

Article

Prosodic Manifestation of Syntactic Boundaries in Japanese

Takayuki Konishi

Abstract

This study investigated how prosody influences syntactic processing in Japanese. Kitagawa (2005) claims that the inconsistencies in grammatical judgements reported in previous studies can be explained by taking into consideration what he calls “Emphatic Prosody.” However, it is not clear from his analysis why it is *Emphatic Prosody* that influences syntactic processing. This study refutes his analysis with empirical data and claims instead that it is prosodic boundaries that elicit different syntactic processing. The analysis using the *Tones and Break Indices* transcription system suggests it is Intonation Phrase Boundaries that act as syntactic boundaries.

1. Introduction

In the field of syntax, there sometimes exists disagreement in grammaticality judgements of complex sentences even among native speakers. Watanabe (1992), while arguing that the Wh-phrase internal to Wh-island cannot take matrix scope, admits varying “degrees of unacceptability” among different speakers (p 262). (1) replicates the sentence used for the discussion.

- 1) (?)-??John-wa [Mary-ga nani-o katta-kadouka]
-Top -Nom what-Acc bought-CompWthr
shiritagatte-iru-no?
want.to.know-CompWh
-> ‘What₁ does John want to know [whether Mary bought *t*₁]?’

(Watanabe, 1992; 256-257, 263)

In his paper (Kitagawa 2005), Kitagawa¹ claims that prosody must be taken into

consideration when a grammaticality judgement is made. According to him, (1) is acceptable when it is interpreted with prosodic boundaries as in (1a), where double-slash (//) denotes a prosodic boundary; on the other hand, for (1b), the sentence is unacceptable².

1a) John-wa [Mary-ga // **Nani-o** katta-kadouka] shiritagatte-iru-no?
 what-Acc -Comp_{Whr} -Comp_{Wh}
 -> ‘What₁ is such that John wants to know [whether Mary bought it₁]?’

1b) # John-wa [Mary-ga // **Nani-o** katta-kadouka]// shiritagatte-iru-no?³
 what-Acc -Comp_{Whr} -Comp_{Y/N}
 -> ‘Does John want to know [whether Mary bought what]?’

Kitagawa argues that, in (1a), the Wh-phrase *Nani-o* is associated with matrix Comp *-no* and elicits a Wh-question. In contrast, the narrower prosodic unit in (1b) associates the Wh-phrase *Nani-o* with the subordinate Comp *-kadouka*, which is not possible in Japanese grammar; therefore the sentence is unacceptable.

Kitagawa’s claim that different prosody induces different syntactic boundaries correct. However, it is not clear why *focus prosody*, called “Emphatic Prosody” (e.g. p304) by Kitagawa, induces different syntactic processing. “Emphatic Prosody” seems to denote *prosodic focus* plus *post-focus reduction*. The prosodically focused phrase is produced with higher pitch, and the pitch range of the following phrases is significantly reduced (e.g. Sugahara 2003, Koori 2011). However, *focus* is a completely different aspect of prosody from *phrasing*, i.e. placement of prosodic boundaries in a sentence.

In this paper, I shall discuss why the factor that contributes to the grammaticality judgements is not focus prosody, as claimed by Kitagawa, but is actually prosodic boundaries. In other words, it is different phrasing pattern that contributes to different syntactic parsing. It is important here to note that, while focus prosody elicits different realisation of a single phrasing patter, it does not change the phrasing pattern itself. Accordingly, focus prosody does not seem to have an influence on the syntactic processing, at least not as much as prosodic boundaries do.

2. ToBI framework of prosodic annotation

In order to understand prosodic phrasing in Japanese, it is important to first look at the system to represent Japanese prosody.

The most comprehensive model to represent prosody in languages is the *Tones and Break Indices (ToBI)* transcription system (e.g. Beckman & Ayers, 1997), which was originally proposed to represent prosody of English. The leading models to represent Japanese intonation are ToBI

applied to Japanese, *J_ToBI* (Venditti, 2005), and its extended version, *X-JToBI* (Maekawa et al., 2002; ‘X’ meaning *eXtended*). Hereafter in this paper, they will be collectively referred to as *Japanese ToBI*.

According to Japanese ToBI, Japanese prosodic hierarchy consists of three components: *Words*, *Accentual Phrases (APs)* and *Intonation Phrases (IPs)*, as shown in **Table I**.

