Use of prosodic and lexical cues for disambiguating wh-words in Korean

Jieun Song1, Hae-Sung Jeon2, Jieun Kiaer3

1School of Digital Humanities and Social Sciences, Korea Advanced Institute of Science and Technology, South Korea
2School of Humanities, Language and Global Studies, University of Central Lancashire, UK
3Oriental Institute, University of Oxford, UK

jsong@kaist.ac.kr, hjeon1@uclan.ac.uk, jieun.kiaer@orinst.ox.ac.uk

Abstract

Previous research has shown that the ambiguity of wh-words in Korean can be resolved by prosody. The present study investigated the interplay between prosody and lexical cues in disambiguation. Our written survey results showed that the use of certain adverbs (e.g., a little, once) with a wh-word increases the likelihood of a yes-no question interpretation. The results of our speech production experiment found an interaction of lexical and prosodic cues in the disambiguation. In particular, the presence of a lexical cue affected speakers’ phrasing choice, but not the type of Intonational Phrase (IP) boundary tones or acoustic prominence. The finding supports the proposal that speech production is affected by the amount of linguistic information available for speakers. We further suggest how the phrasing structure could affect speakers’ choice of the IP boundary tone in Korean.

Index Terms: Korean, intonation, semantics-prosody interface

1. Introduction

In Korean, questions with a wh-word are ambiguous between a wh-question and a yes/no polar question as shown in (1).

/at a kajo? ‘where (wh-word) go-ender (verb)’ (1)

a. Where are you going? (wh-question)

b. Are you going somewhere? (yes-no question)

Previous research has suggested that prosody plays a crucial role in disambiguation. Listeners are more likely to interpret an ambiguous question with a wh-word as a yes-no question when the wh-word is made salient by phrasing or focus, or when an H% boundary tone is used [1, 2, 3]. First, Jun and Oh [1] showed that a wh-phrase (i.e., a wh-word followed by a verb) is produced in one Accentual Phrase (AP) in a wh-question, but in two APs in a yes-no question, with the AP boundary between the wh-word and the following verb. The accentual phrasing seems to be an important cue for disambiguating wh-phrases in perception [3, 4].

Second, speakers may use different Intonational Phrase (IP) boundary tones for disambiguation (Fig. 1). H% was the most commonly used for yes-no questions, LH% for wh-questions [1, 5].

Third, wh-words tend to be acoustically more prominent (with greater f0 and amplitude) or narrowly focused in wh-questions compared to in yes-no questions [1, 6]. Prosodic boundaries after a focused item can be deleted in Korean [7, 8]; the narrow focus in the wh-question may cause the deletion of the AP boundary following the wh-word, i.e., speakers produce the wh-word and the following verb as one AP. However, the disambiguating effect of pitch prominence seemed minor compared to that of accentual phrasing when they were independently manipulated in perception studies [3, 9]. The disambiguating role of prominence in perception has not been tested in terms of amplitude.

Figure 1: Schematic representations of the prosodic differences of a yes-no question (left) versus a wh-question (right; e.g., /at a kajo/?) [1]

Furthermore, there is evidence that there are other linguistic cues for disambiguation; for instance, for Kyeongsang Korean, listeners use sentence-ending interrogative particles (e.g., -na, -no) for the disambiguation [10, 11]. However, to our knowledge, the role of lexical cues has not been investigated. In fact, wh-words are often accompanied by a quantifying/degree adverb or adverbial expressions such as /komi/ ‘a little’, /hanpye/ ‘once’, and /kamo/ ‘a moment’. We refer to them as “existential adverbs” in that they presuppose the existence of an event, yet to its vaguely minimum extent. Although the existential adverbs can be used for both wh-questions and yes-no questions, our intuitions as native Korean speakers were that with the existential adverb(s), the wh-phrase is more likely to be interpreted as a yes-no question, asking whether the described event is true or not.

