Wh-intonation and Information Structure in South Kyongsang Korean and Tokyo Japanese
Hyun Kyung Hwang

1. Wh-intonation and the long distance-scrambling puzzle in Tokyo Japanese

This study is an investigation into the relationship between prosody and the semantic scope of wh-phrases in South Kyeongsang Korean (henceforth SKK), spoken in southeastern Korea; Fukuoka Japanese (henceforth FJ), spoken in southwestern Japan; and Tokyo Japanese (henceforth TJ). Its chief innovation is to take information structure into serious account. Unlike standard Korean, SKK is known as a pitch accent language in which the pitch accents are lexically determined, as in Japanese.

TJ has been the most extensively studied among the languages at least in impressionistic terms. However, it is clear that closer experimental investigation is called for, as contradictory arguments are reached, and experimental evidence is generally lacking in the literature. Further, information structure has not been adequately controlled for in previous work. In this paper, in addition to TJ, I would like to explore two more languages; FJ and SKK, to which little attention has been paid in the prosodic literature. FJ and SKK are instructive cases since the two languages exhibit the same correspondence as TJ between the semantic scope of wh-phrases and prosody. Moreover, the intonation pattern triggered by a wh-phrase (henceforth wh-intonation) is particularly unusual in these varieties in that lexical pitch accents are completely lost in the semantic scope of a wh-phrase. These characteristics make it possible that wh-intonation in FJ and SKK is distinguished from $f_0$ boost or compression of a word triggered by other discourse associated factors. In investigating these two languages, I not only seek to identify the correlation between information structure and syntactic structure, and phonetic realization, but also wish to understand what is intrinsic and what is variable in the correlation.

Tokyo Japanese has been described phonetically as exhibiting an $f_0$ boost in a wh-phrase and $f_0$ reduction in the following material. Wh-intonation in TJ where wh-phrases take different scope is shown below. Wh-phrases and their semantic scope are represented by boldface and shading, respectively.

(1) Schematic pitch contours of wh-interrogatives in TJ

\[ [ \text{wh} \ Comp_{[+WH]} ] Q_{[-WH]} ] [ \text{wh} \ Comp_{[-WH]} ] \]

As shown in (1), the domain of $f_0$ reduction correlates with the semantic scope of the wh-phrases. As will be discussed in more detail below, $f_0$ boost or reduction can also be

---

1 The pattern is called Focus Intonation by Ishihara (2004, 2005) and Emphatic Prosody by Kitagawa and Deguchi (2002). The motivation for this terminology is discussed in §3.

2 In the schema, boundary H tones are omitted. Since boundary tones are not of direct concern in this paper, they will not be represented throughout.
triggered by contrastive focus (henceforth FOCUS) or givenness. Yet, there is a debate regarding the \textit{wh}-intonation domain of indirect \textit{wh}-questions in \textit{wh}-long distance scrambling (henceforth LD-scrambling) constructions.\footnote{These LD-scrambling constructions are judged as somewhat degraded (Saito 1989, Watanabe 1992, Ishihara 2004).} As the LD-scrambled \textit{wh}-phrase is placed out of its scope, constructions with \textit{wh}-LD-scrambling are illuminating cases for obtaining a better understanding of the scope-prosody correlation. The structure and opposing claims are given in (2).

(2) Contradicting claims regarding \textit{wh}-intonation in LD-scrambling construction


\begin{tikzpicture}[every node/.style={font=\footnotesize}, level distance=1.5cm, level 1/.style={sibling distance=3cm}]
  \node {nani-o \textcolor{red}{\text{Naoya-wa [Mari-ga nonda-ka]}} \text{imademo oboeteru-ø}?
    \text{what-Acc Naoya-Top Mari-Nom drank-Comp_{[+WH]} even.now remember-Q_{[-WH]}}}
  \node (a) at (0,-1) {nani-o \textcolor{red}{\text{Naoya-wa [Mari-ga nonda-ka]}} \text{imademo oboeteru ø?}};
  \node (b) at (1,-1) {'Does Naoya still remember what Mari drank?'};
\end{tikzpicture}


As shown in (2a), Ishihara (2002) and Kitagawa and Deguchi (2002) observe that $f_0$ compression terminates with the embedded Comp, based on limited phonetic descriptions. This result was not, however, tested by a systematic acoustic study. On the other hand, Ishihara (2004, 2005) proposes a Multiple Spell-out Model (henceforth MSO) which is a cyclic computation grounded on the concept of phase in the Minimalist framework (Chomsky 2000, 2001). MSO predicts that $f_0$ compression continues to the end of the matrix clause as in (2b) with the subordinated scope interpretation. According to Ishihara (2004, 2005), the experimental results he obtained support this interpretation. The claim is important in two respects: first, it was based on systematic instrumental study. Second, if the results are correct, the scope-intonation correlation is no longer observed in \textit{wh}-LD scrambling cases. That is the reason that Ishihara refers to this as the ‘mismatch’ case. However, Ishihara acknowledges that the intonation pattern in (2a) also seems possible, and in that case, some additional mechanism is required since MSO would not allow the contour in (2a) (Ishihara 2004, p113, footnote 27).

Notice that the \textit{wh}-intonation pattern in TJ is extremely similar to the intonation pattern incurred by FOCUS. This is reflected in the terminologies for the pattern in previous studies—Focus Intonation (Ishihara 2004, 2005) or Emphatic Prosody (Kitagawa and Deguchi 2002). In this paper, however, the prosodic manifestation of other information context is shown to be distinct. Also, it is argued that the domain of \textit{wh}-intonation in the construction can be identified by taking information structure into account. Further, the nature of the prosodic scope marking can be better understood when we explore the expanded list of languages. This approach ultimately provides a good testing ground for the syntax-prosody mapping across languages.

