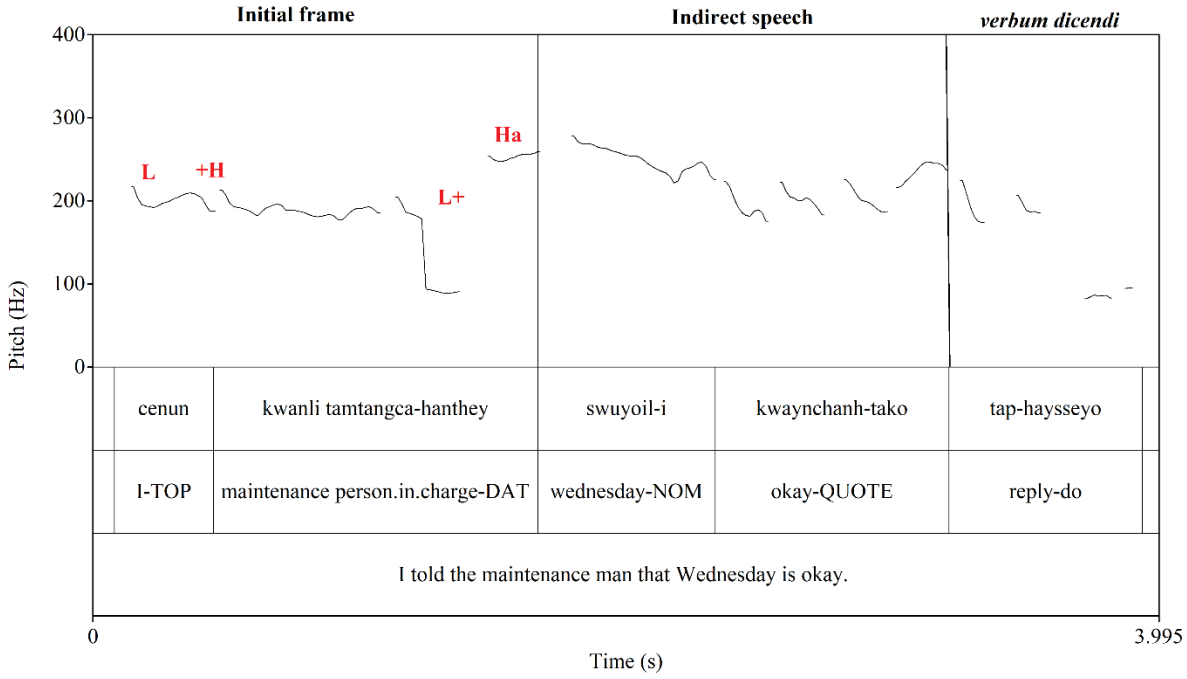
6.2 Indirect Speech

The prosodic boundaries and use of prosodic cues in indirect speech examples can also be explained by the Match Theory analysis. Recall that prosodic cues only occurred at the initial frame boundary in IDS examples. The initial frame of IDS examples is assumed to function as proposed for DS examples—φ-phrase edges predict the boundary strength. The issue with this analysis arises when the initial frame contains more than one word, where each word forms its own φ_{min} that embeds within a $\varphi_{non\text{-}min}$, shown in (16). This means that there are two adjacent $\varphi_{non\text{-}min}$ edges, predicting a strong boundary. Yet in some indirect speech examples, a weak boundary occurs at the edge of the multi-word initial frame, and in others, the predicted strong boundary occurs. This can be explained without modifying the structure proposed so far—multi-word initial frames may be parsed as φ_{min}. Speech rate can affect the size of a phrase, and therefore a multi-word initial frame may be pronounced with one phrasal contour, making the multi-word initial frame a φ_{min}. Figure 7 illustrates the single φ_{min} contour of a multi-word initial frame.

Figure 7 is an example of a multi-word φ_{min}, with a tonal contour of LHLH. The φ_{min} consists of three words: *cenun* 'I', *kwanli* 'maintenance', and *tamtangcahanthey* 'person in 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