Table I Prosodic units in Japanese ToBI transcription system

Word	The smallest unit, roughly identical to morphological word
Accentual Phrase (AP)	A unit consisting of one or more Words and bears maximum of one pitch accent
Intonation Phrase (IP)	A unit consisting of one or more APs, in which pitch range is defined

An *Accentual Phrase (AP)* consists of one or more words and bears maximum of one (i.e. zero or one) pitch accent⁴. Therefore, in Japanese, whenever there are more than one pitch accent, there are more than one AP at the same time. Unaccented words usually combine with a neighbouring accented word to form an AP (Venditti, 2005). **Figure 1** is an example of a three-word AP (*uma no re:su* ‘horse race’) and a two-word AP followed by one-word AP (*kame no re:su* ‘tortoise race’). In the left example, unaccented words, *uma* and *no*, on the left can combine with neighbouring accented word *re:su* to form an AP. However, in the right example, *kame*, which is itself an accented word, cannot combine with another accented word *re:su* to form a single AP, although it can still combine with *no*.

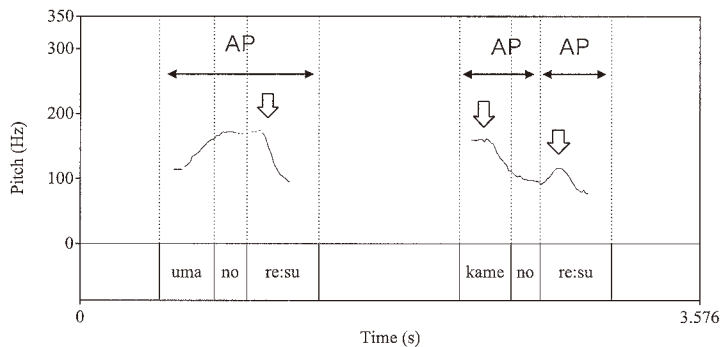


Figure 1 An example of a three-word AP (*uma no re:su* ‘horse race’) and a two-word AP followed by one-word AP (*kame no re:su* ‘tortoise race’) The lines represent pitch-contours and the arrows mark lexical accent.

The higher prosodic unit consisting of one or more accentual phrases is called an *Intonation Phrase (IP)*. IP is the domain in which pitch range is defined. In other words, a new IP is always realised with a new pitch range. The hierarchical relationship is graphically represented in **Figure 2**.

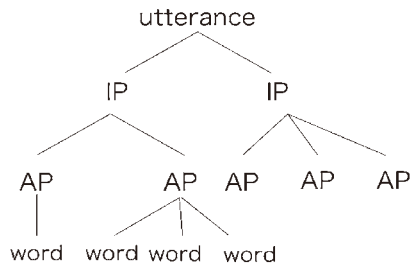


Figure 2 Prosodic hierarchy of Japanese
(IP = Intonation Phrase, AP = Accentual Phrase)

The very factor that defines the pitch peak of each AP is a phenomenon called *downstep*. Because of downstep, pitch peaks of APs following an accented AP are lowered (**Figure 3**). The commonly cited metaphor to describe downstep is a successive “staircase-like” fall of pitch peaks (e.g. Venditti, 2005). Here each *stair* is equivalent to an accented AP and the height of the whole *staircase* corresponds to the pitch range of the IP. In other words, Japanese IP consists of one or more Aps, and downstep is observed for the IP-internal APs.

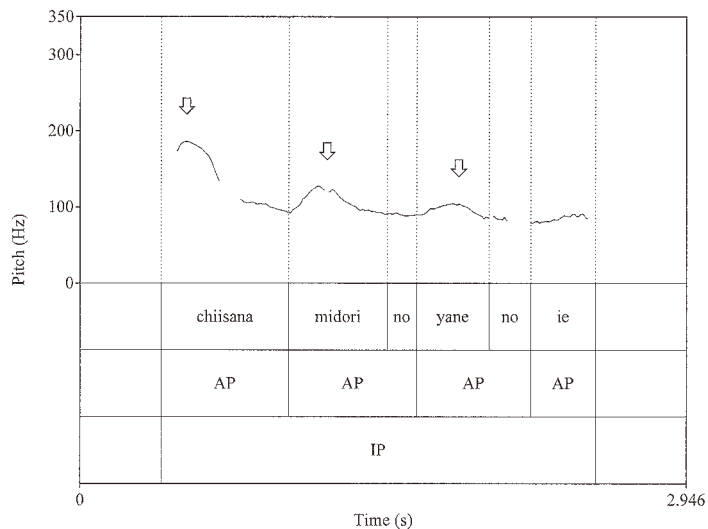


Figure 3 An example of downstep. Each accent is marked with the white arrow. See that the second accent has a lower peak than the first accent; and the third has the lower peak than the second.