The experimental results in the present study show that the scope-prosody correlation is indeed maintained even in the LD-scrambling constructions, as claimed by Ishihara (2002) and Kitagawa and Deguchi (2002). However, the intonation pattern
in (2b) is also possible when the post-subordinate clause material is discourse-given. In order to motivate the new approach adopted here, I review the previous studies on the \textit{wh}-scope marking in LD–scrambling constructions and their limitations in §2, before I introduce the prosodic scope marking in SKK and FJ in §3. The methodology and the results of the current study are presented in §4, and the theoretical implications are discussed in §5. The study is concluded in §6.

2. Limitations of previous studies
Not much attention has been paid to the domain of \textit{wh}-intonation in the construction with LD-scrambling. As noted above, Ishihara (2002) and Kitagawa and Deguchi (2002) claimed that the domain of \textit{wh}-intonation stops with the embedded Comp based on impressionistic observation. Although the observation was not tested systematically, it is worthwhile to examine the model Kitagawa (2005) developed, which is developed from Kitagawa and Deguchi (2002). To account for the scope-prosody correlation, they introduce a syntactic analysis relying on the operation Agree (Chomsky 2000). The features which Kitagawa assumes undergo Agree are given in (3).

(3) Features on \textit{wh}- and focused phrases (Kitagawa 2005)
\begin{itemize}
\item[a.] \textit{WH}-P – Focus \textit{Wh}-phrase
\item[b.] \textit{Wh}-P – Non-focus \textit{Wh}-phrase
\item[c.] \textit{FP} – Non-\textit{wh focus} phrase (both presentational and contrastive)
\item[d.] \textit{Wh} – Reference to \textit{Wh-in} general as in “\textit{Wh}-questions, \textit{Wh}-phrase, \textit{Wh-in-situ}”
\end{itemize}

The scope of \textit{WH}-Ps, \textit{Wh}-Ps and \textit{FPs} is determined when each of them is associated with a specific formal feature of Comp, as summarized in (4).

(4) Features on Comp (Kitagawa 2005)
\begin{itemize}
\item[a.] COMP [\textit{wh}]: COMP with a \textit{wh}-feature (= an interrogative feature) is \textit{unselectively} associated with one or more \textit{wh}-Ps.
\item[b.] COMP [\textit{F}]: COMP with an \textit{F}-feature (= an emphatic feature) is associated with an \textit{FP}.
\item[c.] COMP [\textit{WH}]: COMP with a \textit{WH}-feature (= an interrogative emphatic feature)
\end{itemize}

While \textit{WH}-P appears on \textit{wh}-phrases which exhibit \textit{wh}-intonation, \textit{Wh}-P appears on \textit{wh}-phrases nested inside another \textit{wh}-intonation domain. Recall that $f_0$ of post-\textit{wh} material is compressed in TJ. Thus, \textit{Wh}-P is not prosodically prominent. However, it is not clear what the motivations are for the three different ‘\textit{wh}’ features – \textit{WH}-P, \textit{Wh}-P and \textit{Wh}, and whether the ‘focus’ in \textit{WH}-P and \textit{Wh}-P involve both presentational and contrastive focus as \textit{FP} does. Note that what is discussed as ‘focus’ in the literature is in fact contrastive focus, and a \textit{wh}-phrase is not always contrastively focused. Further, it appears that Kitagawa does not differentiate \textit{wh}-intonation from focus intonation; this is necessary in that the prosody of \textit{wh}-phrases and non-\textit{wh} focused phrases is distinct in SKK and FJ. This point is further discussed in §3. The same limitation is observed in Ishihara’s work (2004, 2005).

Ishihara (2004, 2005) proposes a model based on cyclic computation of prosody for \textit{wh}-scope marking. The idea behind this proposal is that \textit{wh}-scope marking is not a
result of the direct phonology-semantics interaction but triggered by phase-by-phase computation of syntax (Ishihara 2004). The fundamental insight of this account is that wh-intonation is assigned from a wh-phrase to the right edge of some Spell-Out domain. After a phase is constructed, the complement of the phrase head is available to the phonology via Spell-Out. Let us consider phases and the approach to Spell-Out as adopted by Ishihara from Chomsky (2001a, b).

(5) Phrase and Multiple Spell-Out (Chomsky 2001a,b)

a. CPs and vPs are phases.

b. When a syntactic derivation reaches a phase (CP/vP) in the syntax, the complement of the phase head (TP/VP) is transferred to the interface levels. The phonological part of the Transfer is called Spell-Out.

\[
\begin{array}{c}
\text{[CP (Spec) C [TP (Spec) T [vP (Spec) v [VP . . . ]]]]}
\end{array}
\]

phase Spell-Out phase Spell-Out

Two focus features are assumed: one on a Q-particle (FOC_Q), the other on a wh-phrase (FOC_WH). After the two features establish an Agree relation, FOC_WH enters into the interface level. At each Spell-Out domain, FOC_WH found at the interface level triggers creation of wh-intonation, ending at the right edge of the Spell-Out domain. Note that Comp itself is outside of the Spell-Out domain as TP, not CP, undergoes Spell-Out. In order to include Comp in the wh-intonation domain, Ishihara suggests that it is phonologically cliticized to the preceding phrase.

Let us now turn to the prediction of MSO for the LD-scrambling constructions. The relevant simplified tree structure is displayed below. Successive scrambling of the wh-phrase and detailed structure are omitted for the sake of brevity.

(6) Prediction of MSO for LD-scrambling constructions
Ishihara assumes that the *wh*-phrase moves cyclically (Spec vP1 → Spec CP1 → Spec vP2 → Spec CP2). As a result, at each cycle of Spell-Out, the *wh*-phrase escapes from the domain by cyclic movement. Thus *wh*-intonation is not created even at the CP2 phase. To assign *wh*-intonation in this construction, MSO assumes an additional process, called Root Spell-Out. It requires that the entire matrix CP (CP2) be sent to Spell-Out at the last stage of the derivation. As a result, in the case of LD-scrambling, the entire clause including the *wh*-phrase is sent to Spell-Out, and *wh*-intonation is created. Yet, the motivation for Root Spell-Out is not clear.