To examine whether the presence of adverbial expressions increases the likelihood for the yes-no question interpretation without prosodic cues, first, we carried out a paper-based survey. Then we carried out a speech production experiment to investigate whether the presence of adverbial expressions affects the way speakers use prosody for disambiguation. In the experiment, speakers were recorded reading questions containing a wh-word and an adverbial expression (existential vs. others) when the meaning of the question could be determined in the dialogue (i.e., wh- or yes-no question). We analyzed accentual phrasing, IP-final boundary tones, and prominence of the wh-word. It is highly plausible that speakers utilize multiple sources of information (e.g., lexical, prosodic, and contextual) to convey the intended meaning of a sentence, and the multiple cues can interact with one another [12, 13]. For example, speakers are less likely to place phrasal stress on a word when they provide a concurrent semantic cue in the utterance or discourse context [12, 13].

First, we expected that speakers may not use accentual phrasing if the adverb could play the disambiguating role. Second, we hypothesized that the presence of the adverbial expression would suppress the speakers’ need to use different
boundary tones for different interpretations. Lastly, the presence of the adverbial expression may affect the prominence of the wh-word, reducing its functional load in disambiguation.

2. Methods

2.1. Survey

2.1.1 Participants

Sixty-five native speakers of Korean completed a questionnaire. They were university students in Seoul, aged between 22 and 24 with no self-reported language impairments.

2.1.2. Stimuli & Procedure

Forty ambiguous questions with a wh-word were presented in the Korean script (e.g., / ammon tawmk*an tɔmolsuis*ajọ? “Can I see you for a moment?”/ “When can I see you for moment?”). Each question was preceded by a brief description of interlocutors (e.g., “a professor says to her/his student”) which helped the participants interpret the question in a communicative context. Twenty sentences had an existential adverb between the wh-word and the verb, as shown in Table 1, while the other 20 sentences did not. In the questionnaire, the pragmatic marker /tɔm/ ‘a little’ indicating politeness or softening [14] was combined with another adverbial expression (e.g., /tawmk*an tɔm/) as often used in naturally produced sentences. There were fillers which were questions without a wh-word. Participants were instructed to write down an appropriate answer for each question.

2.2. Speech production experiment

2.2.1. Participants

Four native speakers of Seoul Korean (all female) took part in the production experiment. They had no self-reported language or hearing problems and were aged between 22 and 33 (mean: 29). They had lived in English-speaking countries for 32.6 months on average (range: 6 to 107 months) as an adult.

2.2.2. Stimuli

Seventy-two target sentences with a wh-word (“who”, “when”, “where”, “what”; /nuka/, /stāri/, /stī/ or /stīs/, /mawaka/) were created (4 wh-words × 3 adverb types × 3 adverbs × 2 sentences). The sentences were classified into three categories: 1) with existential adverbs, 2) with other adverbs (each sentence had one of the 6 adverbs shown in Table 1), or 3) without adverbs. Other adverbs were frequency adverbs that presuppose more frequent occurrences of an event; they were not expected to trigger a bias in the wh-word interpretation. For instance, /stī tɔm kajo? ‘where (wh-word) a little (existential adverb) go-ender (verb)’ is usually interpreted as ‘do you go somewhere at all?’ whereas, /stī tɔŋkɔŋ kajo? ‘where (wh-word) often (other adverb) go-ender (verb)’ would be interpreted as either ‘where do you often go?’ or ‘do you often go somewhere?’. The ‘other adverbs’ condition was used in the speech production experiment to compare utterances with an equal number of syllables (because it affects phrasing) and identical syntactic structure. Adding an adverb increases the utterance length; as Seoul Korean speakers tend to keep an AP to have five or fewer syllables, the presence of the adverb can lead speakers to adjust their phrasing strategies [15]. The final verb always consisted of three syllables and did not begin with a strong consonant which triggers AP-initial high pitch (see [16]).

A dialogue with a question-answer pair was created to be presented on the computer screen (e.g., Q with either A1 or A2; see (2) with the wh-word, adverb and verb underlined). Speakers could disambiguate each question based on the answer; the answer was yes or no (A1), or it contained the specific information asked (A2, e.g., a place, a thing, a person, a time).

Q. /samusile nuka tawmk*an is*ajọ?
   ‘office-locative who-nominaive for a moment be-end’
   Will someone be in the office for a moment? (with A1)
   Who will be in office for a moment? (with A2)

A1. /ne kulAkajjo/
   Yes. (Someone) will be.