Figure 7

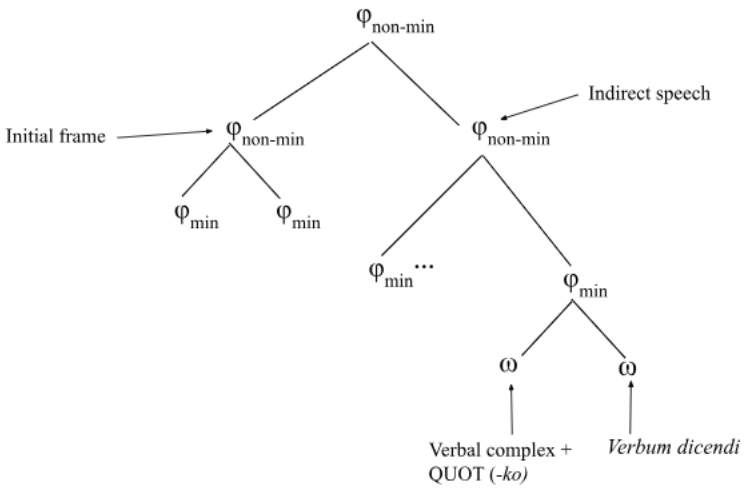
S4 Multi-word φ_{min}



that a slower speech rate would result in multi-word initial frames forming a $\varphi_{\text{non-min}}$, but that in such a case a strong boundary would occur, consistent with the predictions made in this analysis, as seen earlier in Figure 4.

So far, I have demonstrated that the Match Theory analysis of reported speech is applicable to both DS and IDS. However, the lack of a second prosodic boundary in IDS must be addressed. The structure proposed for DS predicts a prosodic boundary after the DS segment, but a prosodic boundary is not observed there in IDS. Thus, the structure must be revised for IDS. The variation comes from the organization of the quotative, which differs in the two reported speech types.

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



Example 16 is the proposed prosodic structure for indirect speech examples. Much like in DS, the *verbum dicendi* must be absorbed into a φ-phrase in order to not violate EQUALSISTERS. In IDS, the quotative *-ko* is not considered a prosodic word until it affixes to the verbal root; this explains the difference in DS and IDS structures. The φ_{min} phrase is formed by the prosodic words (verbal complex + *-ko* and the *verbum dicendi*).

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. The differences in the reported speech types arise from the organization of the quotative marker, where IDS groups it as a prosodic word in combination with the verbal complex, explaining the lack of a secondary prosodic boundary.

6.3 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 $\varphi_{\text{non-min}}$ edges trigger a strong boundary (i.e., more prosodic cues), whereas a singular $\varphi_{\text{non-min}}$ 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 virtually never indicated in indirect speech. The weak end boundary

of direct speech is predicted by a single $\varphi_{\text{non-min}}$ 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.

7 Limitations

As reported speech prosody in Korean is an understudied area, this paper is intended as a preliminary analysis. As such, there were limitations surrounding the methodology. Most of the reported speech prosody literature analysed corpus data (Estellés-Arguedas, 2015; Günthner, 1999; Klewitz & Couper-Kuhlen, 1999, among others); however, due to this paper's focus on a specific structure, part of the purpose of the research was to test the experimental designs in order to determine which, if any, were successful in eliciting the target reported speech. Although the tasks were designed in an attempt to favour declarative reported speech segments, there were still uses of non-declarative reported speech segments (such as hortative or interrogative reported speech examples). The sample size of utterances analysed was greatly reduced due to the exclusion of sentences missing the target structure, reported segments that were not declarative, and sentences including speaker hesitation at boundaries. In addition, pitch reset was not able to be investigated, as there were acoustic confounds: aspirated consonants were not controlled for. Future research may benefit by expanding the scope to include all varieties of reported speech (hortatives, imperatives, etc.), and rather than elicitation of reported speech, an analysis of corpus data may provide better insight into the prosodic cues used by speakers. This study also lacks a control condition. This study only investigated reported speech examples; further research including non-reported complement clauses is needed to determine whether the prosodic cues observed here, attributed to reported speech, pattern differently than the prosody of a non-reported complement clause. In addition, the recordings were done on Zoom, which can affect the duration and intensity measurements in Praat. Further, the durations of pauses and lengthening were not quantified using acoustic thresholds: future research providing measurements of these cues, as well as recordings done in a better environment, should be explored.

8 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 in 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 locations of the boundaries were consistent across speakers, with the onset of reported speech 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 segment 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 Section 6, but some inconsistencies with the boundaries, such as lengthening or pauses seemingly appearing at the edges of Accentual Phrases, led to the proposal of a recursive analysis of reported speech prosody.

In Section 6, 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 led me to propose that the smallest unit with an intonational contour is not an Accentual Phrase, but rather a φ_{min}, and that the strength of prosodic boundaries differs depending on whether the edges of φ-phrases are 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 $\varphi_{non\text{-}min}$ 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.