Further pursuing experimental evidence for his model, Ishihara argues that it receives experimental support. Ishihara (2004, 2005) tests four types of sentences whose structures are shown below.

(7) Configurations of four tested sentence types
a) [Sub_m [Sub_e DP V_e-Comp] target V_m]
   b) [Sub_m [Sub_e wh V_e-Comp] target V_m]
   c) [DP Sub_m [Sub_e tDP V_e-Comp] target V_m]
   d) [wh Sub_m [Sub_e twh V_e-Comp] target V_m]

(7a, b) are declaratives with canonical order, where a DP or a *wh*-phrase is the object of the embedded clause. The embedded object – DP or *wh* – is LD-scrambled in (7c,d), respectively. The target phrase is immediately following the embedded clause. What Ishihara measures are the peak $f_0$ s of the embedded verb ($V_e$) and the target phrase. Based on the statistically significant difference between the peak $f_0$s of DP-scrambled and *wh*-scrambled (c and d in 7), Ishihara concludes that *wh*-intonation in (7d) continues to the matrix material in *wh*-LD-scrambling constructions.

The results are instrumentally derived, yet the between speaker differences cast doubt on Ishihara’s claim. Among four speakers, only one speaker (YY) exhibited significantly lower $f_0$ of the target phrase compared to the other three. For speaker CS and KS, $f_0$ difference between DP and *wh*-scrambled types was not significant. Speaker AH exhibited quite interesting results; the target phrase in *wh*-scrambled sentences yields lower $f_0$ compared to DP-scrambled cases, but that in *wh* in-situ were also lower than DP embedded ones.

Perhaps more importantly, Ishihara does not take into account the possible effects of contextual information even though it is extremely crucial, particularly in TJ as $f_0$ boost or compression may be incurred by discourse information such as FOCUS or givenness (Sugahara 2003, Kubozono 2007). However, when it comes to the distinction between *wh*-intonation and FOCUS intonation, none of the previous literature has differentiated between them.

Finally, this particular MSO approach faces the limitation that it cannot be extended to SKK and FJ, which exhibit similar prosodic scope marking in *wh*-interrogatives. This problem is explored in detail in what follows.

3. *Wh*-intonation in SKK and FJ
The intonation pattern displayed by FOCUS in SKK and FJ is compatible with that in TJ; $f_0$ boost on FOCUSed material and post-FOCUS $f_0$ compression. The contrast

---

4 Although Ishihara reports only the averages of normalized data which make validating his claim difficult, he kindly provided me his actual data upon my request.
between pitch contours with and without FOCUS is confirmed by the SKK declarative sentence in (8).

(8) Minho-nun [Ywumi-ka manul-ul mekessun-ci] mwul-ess-ta
Minho-Top Ywumi-Nom garlic-Acc ate-Comp[+WH] ask-Past-Dec
‘Minho asked whether Yumi ate garlic.’

The overlapped pitch contours with and without FOCUS assigned on the embedded DP manul-ul ‘garlic-Acc’ are presented in Figure 1.

![Figure 1. Pitch contours of a declarative with and without FOCUS in SKK (uttered by SKK speaker F2\(^5\))](image)

The vertical dotted lines mark phrasal boundaries. As shown in the overlapped pitch contours, the presence of FOCUS on the embedded object triggers \(f_0\) boost of the phrase and post-FOCAL compression.

As for \(wh\)-intonation, on the other hand, FJ and SKK exhibit a distinctive intonational pattern for \(wh\)-scope marking which is clearly distinguishable from FOCUS marking. Interestingly, a prosodic constituent is formed from a \(wh\)-phrase to the complementizer that binds the \(wh\)-element; it consists of a rise during the \(wh\)-phrase followed by a high flat interval with a fall at the end (Gim 1978, Kubo 1993). This is different from TJ in that all elements inside the constituent completely lose their lexical pitch accents (Kubo 1993; Smith, to appear). Lexical pitch accent loss was instrumentally tested in Hwang (2006) for SKK, and in Smith (2007) for FJ. The \(wh\)-intonation pattern of FJ and SKK corresponding to (1) in TJ are schematically shown in (9).

\(^5\) All pitch contours in the current paper were made using Praat, and speakers’ gender is indicated by F (female) or M (male).
Despite the fact that FJ and SKK are varieties of different languages, they exhibit exactly the same pitch pattern, distinguished by a high flat $f_0$ contour assigned to the scope of a $wh$-phrase. In comparison SKK and FJ with the TJ case, notice that the three languages have the property of scope-prosody correspondence in common. In FJ and SKK, however, $wh$-intonation and $f_0$ reset/reduction are clearly distinguishable. The examples above show that FOCUS intonation and $wh$-intonation are independent effects in SKK and FJ indicating that we need to tease the two apart. For this reason, I use the terminology $wh$-intonation rather than FOCUS intonation for the prosodic pattern of scope marking. More importantly, the existence of distinct $wh$- and FOCUS intonations gives us a tool to separate out the effects of $wh$-scope marking and information structure, particularly in contexts where the $wh$-phrase is scrambled or $wh$-interrogatives are structurally distinct.

Recall that MSO predicts that $wh$-intonation terminates with matrix Comp while the LD-scrambled $wh$-phrase takes embedded scope. If the prediction of MSO for the LD-scrambling construction applied correctly to these varieties as well, there would be high plateau all the way to the end of a matrix clause independent of the embedded semantic scope in SKK and FJ. However it has been shown that this is not the case in both SKK and FJ (Kubo 2005, Hwang 2006), indicating that a new approach is necessary which can fully account for the data incorporating the relevant languages. In the following section, I test the LD-scrambled constructions in various discourse contexts in the languages.