The absence of downstep in two successive APs means that the pitch range is reset at their boundary. As shown in **Figure 4**, the reset of the pitch range therefore marks an IP boundary which exists between the two successive APs. Hereafter AP boundaries are represented with ‘2’ and IP boundaries (and therefore AP boundaries at the same time) are represented with ‘3.’ Word boundaries are marked with ‘1.’ These are called *Break Indices (BIs)* in the notation of ToBI.

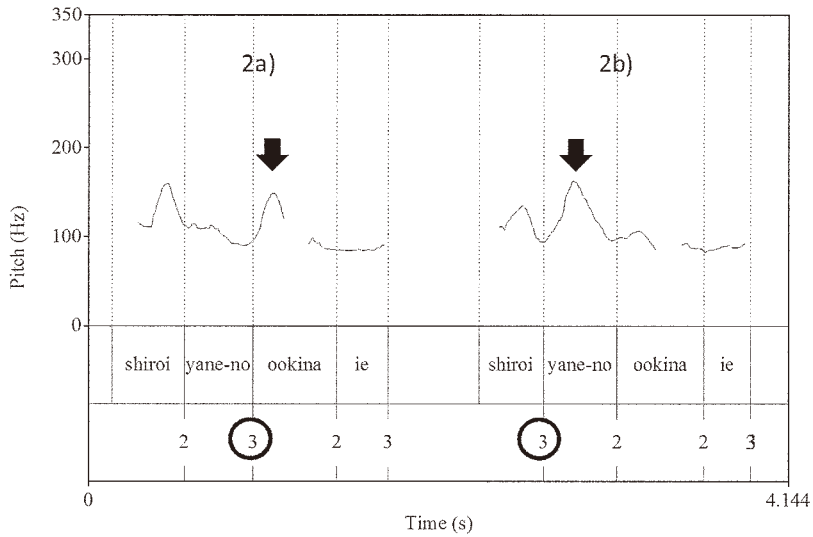
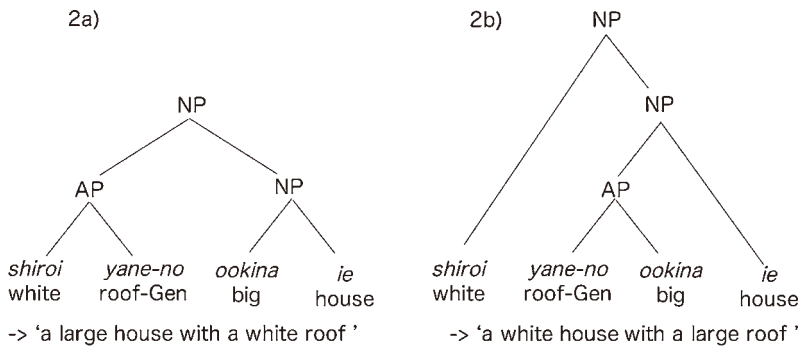


Figure 4 *shiroi yane-no ookina ie* with two different phrasing patterns (male Japanese voice)
The arrows mark the reset of the pitch range.

The prosodic boundaries and syntactic boundaries usually match. For example, the left phrasing pattern of **Figure 4** will induce the syntactic parsing in (2a) below, while the right phrasing pattern will induce the one in (2b). It has actually been reported in previous studies that different phrasing patterns are used to disambiguate syntactically ambiguous sentences (e.g. Venditti et al., 2014).



3. Focus prosody in Japanese

Firstly, it is quite obvious that the kind of prosody used for disambiguation in Japanese can be applied to the manifestation of *focus*. By placing an IP boundary, some phrases are made prosodically salient. For example, the first utterance in **Figure 5** is the normal one, with the pitch peak of *bi.ru* lowered by downstep. On the other hand, the second intonation is used when the

speaker aims to contrast ‘beer’ with something else, e.g. wine. Here, downstep is absent and the pitch range is reset between the two phrases; i.e. the pitch peak of ‘bi:ru’ is as high as that of ‘John-no.’