A2. /kimpisaka is*ulkajjo/
   Secretary Kim will be.

Table 1: The type of adverbs in the sentence

<table>
<thead>
<tr>
<th>Type</th>
<th>Adverb</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Existential</td>
<td>/tɔm/</td>
<td>a little</td>
</tr>
<tr>
<td></td>
<td>/hanpan/</td>
<td>once</td>
</tr>
<tr>
<td></td>
<td>/tawmk*an/</td>
<td>a moment</td>
</tr>
<tr>
<td></td>
<td>/tɔŋkɔŋ/</td>
<td>often</td>
</tr>
<tr>
<td></td>
<td>/hanŋaj/</td>
<td>always</td>
</tr>
<tr>
<td></td>
<td>/tawtu/</td>
<td>often</td>
</tr>
</tbody>
</table>

2.2.3. Procedure

To ensure that the speaker understood the intended meaning of the question (i.e., wh vs. yes/no Q), they were first asked to read the question and the answer presented on the screen before the recording started.

For recording, each participant read the question from the list and, the experimenter who was a native speaker of Seoul Korean read the answer. This procedure was repeated twice in order to ensure that speakers understood the intended meaning of the question, and the second production was used for analysis. The presentation order of the dialogues was randomized for each speaker. The recording took place in a sound-proof booth in University College London, London, using a RODE NT1-A microphone, with a sample rate of 44100Hz and a 16-bit quantization rate.

2.2.4. Data annotation and measurements

First, word boundaries were automatically marked using a forced aligner [16] and then hand-corrected. The first and second authors then auditorily and visually examined each utterance and annotated AP boundaries, IP boundaries, and IP-boundary tones, using Praat. The annotation criteria followed the K-ToBI labelling conventions [18]. For differentiating H%, LH%, and HL%, the annotators examined the timing of the f0 rise; for H%, the pitch starts to rise in the penultimate syllable, whereas for LH%, the rise starts in the final syllable [18]. The inter-annotator agreement rate was 79.5 % for the presence of an AP boundary between the wh-word and the adverb and 81.5% for that between the adverb and the final verb. The inter-annotator agreement rate for IP boundary tones was 85.4%. The two annotators reexamined all the utterances with disagreement together to reach an agreement. Prominence of the wh-word was measured by the maximum intensity and f0 using a Praat script.
3. Results

3.1. Survey

The majority of participants opted for the wh-question interpretation without the existential adverb. As Fig. 2 shows, the most common response type was a wh-answer (i.e., providing specific information on who, when, where and what), followed by a yes-no answer and a “both” answer (i.e., saying ‘yes’ or ‘no’ and providing specific information on who, when, where and what) for this condition; the percentage of occurrence was 71.29%, 17.09%, and 11.62%, respectively.

However, the interpretation bias changed when the existential adverb was present: the most common response type was a yes-no interpretation (Fig. 2, wh-question interpretation, 29.46%; yes-no question, 48.44%). This change was significant in a simple regression analysis in which the ratio of the yes-no to the wh-interpreter was the dependent variable, and adverb type (existential adverbs vs. no adverbs) was an independent variable, F(1, 32) = 4.96, p = 0.03, R² = 0.13, b = -9.61. In addition, the number of “both” answers also increased in the presence of the existential adverb (Fig. 2, 22.10 %) compared to when there was no adverb. A simple regression analysis found that the effect of the presence of the adverb on the ratio of the ‘both’ answer to the wh-question interpretation did not reach the significant level (p = 0.086).

![Figure 2: Percentage of occurrence of responses by answer type and adverb type in the questionnaire](image)

The survey results confirmed that there is a bias towards interpreting a wh-phrase towards a yes-no question in the presence of the existential adverb. That is, the presence of the existential adverb overall reduces the likelihood for the ‘wh-question’ interpretation that is otherwise high.