4. Experiment

4.1 Methods

4.1.1 Materials
In order to examine the domain of $wh$-intonation when $wh$-LD-scrambling has occurred, two object types (DP or $wh$-phrase) and two object positions (embedded or LD-scrambled) were tested. This was done to replicate and extend the data of Ishihara (2004, 2005) by putting the same types of sentences in different contexts. The structures of the four types of target sentences are repeated in (10).

$\begin{align*}
(10) & \text{a) } [CP1 \ [CP2 \ DP \ Comp] \ I.O. \ Dec] \\
& \text{b) } [CP1 \ [CP2 \ wh \ Comp] \ I.O. \ Dec] \\
& \text{c) } [CP1 \ DP \ [CP2 \ t_{DP} \ Comp] \ I.O. \ Dec] \\
& \text{d) } [CP1 \ wh \ [CP2 \ t_{wh} \ Comp] \ I.O. \ Dec]
\end{align*}$

All four types are affirmative sentences embedding either yes/no or $wh$-questions. (10a) and (10b) are sentences with canonical word order (DP/$wh$ embedded). DP and $wh$-
phrase are LD-scrambled in (10c) and (10d), respectively (DP/wh scrambled). The target phrases which are indirect objects (I.O.) immediately following embedded clauses are indicated by underlining. Unlike the material tested by Ishihara (2004, 2005) and Hwang (2006), adverb *imademo* ‘still’ was not used because of its focal meaning. Three different sets were tested where the embedded verbs were varied: ‘eat’, ‘read’ and ‘make’ for SKK, ‘drink’, ‘read’ and ‘knit’ for Japanese. These verbs were selected in consideration of segmental context and word frequency. Only accented words were used, as different behavior between accented and unaccented words has been reported (Kubozono 1989, Sugahara 2003 among others).

Each sentence was preceded by a question in a dialogue of different information structure. One set of dialogues where a *wh*-phrase is embedded in the target sentence in SKK is listed in (11).

(11) Example dialogues with varying context

a) FOCUS

Experimenter: *Minho-nun way kulekey Ywumi-hantey kwansim-i manh-no?*

Minho-Top why so Yumi-Dat interest-Nom plenty-Q

Ywumi-ka me-lul ilkessnun-ci Ywumi unni-hantey mwule-pwassta-

mye?

Yumi-Nom what-Acc read-Comp Yumi sister-Dat ask-tried-Q

‘Why is Minho so interested in Yumi?

He asked Yumi’s sister what Yumi read, right?’

Speaker: *uuung, Ywumi unni-ka aniko…*

No, Yumi sister-Nom is not

*Minho-nun Ywumi-ka me-lul ilkessnun-ci*

Minho-Top Yumi-Nom what-Acc read-Comp

*Yengwu-hantey mwule-pwassta-nta*

Yeongwu-Dat ask-tried-Quot

‘No, it’s not Yumi’s sister.

I heard that Minho asked Yengwu what Yumi read.’

b) New

Experimenter: *Minho-nun me-lul hassta-ko?*

Minho-Top what did-Quot

‘You said Minho did what?’

Speaker: *Minho-nun Ywumi-ka me-lul ilkessnun-ci*

Minho-Top Yumi-Nom what-Acc read-Comp

*Yengwu-hantey mwule-pwassta-nta*

Yeongwu-Dat ask-tried-Quot

‘I heard that Minho asked Yengwu what Yumi read.’

c) Given

Experimenter: *Minho-nun Yeongwu-hantey me-lul mwule-pwassta-no?*

Minho-Top Yeongwu-Dat what ask-tried-Q

‘Do you know what Minho asked Yeongwu?’

Speaker: *Minho-nun Ywumi-ka me-lul ilkessnun-ci*

Minho-Top Yumi-Nom what-Acc read-Comp

*Yengwu-hantey mwule-pwassta-nta*

Yeongwu-Dat ask-tried-Quot

‘I heard that Minho asked Yengwu what Yumi read.’
Note that the target sentence which is uttered by a speaker is identical across the contexts. Varying discourse contexts provided by the experimenter’s questions results in the assignment of distinct information structural properties on the target phrase. In the FOCUS context given in (11a), as the experimenter asks a question with an incorrect I.O., FOCUS is expected to be assigned on the I.O. in correcting the information. While the I.O. is discourse-new in (11b), it is given information in (11c).

4.1.2 Participants and recording

Three female and three male speakers of each language participated in the recording. They ranged in age from 20 to 33 years at the time of recording. All were born and grew up in the respective linguistic target area, and had no history of speech or hearing impairment.

Recordings were made in a sound-attenuated booth at Cornell University for TJ, at Smyrna church in the city of Changwon for SKK, and at Kyushu University in the city of Fukuoka for FJ. A portable Marantz digital recorder (PMD 660) and a SHURE SM 57 microphone were used for the recordings.

I briefed the speakers on the procedure of the recording. Since the target materials were embedded in a dialogue, they were informed that I would ask a question, and they would answer the question by reading the given materials written in Korean or Japanese orthography. The experimenter’s questions (i.e., contexts) were not included in the written scripts, so that they had to pay attention to the questions uttered by the experimenter. In providing context by asking the questions, efforts were made for consistent production. The speakers were instructed to give natural renditions at a comfortable speed. Practice time was given in order to eliminate unnaturalness. When the speakers misread the materials, or inserted unnatural pauses, they were asked to repeat the sentence before proceeding to the next. A total of four repetitions were elicited from each speaker, and a short break was given between the repetitions. Recording session for each speaker lasted approximately one hour including practice time and break. All speakers were paid for their participation at the end of the recording session.6

4.1.3 Measurements and analyses

The data were digitized with a 22,050 Hz sampling rate and 16-bit quantization. Labeling and measurements were made using Praat version 5.0.03. Only three renditions out of four repetitions were analyzed. In most cases, the second, third and fourth repetitions were used unless their quality was substantially low by disfluency, unnaturalness or background noise. When an outlier \( f_0 \) value was present due to consonantal perturbation in the later repetitions, the first repetition was used.