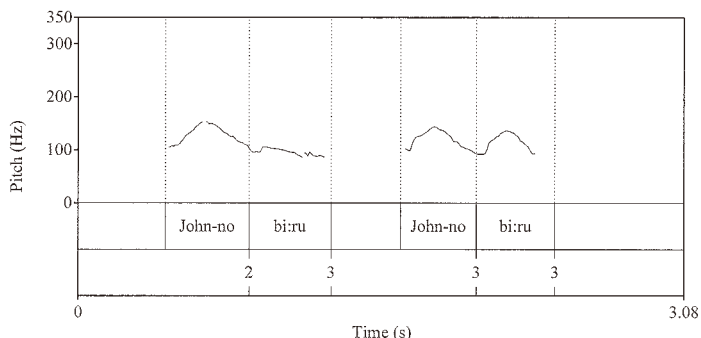


Figure 5 Prosodic disambiguation of *John-no bi:ru* ‘John’s beer.’

The first utterance is the default one while the second utterance is the one in which beer is contrasted with something else; e.g. (*Not John’s wine but*) *John’s beer*. See *bi:ru* in the second utterance is uttered as a separate IP.

Secondly, it is claimed in some studies (e.g. Sugahara 2003, Koori 2011) that focused phrases in Japanese are actually uttered more prominently compared to their original pitch height. In addition, Koori (2011) claims that the gap between a pitch accent and the phrase end is filled with a flat low pitch contour if there are no other pitch accents in between. Such compression of pitch is called *post-focus reduction*. In the experiment of Koori (2011), compression of pitch ranges was observed for the APs following the focused AP.

However, the ToBI transcription system does not have any notation to represent the relative height of each AP. In addition, to date few studies have discussed the effect of the relative height of each AP. However, at least when a linguistic phenomenon is analysed, it should be borne in mind that phrasing (which has the demarcation function) and focus prosody (which is related to relative prominence of each AP) must be clearly distinguished.

4. Yes-No question and Wh-question in Japanese

Unlike English, which implements do-support and subject-auxiliary inversion to form a question, Japanese Yes-No question and Wh-question are formed without any change in word order.

For a Yes-No question, a complementiser *-no* (in case of non-honorific speech) is added at the end of the sentence. Compare (3) and (4). When articulated, Japanese interrogative sentences have sentence-final rising intonation.

Wh-question is formed by a Wh-word and complementiser *-no*. Compare (3) and (5). Again, it has sentence-final rising intonation.

- 3) John-wa sushi-o tabeta
 -Top sushi-Acc ate
 -> 'John ate sushi.'
- 4) John-wa sushi-o tabeta-no?
 -Top sushi-Acc ate
 -> 'Did John eat sushi?'
- 5) John-wa nani-o tabeta-no?
 -Top what-Acc ate-Comp_{wh}
 -> 'What did John eat?'

Kitagawa cites Deguchi & Kitagawa (2002) and Ishihara (2002) to claim that “at least in the Tokyo dialect, Wh-questions in Japanese must generally be accompanied by what he calls ‘Emphatic Prosody’” (p304) and that “Wh-words themselves generally carry prosodic prominence in Wh-questions in Japanese, unlike, for example, in English” (p305). He further argues that in Emphatic Prosody, the accent is followed by post-focus reduction” which “virtually, though not entirely, suppresses all lexical accents up to the end of some clause by compressing their pitch and amplitude ranges” (p305).

Although not empirically attested by any existing research, it seems quite reasonable to assume prosodic salience of Wh-words in, since the Wh-word is pragmatically most salient in Wh-questions. However, whether this salience is the one manifested by focus prosody (i.e. Emphatic Prosody in Kitagawa’s terminology) is questionable.

5. Prosodic boundaries and syntactic disambiguation

Kitagawa claims that the varying degrees of unacceptability among different speakers of sentence (1) (repeated as (6) below), reported by Watanabe (1992, p262), are due to different prosodic realisations of the sentence by each speaker: *local prosodic focus* that ends at the subordinate Comp –*kadouka* and *global prosodic focus* that ends at the matrix Comp –*no*⁵. Kitagawa claims that, when the sentence is parsed with *local prosodic focus*, the Wh-phrase *nani-o* is associated with the subordinate Comp –*kadouka*, which results in ungrammaticality (6b). The underlying assumption here is that the domain of focus prosody coincides with the scope domain of the Wh-question (Deguchi & Kitagawa 2002; Ishihara 2002).