3.2. Speech production experiment

Speakers placed an AP boundary before the final verb more frequently when there was an adverb (both ‘existential’ and ‘other’ adverb conditions) compared to the no-adverb condition (Fig. 3). This is probably because they were reluctant to produce the longer wh-phrase with an adverb without starting a new AP before the verb. To compare accentual phrasing in the two adverb conditions (i.e., existential vs. other), a logistic mixed-effects regression analysis was performed with the presence of an AP boundary as the dependent variable, and adverb type (i.e., existential vs. other) and intended interpretation (i.e., yes-no vs. wh) as independent variables. Speaker and item (the six adverbs shown in Table 1) were included in the model as random effects. Because an AP boundary could be placed before and after the adverb, separate models were fitted for the AP boundary between the wh-word and the adverb (mod 1) and that between the adverb and the verb (mod 2). Lme4 [19] was used in R for statistical analyses. The significance of each factor was calculated by comparing two models with and without the relevant factor (α = 0.05).

There was a main effect of the intended interpretation in both models [mod 1: $\chi^2(1) = 27.42$, p < 0.001; mod 2: $\chi^2(1) = 66.19$, p < 0.001], indicating that speakers were more likely to produce the AP boundary between the wh-word and the adverb and between the adverb and the verb for a yes-no question. The main effect of adverb type was not significant in both models. However, the intended interpretation × adverb type interaction effect was significant for the AP boundary between the wh-word and the adverb, $\chi^2(1) = 4.05$, p = 0.04 (mod 1). Specifically, for the yes-no question interpretation, the AP boundary was less likely to be placed for the existential adverb condition than the other adverb condition (upper left graph of Fig. 3). That is, speakers were less likely to mark a yes-no question by accentual phrasing when there was a lexical item potentially indicating a yes-no question. On the other hand, for the wh-question interpretation, the presence of the existential adverb did not reduce the frequency of the AP boundary occurrences (the same graph of Fig. 3). The adverb type × intended interpretation interaction was not significant for the boundary between the adverb and the verb, p = 0.19 (mod 2).

![Figure 3: Bar plots with error bars showing ±1 standard deviation for the proportion of AP boundary occurrence by adverb type and answer type (top) and those for the no-adverb condition by answer type (bottom)](image)

For IP boundary tones, Fig. 4 shows that H% was the most commonly used for both yes-no and wh-interpretations (all HL% tones were produced by one speaker). This is inconsistent with the finding of the previous study [1] which examined 4 speakers, showing that LH% is a dominant type of boundary tone used for wh-questions. In order to investigate the effects of adverb type and the intended interpretation, a logistic mixed-effects regression analysis was performed with the IP boundary tone (H% vs. LH%) as the dependent variable and adverb type (existential vs. others) and the intended interpretation (yes-no vs. wh) as independent variables. Speaker was included in the model as a random effect. The main effect of the intended interpretation was significant, $\chi^2(1) = 4.61$, p = 0.03; in fact, H% was used more frequently with a wh-answer than with a
yes-no answer (mean percentage of occurrence: wh, 90%; yes-no, 84%). The main effect of adverb type was marginally significant, $\chi^2(2) = 5.63$, $p = 0.0598$. The interaction of the intended interpretation and adverb type was not significant, $p = 0.64$. By conducting planned comparisons in R\(^1\), we examined the effect of adverb type further; the results showed that the difference between the existential and other adverb condition (contrast 1) was not significant, $p = 0.34$, but the difference between the two adverb conditions and the no-adverb condition (contrast 2) was significant, $b = 1.01, z = 2.171, p = 0.03$. That is, there was a tendency for speakers to use LH\% more frequently in the no-adverb condition than in the two adverb conditions.