Phrase boundaries were manually marked on each utterance. In measuring fundamental frequencies, both peak and valley \( f_0 \) values of each phrase were extracted in order to discern the best reference point. The labeling and measurements are illustrated in Figure 2.

---

6 The data collection was funded by a Research Travel Grant from the Cornell University Graduate School.
The token shown in Figure 2 is a *wh*-embedded type uttered in the new context by a male speaker of SKK. Recall that ‘new’ refers to a context where the *wh*-question has not been asked in prior discourse. Although there are some bumps, the high flat pitch pattern is observed from the end of the embedded *wh*-phrase to the syllable preceding the Comp. While the $f_0$ peak of the target phrase is the primary concern of this study, pitch patterns of the material preceding the target were also examined. Since pitch range of even a single speaker can vary from token to token, comparing only the absolute $f_0$ peak values without considering a reference point can be misleading. Yet, it is a non-trivial problem to factor out the pitch range variation. Recall that unjustified normalization may distort the data.

To circumvent the possibility of distortion by normalization, the reference point was carefully chosen. From the peak and valley $f_0$ values of the preceding material, maximum $f_0$, minimum $f_0$ and $f_0$ range of the preceding material were obtained. In the preceding material, maximum $f_0$s were detected in the embedded subject for the constructions with canonical word order. For DP-scrambled constructions, the $f_0$ peaks of the scrambled DPs were the maximum $f_0$s. Minimum $f_0$ values were observed at the end of the embedded Comp for all the utterances for TJ. For SKK, some were found at the beginning (L tone) of the embedded *wh*-phrase. For the *wh*-embedded constructions in SKK, L tone of the *wh*-phrases and the comp were extremely similar, only 2-3 Hz difference. Notice that the two $f_0$s are quite similar in Figure 2.

In order to determine the most reliable reference point, the values were compared among different contexts, sentence types, and speakers. Since no speaker variation was found, the data were pooled across speakers; the means and standard deviations (in parentheses) are given in Table 1.
Table 1. Maximum $f_0$ (Max $f_0$), minimum $f_0$ (Min $f_0$) and $f_0$ range in SKK and TJ

<table>
<thead>
<tr>
<th></th>
<th>Max $f_0$</th>
<th>Min $f_0$</th>
<th>$f_0$ Range</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SKK</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F_{wh\ emb}$</td>
<td>190.6 (66.7)</td>
<td>124.9 (43.9)</td>
<td>65.7 (29.6)</td>
</tr>
<tr>
<td>$F_{wh\ scrmb}$</td>
<td>191.2 (68.9)</td>
<td>127.2 (45.1)</td>
<td>63.9 (27.8)</td>
</tr>
<tr>
<td>$F_{DP\ emb}$</td>
<td>192.2 (67.1)</td>
<td>123.6 (42.2)</td>
<td>68.6 (31.1)</td>
</tr>
<tr>
<td>$F_{DP\ scrmb}$</td>
<td>191.0 (67.9)</td>
<td>122.2 (42.1)</td>
<td>68.8 (32.2)</td>
</tr>
<tr>
<td><strong>MEAN</strong></td>
<td>191.3 (67.2)</td>
<td>124.5 (43.1)</td>
<td>66.8 (30.1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>TJ</strong></th>
<th>Max $f_0$</th>
<th>Min $f_0$</th>
<th>$f_0$ Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_{wh\ emb}$</td>
<td>239.7 (78.8)</td>
<td>117.9 (37.4)</td>
<td>121.8 (51.2)</td>
</tr>
<tr>
<td>$F_{wh\ scrmb}$</td>
<td>264.6 (86.8)</td>
<td>115.8 (39.1)</td>
<td>148.8 (56.3)</td>
</tr>
<tr>
<td>$F_{DP\ emb}$</td>
<td>242.8 (79.5)</td>
<td>118.8 (39.3)</td>
<td>124.0 (48.0)</td>
</tr>
<tr>
<td>$F_{DP\ scrmb}$</td>
<td>250.0 (81.1)</td>
<td>117.4 (39.9)</td>
<td>132.6 (50.0)</td>
</tr>
<tr>
<td><strong>MEAN</strong></td>
<td>249.3 (81.6)</td>
<td>117.5 (38.7)</td>
<td>131.8 (52.2)</td>
</tr>
</tbody>
</table>

As shown in Table 1, minimum $f_0$ values remain fairly constant across the different conditions and languages, indicating that this is the best reference point, as influences of the conditions are directly reflected by the $f_0$ values of the target with reference to minimum $f_0$, regardless of the change of $f_0$ range. In order to confirm statistical significance, one-way ANOVAs using generalized linear model were performed.\(^7\) The dependent variable considered was minimum $f_0$ and the independent factors were context, object type (DP or $wh$), object position (embedded or scrambled) and interaction of the object type and position. There were no significant differences between minimum $f_0$s in different contexts (F(2, 642)=0.7702, p=0.6804 for SKK, F(2, 642)=0.8140, p=0.6656 for TJ), Object type (F(1, 642)=2.0140, p=0.1559 for SKK, F(1, 642)=0.0011, p=0.9730 for TJ), object position (F(1, 642)=0.2154, p=0.6425 for SKK, F(1, 642)=0.0422, p=0.8372 for TJ), and interaction of object type and position (F(1, 642)=0.7702, p=0.6804 for SKK, F(2, 642)=0.8140, p=0.6656 for TJ), Object type (F(1, 642)=2.0140, p=0.1559 for SKK, F(1, 642)=0.0011, p=0.9730 for TJ), object position (F(1, 642)=0.2154, p=0.6425 for SKK, F(1, 642)=0.0422, p=0.8372 for TJ), and interaction of object type and position (F(1,

\(^7\) Statistical analyses were carried out using JMP version 7.0.
Thus, the ratio of minimum $f_0$ over $f_0$ peak of the target phrase was calculated for each token. The results of the experiment are presented in the following section.