- 6) (?)-??John-wa [Mary-ga nani-o katta-kadouka]
 -Top -Nom what-Acc bought-CompWthr

shiritagatte-iru-no?

want.to.know-CompWh

-> 'What₁ does John want to know [whether Mary bought *t*₁]?'

(Watanabe (1992: 256-257, 263))

6a) John-wa [Mary-ga NANI-o katta-kadouka shiritagatte-iru-no]?

6b) *John-wa [Mary-ga NANI-o katta-kadouka] shiritagatte-iru-no?

Kitagawa is right in claiming that the Wh-phrase is associated either with the matrix Comp *-no* (6a) or the subordinate Comp *-kadouka* (6b) depending on the prosodic pattern. However, it is in fact the IP boundaries (i.e. BI=3), not prosodic focus, that are relevant here.

Kitagawa claims that, for both (6a) and (6b), the Wh-word *NANI* 'what' bears prosodic focus, whether it is a global prosodic focus (6a) or a local one (6b). As a matter of fact, if prosodic focus was placed on *NANI*, the essential constituent of Yes-No question, *shiritagatte-iru-no* 'Does (he) want to know' cannot bear the primary prominence in the sentence. Accordingly, it is highly unlikely that (6b) is interpreted as a yes-no question even with the *local prosodic focus* as claimed by Kitagawa.

On the other hand, although the interpretation with *global prosodic focus* (6a) is theoretically possible, the actual utterance with post-focus reduction up until the very end of the matrix Comp seems physiologically very unlikely. **Figure 6** is the pitch contour of another example with similar intonation pattern given by Kitagawa (in the original source indexed (18) c, p316). The reduced part of the actual recording, which can be accessed at <http://www.iub.edu/~ykling/SoundGallery/EL/index.html>, sounds rhythmically unnaturally compressed. Therefore, the logic behind Kitagawa's claim, although theoretically possible, is actually very unlikely.

7) *Dare-ga Dare-to atteita-kadooka kimi-ni tazuneta-no*
 'who-Nom who-with seeing-Comp_{Whr} you-Dat asked-Comp_{Wh}

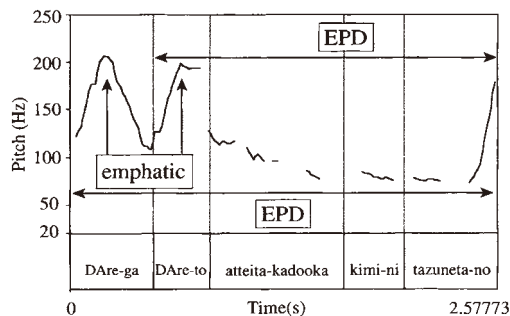


Figure 6 An example of post-focus reduction in Kitagawa (2005) quoted with the permission of the publisher

Furthermore, it is important to note that the empirical evidence of post-focus reduction, i.e. non-realisation of accents on lexically accented words following a focused AP, is found only for the *contrastive focus* (Koori 2011) while the Wh-word bears *information focus* (Gundel & Fretheim, T. 2004). Sugahara (2003), who analysed information focus, defines the *reduction* as a compression of the pitch ranges of the downstepped phrases, not the complete absence of lexical accents. In **Figure 7**, the section marked 'POST-FOCUS' still has two pitch peaks, proving that it still bears accents.

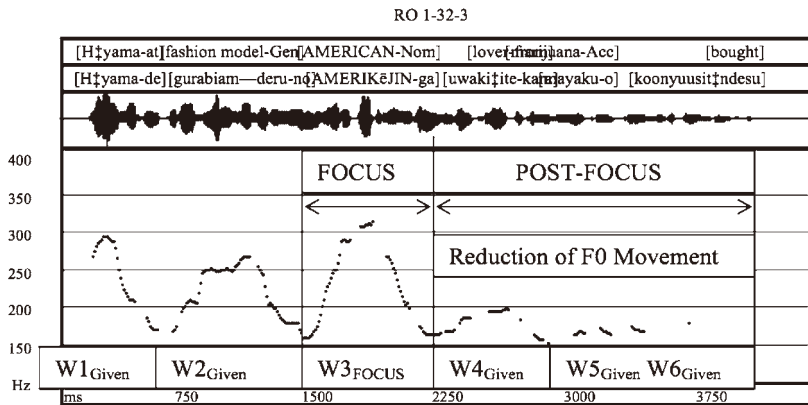


Figure 7 The *post-focus reduction* referred to in Sugahara (2003; p183) quoted with the permission of the author

Note that the section marked 'POST-FOCUS' still bears some pitch peaks, i.e. accents.