---

**Figure 4: Bar plots with error bars showing ±1 standard deviation for the proportion of occurrence of IP boundary tones by adverb type**

For analyzing pitch and intensity of the wh-word, questions with /hanp\/ ‘once’ and /han\/ ‘always’ were excluded because the word-initial /h/ would trigger an AP-initial H tone unlike in other sentences. Mixed-effects regression analyses were conducted with the adverb type and the intended interpretation as independent variables and speaker and adverb as random effects. The dependent variable was the maximum f0 (mod 1) or the maximum intensity (mod 2) in the wh-word. Both the f0 and intensity of the wh-word were larger in wh-questions than in yes-no questions [mod1: $\chi^2(1) = 36.62$, $p < 0.001$; mean (standard deviation) of maximum f0 in Hertz – Wh: 227(62.7), yes-no: 204(44.4); mod2: $\chi^2(1) = 10.79$, $p = 0.001$; mean (standard deviation) of maximum intensity in dB SPL, Wh: 71.1(4.16), yes-no: 70.2(4.53)]. This finding is consistent with the accentual phrasing pattern that speakers were less likely to place an AP boundary after the wh-word for a wh-question; dephrasing is probably more likely to occur after the prominent wh-word in wh-question interpretations [7, 8]. On the other hand, no effects related to adverb type were found, $p > 0.05$.

4. Discussion & Conclusions

The present study investigated how the ambiguity of wh-words in Korean can be resolved by a combination of prosodic and lexical cues. The main finding was that there was an interaction between lexical cues and prosody; the presence of a lexical cue seems to reduce the need for using accentual phrasing for disambiguation. To summarize the results, first, without prosodic cues, existential adverbs (e.g., /kom/ ‘a little’ and /hanp\/ ‘once’) shifted the interpretation of an ambiguous question from a wh-question to a yes-no question. Then in speech production, speakers would produce an AP boundary after the wh-word without an existential adverb (i.e., without any adverb or with a non-existential adverb), but its presence seemed to reduce the need to produce the AP boundary. Furthermore, the results suggest that phrasing is what Korean speakers manipulate in relation to the lexical cues rather than acoustic prominence on the wh-word; the wh-words were often produced with higher salience in f0 and intensity or narrow focus in wh-questions compared to yes-no questions, but the type of the adverb had no effect. Similar ideas were proposed in the smooth signal redundancy hypothesis [12, 13] which suggests that there is an inverse relationship between language redundancy (e.g., predictability based on lexical, semantic, syntactic, or discourse information) and prosodic redundancy. For example, words or syllables tend to be more phonetically reduced or produced in fewer phrases when they are more predictable in the context [13]. Similarly, speakers may be more likely to produce AP boundaries for the yes-no question interpretation when the intended meaning is less predictable in the absence of lexical cues.

Second, the finding has implications in understanding how speakers choose IP boundary tones. In previous studies, the reason for listeners’ preference for the LH\% tone in wh-questions and for the H\% tone in yes-no questions was not clearly explained. Our results showed that H\% was used dominantly when any adverb was present regardless of the intended interpretation. This may show that the choice of the boundary tone is affected by speakers’ intended phrasing structure. That is, in previous studies, speakers may have produced LH\% for the wh-questions, because the wh-phrase (e.g., a wh-word followed by a verb) formed an IP-final AP as a whole. Korean speakers have a strong tendency to produce an AP-final rise [see 16]; for instance, in /\ati\koj\a/ (yes-no question, with the AP boundary marked by #), speakers may be under pressure to produce the AP-final rise for each AP, causing strict alignment between the four tones and the four syllables (see Fig. 1), leading to H\% with the rise beginning in the penultimate syllable. On the other hand, in /\ati\koj\a/ (wh-question), the AP length probably allows some flexibility in aligning the second H (H+ in Fig. 1), potentially delaying the onset of the AP/IP-final rise to the final syllable. Then, for our sentences with an adverb, speakers were more likely to produce a short IP-final AP, and this would have led speakers to produce H\% rather than LH\%. As the present study examined only four speakers, our proposal is tentative. However, the potential relationship between the length of the IP-final AP and speakers’ IP boundary tone choice deserves further investigation.

All in all, the findings support the proposal that speech production is affected by the amount of linguistic information available for speakers. In addition, we suggest a possibility that speakers’ choice of the IP boundary tones may be affected not only by their pragmatic intention but also by their intended phrasing structure of the utterance.

---

\(^1\) The weights for contrast 1 were 0 (no adverb), +1 (existential), and -1 (other). The weights for contrast 2 were -2 (no adverb), +1 (existential), and +1 (other).
5. References