It should be noted that loss or change of $wh$-intonation is observed among younger speakers of FJ. The FJ speakers in the current experiment, all undergraduates at Kyushu University, exhibited anomalies with respect to the prosodic pattern. One female speaker (FJF3) does not have the high plateau: she consistently produced a $wh$-phrase as H*L accent resulting in pitch contours similar to the TJ pattern. Four speakers uttered a $wh$-phrase as unaccented, yet no accent deletion was followed indicating the lack of $wh$-intonation. Only one male speaker (FJM3) shows the high flat contour for the $wh$ LD-scrambling sentences from the second repetition suggesting that proficiency is necessary for producing the pattern. A pitch contour for the $wh$ LD-scrambling context uttered by the speaker is illustrated in Figure 3.

As shown above, a high flat pitch contour with declination is observed from the scrambled $wh$-phrase in the sentence initial position to the end of the embedded clause (Ve). While the lexical accent of the material between the $wh$-phrase and the embedded verb was deleted, that of the embedded verb was not, exhibiting a pitch fall on the accented mora. It is noteworthy that Kubo (1996) describes the intonation pattern of indirect $wh$-questions as bearing a penultimate accent when the Comp is –$ka$. It suggests that large number of speakers including older generations should be tested as even this speaker’s high plateau is not exactly comparable to what has been reported. Thus, only SKK and FJ data are further analyzed and discussed hereinafter.

4.2. Results
4.2.1 $wh$-intonation in LD-scrambling constructions
Before turning to target $f_0$ with reference to the minimum $f_0$, we need to first confirm that there is $wh$-intonation in $wh$-LD scrambled constructions. While this fact is obvious in SKK since $wh$-intonation is realized as a high plateau, in TJ a careful comparison is required to examine whether there is $f_0$ compression in $wh$-LD scrambled constructions compared to the DP-scrambled counterpart. The intonation patterns of a $wh$-scrambled and a DP-scrambled construction in SKK are shown in Figure 4 below.
Both utterances in Figure 4 were produced in the FOCUS context. However, regardless of the varying context, *wh*-scrambled constructions consistently exhibited *wh*-intonation, while ordinary implementation of pitch accents was observed in constructions where non-*wh*-DPs were scrambled. Observe that *wh*-intonation is shown from the scrambled *wh*-phrase to the end of the embedded clause. Thus, by measuring $f_0$s of the scrambled *wh*-phrase and the embedded verb, it can be shown that the high plateau is assigned from the *wh*-phrase to the Comp where it takes scope for SKK. Recall the configurations of the LD-scrambling constructions, which are repeated in (12).

(12) a) $[\textbf{wh} \quad [\text{Sub}_m [\text{Sub}_e \quad t_{wh} \quad \text{V}_e-\text{Comp}] \quad \text{Target} \quad \text{V}_m-Q]$

b) $[\text{DP} \quad [\text{Sub}_m [\text{Sub}_e \quad t_{DP} \quad \text{V}_c-\text{Comp}] \quad \text{Target} \quad \text{V}_m-Q]$

Since all speakers exhibited the same pattern, the peak $f_0$ values of the scrambled phrases and the embedded verbs were pooled; the means are given in Table 2.
Table 2. $f_0$ peak of the Scrambled Phrase (SP) and the embedded verb (Ve) in SKK

<table>
<thead>
<tr>
<th></th>
<th>Scrambled Phrase (SP)</th>
<th>Ve</th>
<th>Ve/SP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$wh$</td>
<td>213.95</td>
<td>212.28</td>
<td>1.00</td>
</tr>
<tr>
<td>DP</td>
<td>237.71</td>
<td>182.68</td>
<td>0.79</td>
</tr>
</tbody>
</table>

As presented in Table 2, there were found no signs of declination or downstep between $f_0$ peaks of the scrambled $wh$-phrases and the embedded verbs while $f_0$ peaks of the embedded verbs in DP-scrambled constructions exhibited the expected declination. The pitch patterns of the LD-scrambled sentences are schematically shown in Figure 5 below.

![Figure 5. Schematic pitch contours of LD-scrambled sentences in SKK](image)

Table 3. $f_0$ peak of the embedded subject (Sube) and the minimum $f_0$ (Min $f_0$) in TJ

<table>
<thead>
<tr>
<th></th>
<th>Sube</th>
<th>Min $f_0$</th>
<th>Sube/Min $f_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$wh$</td>
<td>192.49</td>
<td>117.98</td>
<td>1.63</td>
</tr>
<tr>
<td>DP</td>
<td>204.63</td>
<td>117.63</td>
<td>1.74</td>
</tr>
</tbody>
</table>

The pitch contours in Figure 5 indicate that $wh$-intonation is present not only in $wh$-in situ but also in $wh$-LD-scrambling constructions.

In TJ, however, the presence of $wh$-intonation cannot be explored in the same way. Recall that $wh$-intonation is marked by $f_0$ compression in TJ. Thus, a peak $f_0$ of a certain point between the scrambled phrases and the Comp should be compared with reference to the minimum $f_0$. The peak $f_0$ of the scrambled $wh$-phrase is not adequate for comparison since the $f_0$ peak of $wh$-phrases show a peak $f_0$ boost. Also, the topic (the subject of the matrix clause) or the embedded verb phrases yield somewhat reduced $f_0$ values even in DP cases. Therefore, the $f_0$ peaks of the embedded subjects were compared. As motivated before, a reference point –minimum peak $f_0$– is necessary for more accurate comparison. The peak $f_0$s of the embedded subjects with reference to the minimum $f_0$ and the schematic pitch contours are given in Table 3 and Figure 5, respectively.
As shown in Table 3 and Figure 6 above, across speakers and the contexts, $f_0$ peaks of the embedded subjects after *wh*-phrases were compressed compared to those after *DP*-phrases, indicating that *wh*-intonation is observed in *wh*-LD-scrambled constructions. Let us now turn to the results of the target phrase as a function of the context.