In the discussion hereafter, I will use (8) below to analyse the effect of prosodic boundaries on syntactic processing. (8) is the same as (6) except that *-kadouka* has been replaced by *-ka*. As mentioned by Kitagawa, both *-kadouka* and *-ka* have the meaning 'whether' and the constructions of (6) and (8) are "essentially identical" but the replacement with *-ka* will increase acceptability (p311).

8) John-wa Mary-ga *nani-o katta-ka shiritagatte-iru-no*

Firstly, if there is no AP boundary (i.e. if the BIs between *nani-o* and *katta-ka*, and *katta-ka* and *shiritagatteiru-no* are both 1)⁶, *nani-o katta-ka shiritagatte-iru-no* will be parsed as one unit; therefore, the Wh-phrase *nani* will be associated with the IP final Comp (i.e. the matrix comp) *-no*. **Figure 8** represents the pitch contour of the utterance. The rise of pitch at the end of the sentence is interrogative intonation.

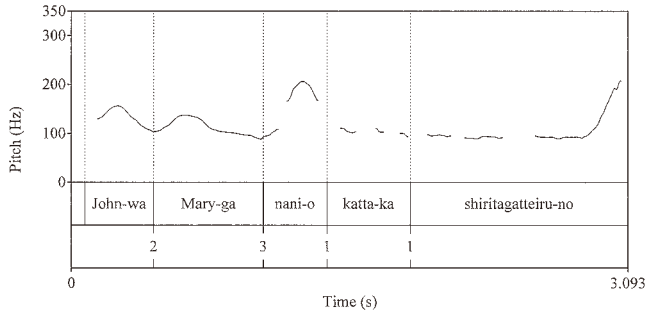


Figure 8 The pitch contour of 'nani-o (1) katta-ka (1) shiritagatteiru-no' (male Japanese voice)

On the other hand, if there is an IP boundary between *katta-ka* and *shiritagatteiru-no* (**Figure 9**), the IP boundary creates the syntactic boundary, and *nani-o katta-ka* and *shiritagatteiru-no* will form two separate constituents; therefore, *nani* will be associated with the IP-final comp (i.e. subordinate Comp) *-ka*. Even when there is an AP boundary (BI=2) between *nani-o* and *katta-ka* (i.e. When the BIs are 2 and 3 respectively; **Figure 10**), the sentence will still induce the same syntactic parsing and it will have the Wh-scope over the subordinate phrase.

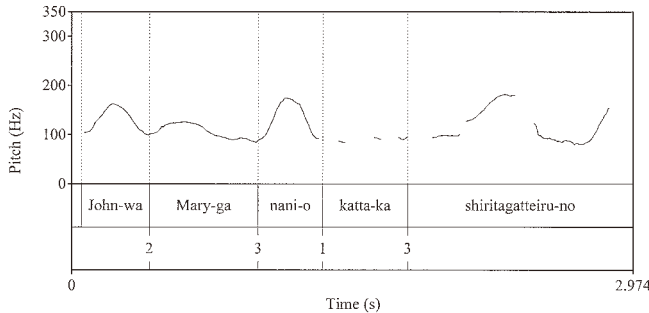


Figure 9 The pitch contour of 'nani-o (1) katta-ka (3) shiritagatteiru-no' (male Japanese voice)

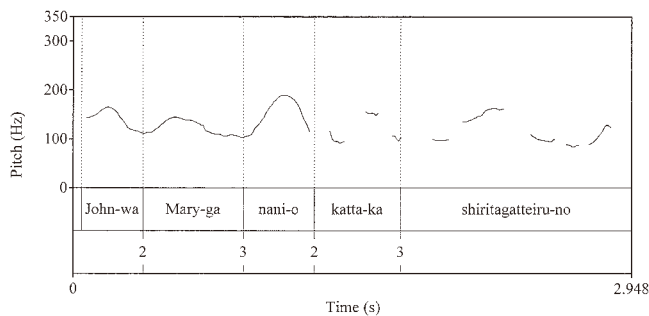


Figure 10 The pitch contour of 'nani-o (2) katta-ka (3) shiritagatteiru-no' (male Japanese voice) Note that the peak of the phrase 'shiritagatteiruno' is as high as that of the preceding phrase 'katta-ka' (i.e. the reset of the pitch range).