Acknowledgements

I would like to thank my supervisor, Dr. Emily Elfner, for her invaluable guidance in writing this paper. Thank you for helping me when I was confused and wading in audio data, and pointing me in the direction I needed when I was lost. This research has been submitted and accepted as my Major Research Paper, for partial fulfilment of the Master of Arts in Linguistics at York University, provided here in a condensed version.

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

Boersma, P., & Weenink, D. (2024). *Praat: Doing phonetics by computer* (Version 6.4.19) [Computer program]. https://www.praat.org

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

Buchstaller, I. (2017). Reported speech. In A. Barron, Y. Gu, & G. Steen (Eds.), *The Routledge handbook of pragmatics* (1st ed., pp. 399–416). Routledge. https://doi.org/10.4324/9781315668925-30

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

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

Hirschberg, J., Beňuš, Š., Gravano, A., & Levitan, R. (2020). Prosody in discourse and speaker state. In C. Gussenhhoven & A. Chen (Eds.), *The Oxford handbook of language prosody* (pp. 468–476). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198832232.013.31

Holt, E. (2009). Reported speech. In S. D'hondt, J.-O. Östman, & J. Verschueren (Eds.), *The pragmatics of interaction* (pp. 190–205). John Benjamins Publishing Company. https://doi.org/10.1075/hoph.4.11hol

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* (pp. 280–303). University of Toronto Press.

Jeon, H.-S. (2015). Prosody. In L. Brown & J. Yeon (Eds.,) *The handbook of Korean linguistics* (pp. 41–58). John Wiley & Sons. 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

Kasimir, E. (2008). Prosodic correlates of subclausal quotation marks. *ZAS Papers in Linguistics*, *49*, 67–77. https://doi.org/10.21248/zaspil.49.2008.364

Kim, S. (2012). *Interaction, grammar, and stance in reported speech* [Doctoral dissertation, University of California, Los Angeles]. 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*, 59–79.

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., M., Jr., & 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]. ProQuest Dissertations & Theses. https://www.proquest.com/docview/305339124/abstract/2AA14FE6EEE042D8PQ/1

Pierrehumbert, J. B., & Beckman, M. E. (1988). *Japanese tone structure*. The 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

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. https://doi.org/10.1002/9781444343069.ch14

Yatziv-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* (pp. 117–169). John Benjamins Publishing Company. https://doi.org/10.1075/scl.68.04mal

Yeon, J., & Brown, L. (2019). Quotations. In *Korean: A comprehensive grammar* (2nd ed., pp. 446–468). Routledge. https://doi.org/10.4324/9781315160351-10

Yoon, J. H. S. (2018). Korean syntax. In M. Aronoff (Ed.), *Oxford research encyclopedia of linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.171

Appendix A

Romanization Conventions

Table A1

Consonants

Hangul	Yale	IPA
ㅁ	m	/m/
ㄴ	n	/n/
ㅇ	ng	/ŋ/
ㅂ	p	/p/
ㅃ	pp	/p͈/
ㅍ	ph	/pʰ/
ㄷ	t	/t/
ㄸ	tt	/t͈/
ㅌ	th	/tʰ/
ㄱ	k	/k/
ㄲ	kk	/k͈/
ㅋ	kh	/kʰ/
ㅈ	c	/tɕ/
ㅉ	cc	/t͈ɕ/
ㅊ	ch	/tɕʰ/
ㅅ	s	/s/
ㅆ	ss	/s͈/
ㅎ	h	/h/
ㄹ	l	/ɭ/, /ɾ/

Table A2

Vowels

Hangul	Yale	IPA
ㅏ	a	/a/
ㅐ	ay	/ɛ/
ㅑ	ya	/ja/
ㅒ	yay	/jɛ/
ㅘ	wa	/wa/
ㅙ	way	/wɛ/
ㅓ	e	/ə/
ㅔ	ey	/e/
ㅕ	ye	/jə/
ㅖ	yey	/je/
ㅝ	we	/wə/
ㅞ	wey	/we/
ㅗ	o	/o/
ㅚ	oy	/ø/
ㅛ	yo	/jo/
ㅜ	(w)u	/u/
ㅟ	wi	/y/
ㅠ	yu	/ju/
ㅡ	u	/ɯ/
ㅢ	uy	/ɯi/
ㅣ	i	/i/

Appendix B

K-ToBI Labels (Jun, 2022)

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, and is thus not marked with the IP boundary tone diacritic, %.*