4.2.2 Relative $f_0$ of the target depending on information structure
The peak $f_0$s of the target phrase with reference to the reference point were pooled across speakers as all the speakers of each language exhibited the same pattern. The mean ratios of the minimum $f_0$ (i.e. reference) over the target $f_0$ peak are shown in Figure 7, separated by language.
There are appreciable differences in the target $f_0$s and the ratios among the contexts, yet, there are no appreciable differences among the four different sentence types in both languages. Somewhat higher ratio values for the $wh$-scrambled cases in SKK can be attributed to the greater minimum $f_0$ rather than the lower target $f_0$. Recall that the pitch realization of the high plateau is quite high, and the duration of the Comp (-ci) is short when falling tone is assigned. This triggers the undershot of L tone when followed by a H initial word. For the Given context, the undershot effect is weak since the $f_0$ excursion size of the target phrase following is small.

Overall, across the speakers and the languages, the New cases exhibit more variation in $f_0$ peaks of the target. This is because quite a lot of tokens overlap with the values in the Given cases; sometimes speakers accommodate the discourse-new indirect object as given probably because of the relatively small importance of the phrase in terms of conveying information.
In order to test the significance of the differences statistically, one-way ANOVAs using a generalized linear model were performed. The dependent variable considered was the ratio of minimum $f_0$ over target $f_0$. Independent factors were context, object type (DP or $wh$), object position (embedded or scrambled) and interaction of the object type and position.

In SKK, as expected from Figure 7, there is a significant effect of context on the ratio ($F(2, 642)=540.0635, p<0.0001$). The post-hoc comparisons indicate that all three contexts are differentiated (FOCUS $>$ New $>$ Given). On the other hand, there was no significant difference in the ratios depending on object type ($F(1, 642)=0.3424, p=0.5585$) or position ($F(1, 642)=1.9619, p=0.1613$). The interactions of object type and position was not significant ($F(1,642)=0.4829, p=0.4671$).

In TJ, one-way ANOVAs show that there is a significant effect of context on the ratios of minimum $f_0$ over target $f_0$ ($F(2,642)=638.1283, p<0.0001$), yielding significant differences in the ratios among the three contexts (FOCUS $>$ New $>$ Given). It is confirmed that object type or position does not play a significant role for the ratios in TJ ($F(1,642)=0.0244, p=0.9156$ for type; $F(1,642)=0.0112, p=0.9156$ for position). Also, the interaction between object type and position was not significant ($F(1, 642)=0.2301, p=0.6314$).

5. Discussion
5.1 Resolving the puzzle
The experimental results of the current study show that the right edge of the $wh$-intonation (i.e. high plateau) in SKK consistently aligns with the embedded Comp regardless of the position of the $wh$-phrase (in situ or LD-scrambled). Also, there was no significant effect of the object type (DP or $wh$-phrase) or the object position (embedded or scrambled) on the $f_0$ excursion size of the target phrase following the embedded Comp. Instead, there was found an immediate effect of discourse-context on the $f_0$ of the post-$wh$-intonation material. These prosodic patterns of SKK in the varying contexts enable us to resolve the puzzle with $wh$-LD scrabbling constructions in TJ. The two claims concerning the issue are repeated below.

(13) The puzzle regarding the domain of $wh$-intonation in the LD-scrambling construction

\[
\text{nani-o Naoya-wa [Mari-ga nonda-ka] imademo oboeteru-ø?}
\]
what-Acc Naoya-Top Mari-Nom drank-Comp[+WH] even.now remember-Q[-WH]


\[
\text{nani-o Naoya-wa [Mari-ga nonda-ka] imademo oboeteru ø?}
\]
‘Does Naoya still remember what Mari drank?’

Recall that in the experiment reported in this study, $f_0$ of the material following the scrambled $wh$-phrase is consistently compressed up to the embedded Comp in TJ. Yet, $f_0$ of the material following the Comp is either prominent or reduced depending on its information status. The results suggest that in parallel with SKK, position of the object (embedded or scrambled) or type of the object ($wh$-phrase or DP) does not trigger pitch
reduction or boost of the post-Comp material as in (13a). Thus, it is premature to conclude that the results of Ishihara’s (2004, 2005) experiment show that the wh-intonation domain includes matrix material extending to the end of the matrix clause in wh-LD scrambled sentences with scrambled wh-phrases taking embedded scope.

The remaining question to be considered here is to account for the disparity between the experimental results reported in this paper and prosodic pattern in (13b) reported by Ishihara (2004, 2005). Recall, first, that discourse context was not controlled for in the latter study; this is particularly critical in TJ, since, as we have seen, both discourse-givenness and wh-intonation are realized by f0 reduction in this variety. It is important to keep in mind here that the lack of explicit prior discourse is not necessarily equivalent to a neutral context, especially in wh-question environments (Crain & Steedman 1985, Altmann & Steedman 1988). Thus failure to control for discourse context may be at work in the production of the pattern in (13b).

Recall that Ishihara argued that f0 of the material immediately following the Comp is compressed in a wh-scrambled sentence compared to in a DP-scrambled one. Although this difference is reported as statistically significant, it is difficult to assess the linguistic significance of this result without access to the actual data. It is possible that even a small difference between the two means (wh and DP) can be statistically significant if the standard error of the difference is small as the statistical significance in a T-test is determined by calculating t-ratio (difference between the means / standard error of the difference). The reported statistical significance may simply be a byproduct of normalization. Also, LD-scrambling does not change meaning in either scrambled wh- or DP-scrambled sentences.