The fact that two different patterns of prosody, those in **Figure 9** and **Figure 10**, induce the same syntactic interpretation seems to refute Kitagawa's claim that subordinate Wh-scope requires focus prosody placed on the phrases; subordinate-Wh-scoped syntactic parsing is perfectly possible without any kind of reduction (including downstep) after the accent, which seems to be a necessary condition of prosodic focus.

The problem arises when the boundary between *katta-ka* and *shiritagatte-iru-no* is an AP boundary (BI=2). It is not certain whether the AP boundary is large enough to act as a syntactic boundary. Sentence (8), which has BI 1 between *nani-o* and *katta-ka*, and BI 2 between *katta-ka* and *shiritagatte-iru-no* (**Figure 11**), will be parsed with a syntactic boundary that matches the AP boundary (i.e. the subordinate Wh-scope).

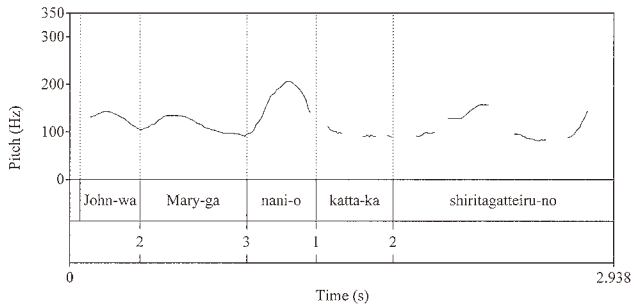


Figure 11 The pitch contour of 'nani-o (1) katta-ka (2) shiritagatteiru-no' (male Japanese voice)

However, a different sentence (9) below with the same kind of prosody (**Figure 12**) is interpreted with matrix Wh-scope, i.e. *nani* is associated with *-no*.

- 9) John-wa Mary-ga nani-o katta-to omotteiru-no
 -Top -Nom what-Acc bought-Comp think-Comp
 -> 'What does John think that Mary bought?'

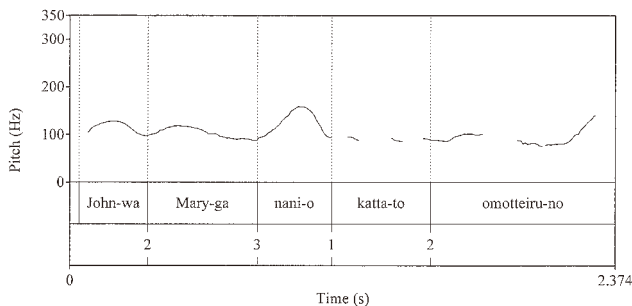


Figure 12 The pitch contour of 'John-wa Mary-ga nani-o (1) katta-to (1) omotteiru-no' (male Japanese voice)⁷

Such difference seems to arise from the difference in the *default* syntactic structures. Kitagawa cites Fodor’s (2002) “Implicit Prosody Hypothesis,” which states that when more than one prosodic pattern are available for a sentence, the reader’s parsing of the sentence will be influenced by “default prosody.”⁸ Sentence (8) is biased towards a Yes-No question (i.e. *nani* associates with *-ka*). This is supported by the higher acceptability of (6) with a local Wh-scope than with a global one. In other words, it is more acceptable as a Yes-No question than as a Wh-question. Remember the only difference between (8) and (6) is that of *-ka* and *-kadouka*, which, according to Kitagawa, is very little. On the other hand, (9) seems to be biased towards Wh-question (i.e. *nani* associates with *-no*) since the local Wh-scope interpretation would be highly unacceptable. In fact, the acceptability of (9) with no AP (or IP) boundary (**Figure 13**) is much higher than (8) with no AP boundary, i.e. the intonation depicted in **Figure 8**.