However, f0 of the target phrase can be compressed both in wh- and DP-scrambling constructions. There are two possible information structures in which the intonation pattern in (13b) are observed. First, as in the Given context in this paper, the post-Comp material is discourse-given, and prosodically reduced, as in (14).

\[
\text{(14) reduction by wh-} \quad \text{reduction by Givenness} \\
\text{nani-o Naoya-wa [Mari-ga nonda-ka] imademo oboeteru-\text{\(\hat{o}\)}}
\]

Another possibility is that FOCUS intonation is implemented on the entire clause by the scrambled phrase (15). It is certainly conceivable that the speakers arbitrarily assigned FOCUS on the scrambled element as FOCUS is sometimes claimed to motivate LD-scrambling (Miyagawa 2001).

\[
\text{(15) reduction by FOCUS on wh-} \\
\text{nani-o Naoya-wa [Mari-ga nonda-ka] imademo oboeteru-\text{\(\hat{o}\)}}
\]

Note that in both cases, the pitch reduction on the material following the embedded clause is triggered by the information status –specifically, discourse-givenness or FOCUS– as an effect independent from wh-intonation. Therefore, the domain of wh-intonation in wh-LD-constructions terminates with the embedded Comp as in (15a), but the intonation pattern in (15b) can be observed in different information structure.

Then, how do we account for the correlation of wh-scope and prosody in formal syntax? The question is discussed in the following section.
5.2 Syntactic account
Two models have been proposed regarding this phenomenon in the literature – MSO (Ishihara 2004, 2005) and a model based on Agree (Kitagawa and Deguchi 2002, Kitagawa 2005). The instrumental data of this study show that the prediction of MSO is not empirically felicitous, and as Ishihara (2004) acknowledged, it is not possible to produce the actually observed intonation pattern in the model. The alternative that I consider to be more promising is the ‘Agree model’. Yet, recall that the invalid assumptions are reflected in the feature set: wh-intonation is ‘focus’ intonation, and a wh-phrase which triggers wh-intonation is prosodically prominent. The feature set in Kitagawa (2005) is repeated below.

(16) Features on wh- and focused phrases (Kitagawa 2005)
   a. \textit{WH-P} – Focus Wh-phrase
   b. \textit{Wh-P} – Non-focus Wh-phrase
   c. \textit{FP} – Non-wh focus phrase (both presentational and contrastive)
   d. \textit{Wh} – Reference to Wh-in general as in “Wh-questions, Wh-phrase, Wh-in-situ”

The assumption cannot be extended to SKK or FJ since wh-intonation is independent from ‘focus’ and wh-phrases are not prosodically prominent in those languages. Further, it is not clear what kind of ‘focus’ is intended in the classification for wh-phrases. Thus, in modifying the model, I propose two features in (17).

(17) a. [wh] – it is inherently assigned on wh-phrases
   b. [FOCUS] – it can be assigned on both wh- and non-wh items by information structure.

Following Kitagawa and Deguchi (2002), I assume the following.

(18) Assumptions
   a. an uninterpretable [wh] feature is assigned on Comp, and interpretable [wh] feature on a wh-element.
   b. an uninterpretable [wh] feature on Comp is deleted through Agree.
   c. The Agreement is interpreted as a semantic scope of a wh-element at LF, and wh-intonation is assigned at PF.

Kitagawa and Deguchi (2002) do not erase the feature at PF without any discussion of why and how it could remain intact. I assume that the uninterpretable [wh] feature is deleted at PF as well, following Chomsky (2001) that every uninterpretable feature should be deleted before entering interface level.

While the phenomenon in direct wh-questions is straightforwardly captured in wh-scrambled structures, another assumption is required for indirect wh-questions: at PF, a [wh] feature pair under Agreement is ‘linearly scanned’ (Kitagawa and Deguchi, 2002). Recall that the best generalization from the data is that wh-intonation is assigned from a wh-phrase to the Comp linearly.
(19) Indirect wh-question with LD-scrambling

\[
\text{embedded wh-scope}
\]

a. LF: [CP nani[\textbf{wh}]\textbf{-}o Naoya\text{-}wa [CP Mary\text{-}ga nani[\textbf{wh}]\textbf{-}o nonda\text{-}ka[\textbf{wh}]]
\text{imademoboeteru}]

b. PF: [CP nani[\textbf{wh}]\textbf{-}o Naoya\text{-}wa [CP Mary\text{-}ga nonda\text{-}ka[\textbf{wh}]]
\text{imademoboeteru}]

As shown in (19), the upper copy, which is outside the domain of Agree, is deleted at LF. On the other hand, the original copy establishes Agreement with the embedded Comp and wh-embedded scope interpretation is obtained. At PF, linear-scanning applies here and the wh-intonation is assigned on the material between the feature pair. This apparatus ‘linear scanning’ is required to assign wh-intonation on the unusual domain. It is noteworthy that the wh-intonation domain in this structure cannot be derived by a purely syntactic mechanism Agree because of the matrix subject nested by the material of the embedded clause, and the linear nature of prosody.

6. Conclusion
In this paper, I have examined intonation patterns of wh-interrogatives in order to resolve the puzzle regarding the domain of wh-intonation in LD-scrambling constructions in SKK, FJ and TJ by taking information structure into account. A phonetic experiment was conducted in which different contexts were provided. The results show that when a wh-element is scrambled out of the embedded clause, the domain of wh-intonation is not the main clause but the embedded clause including the matrix subject, indicating that the earlier proposal (Ishihara 2004, 2005) cannot capture the full set of data. The current findings not only resolve the empirical issue, but also show that in understanding the nature of the syntax-phonology interface, it is crucial to distinguish wh-intonation from intonation conditioned by information structural properties (FOCUS vs. New vs. Given), and wh-intonation and FOCUS intonation should be distinguished.

References
Gim, Cha-gyun (1978), The role of Tones in sentences. Eoneohag 3, 61-79.