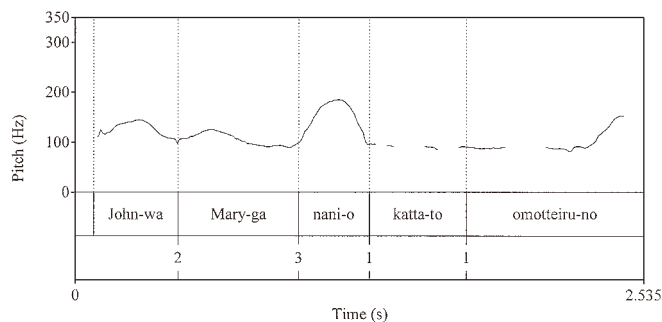


Figure 13 The pitch contour of ‘nani-o (1) katta-to (2) omotteiru-no’ (male Japanese voice)

It can be deduced therefore that an AP boundary does NOT affect syntactic boundary, and that a sentence is parsed in the following process.

- i. IP boundaries act as syntactic boundaries.
- ii. Each constituent segmented by IP boundaries will be further parsed with reference to its *default* syntactic structure. Syntactic boundaries will be defined accordingly.

6. Conclusion

The current paper supports Kitagawa’s general idea that grammaticality judgement of a written sentence is influenced by its hidden prosody. On the other hand, this paper refutes his main point and argues instead that it is not focus prosody, as Kitagawa claims, but Intonation Phrase boundaries that are relevant to syntactic processing. If more than one syntactic interpretation are available, the default syntactic structure will be adopted.

7. References

- Beckman, M. E., & Ayers, G. (1997). Guidelines for ToBI labelling. The OSU Research Foundation, 3.
- Gundel, J. K., & Fretheim, T. (2004). Topic and Focus. In L. R. Horn and G. Ward (Eds.) *The Handbook of Pragmatics*, pp175-196. Blackwell.
- Deguchi, M. and Kitagawa, Y. (2002). "Prosody and Wh-questions," *NELS* 32, pp73-92.
- Fodor, J. D. (2002). "Prosodic Disambiguation in Silent Reading," *NELS* 32, pp113-137.
- Ishihara, S. (2002). "Invisible but Audible Wh-scope Marking: Wh- constructions and Deaccenting in Japanese," *WCCFL* 21, pp180-193.
- Kitagawa, Y. (2005). Prosody, syntax and pragmatics of Wh-questions in Japanese. *English linguistics*, 22(2), pp302-346.
- Koori, S. (2011). The phonetic Characteristics of Broad Focus in Tokyo Japanese. "東京方言における広いフォーカスの音声的特徴." *Onsehgengo-no-kenkyu*, 5, pp13-20.
- Maekawa, K., Kikuchi, H., Igarashi, Y., & Venditti, J. J. (2002). X-JToBI: an extended j-toBI for spontaneous speech. In Proc. ICSLP, pp1545-1548.
- Sugahara, M. (2003). *Downtrends and post-FOCUS intonation in Tokyo Japanese*. Doctoral dissertation. University of Massachusetts Amherst.
- Venditti, J. J. (2005). The J_ToBI model of Japanese intonation. *Prosodic typology: The phonology of intonation and phrasing*, pp172-200.
- Watanabe, Akira (1992) Subjacency and S-structure movement of WH-in-situ. *Journal of East Asian Linguistics*, 1, 255-291.

Endnotes

- 1 Throughout this paper, 'Kitagawa' represents 'Kitagawa (2005)' unless stated otherwise. The page numbers without an author's name also represent those of Kitagawa (2005).
- 2 Prosodic boundaries irrelevant to current analysis are omitted. Shading is added so that the prosodic unit is visually demarcated.
- 3 I follow Kitagawa's notation that # denotes unacceptability of a sentence with indicated prosody.
- 4 Be careful therefore with the use of the terminology in that a phrase without accent can still be referred to as *Accental Phrase*.
- 5 The original terminology adopted by Kitagawa are *Local Emphatic Prosody* and *Global Emphatic Prosody*. However, for the sake of simplification, the current paper refers to them as *local prosodic focus* and *global prosodic focus* respectively.
- 6 Again, it is important to know this phrasing is physiologically very hard.
- 7 Relatively lower peak of the last AP *omotte-iru-no* compared to the one in **Figure 11** *shiritagatte-iru-no* seems to be because of the different lengths of the APs: 9- vs 7-mora. Also, the accent in the former is on the fourth mora whereas that in the latter is on the second mora, which is also expected to have lowered the pitch peak of the latter.
- 8 In this paper, I will not go into detail to discuss what defines "default prosody"; the main purpose of this paper is to refute Kitagawa's claim that it is focus prosody that defines the syntactic processing.

