Intonational Phrasing in Discourse Analysis

Hiroyuki Nagahara

Intonation is a crucial factor in analyzing texts in discourse. One of its most important characteristics is the intonational phrase structure of the text, which determines how a text is organized in terms of intonational breaks from the phrase and sentence level, up to the text level. For example, an utterance might consist of a single intonational phrase without any break in its intonational inflection, or it might consist of several phrases. The intonational phrase organization of an utterance reflects the semantic and discourse-functional structure of the utterance as well as such paralinguistic factors as emotionality and attitudes of the speaker. Therefore, it is crucial to define intonational phrases strictly from the phonological perspective of the target language, then correlate the intonational structure of an utterance to its semantic and discourse-functional structure.

Since languages differ in intonational characteristics, we cannot directly apply the definition based on one language to another language. For example, the definition of the so-called 'intonation unit' in Chafe (1987) and Du Bois et al. (1992) is based on the characteristics of English phonology, and so it is almost absurd to apply the same definition to transcribe intonational breaks for other languages such as Japanese. To study the intonational structure of Japanese, we have to base our definition strictly on the characteristics of Japanese phonology, and that is one of the objectives of this paper.

Another concern of this study has to do with what I call 'transcriber's bias', by which the transcriber's 'phonological judgment' is influenced by the semantic and functional structure of the very utterance s/he is trying to elucidate. This is a problem inherent in any studies involving text transcriptions, particularly in intonation studies because of the close relationship between intonation and semantic

and functional structure. I will address this issue by evaluating a particular transcription system in terms of the degree of fitness between this system and a strictly phonological segmentation method (to be defined in the first part of this paper).

1. Intonational Phrase in Japanese

The theory of prosodic organization is concerned with prosodic categories (organized into a nonrecursive hierarchy distinct from syntactic structure) which serve as domains of phonological and phonetic processes (Selkirk 1984; Nespor and Vogel 1986; Hayes 1989). One of the most influential works concerned with Japanese prosody is Pierrehumbert and Beckman (1988), which I will refer to as P/B hereafter. The prosodic categories assumed in P/B are the utterance, the intermediate phrase, the accentual phrase, and the phonological word. In this study I will follow them except for the term *intonational phrase* in preference to their term *intermediate phrase* for the sake of stylistic felicity. These prosodic categories are organized as follows:

1) Prosodic categories in Japanese

Utterance	[								]
Intonational phrase	[		][	][		][			]
Accentual phrase	[		][	][	][	][		][	]
Phonological word	[	][	][	][	][	][	][	][	]

In the following, $]_u$, $]_i$, $]_a$, and $]_w$ represent the utterance, intonational phrase, accentual phrase, and phonological word boundaries, respectively.

There are two types of tones in Japanese within P/B's framework: boundary tones and pitch accents (Nagahara 2000). Boundary tones are simplex tones such as H's and L's linked to prosodic phrase boundaries, whereas pitch accents (marked with an asterisk) are complex tones of the form H*L linked to the moras of accented syllables. In this usage the term *accent* refers to a lexical property, whereas the term

pitch accent refers to a tone linked to the mora of an accented syllable.

Intonational inflection is further subject to the process known as catathesis or downstep. It is a process in which a pitch accent H*L triggers lowering of the following pitch register. The following pair of adjective-noun phrases illustrates the effect of catathesis. The phrase in Figure 1.1 contains an accented adjective *umái* 'tasty' followed by a noun *mamé* 'beans' in the sentence frame _____ *wa arimasén* 'There are no _____'. The phrase in Figure 1.2 is identical to that in Figure 1.1 except for the unaccented adjective *amai* 'sweet'.

Figure 1.1

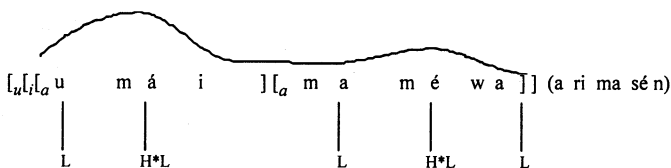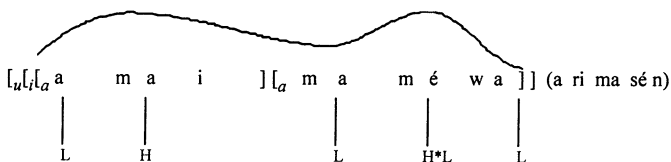A schematic representation of catathesis on the phrase *umái mamé (wa arimasén)* 'There are no tasty beans'

[ᵤ[ᵢ[ₐ u   m á   i   ] [ₐ m   a   m é   w a]] (a ri ma sé n)

     L   H*L   L   H*L   L

Figure 1.2

A schematic representation of lack of catathesis on the phrase *amai mamé (wa arimasén)* 'There are no sweet beans'

[ᵤ[ᵢ[ₐ a   m a   i   ] [ₐ m   a   m é   w a]] (a ri ma sé n)

     L   H   L   H*L   L

In the first phrase (Figure 1.1) the fundamental frequency (F0) of the pitch accent on the noun is lower due to the catathesis triggered by the preceding pitch

accent on the adjective. In the second phrase (Figure 1.2) there is no catathesis effect on the noun's F0 level since there is no pitch accent preceding it.

Catathesis is delimited by intonational phrase boundaries. That is, a new intonational phrase resets the F0 level of the following pitch register. The pair of phrases in the following figures illustrates the difference in pitch inflection between a single intonational phrase (Figure 2.1) and two intonational phrases (Figure 2.2).

Figure 2.1

A schematic representation of catathesis on the phrase *imootó no marí ga* (younger-sister of ball NM$_{[1]}$) 'my sister's ball' with no intonational phrase break
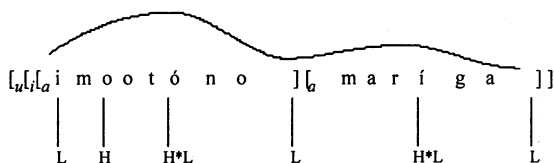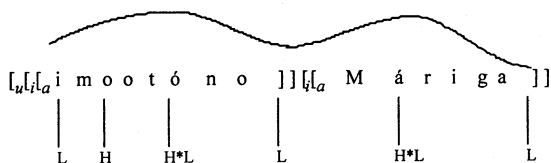

Figure 2.2

A schematic representation of lack of catathesis on the phrase *imootó no Mári ga* (younger-sister of Mari NM) 'Mari, my sister' intervened by an intonational phrase break

In the phrase in Figure 2.1 there is catathesis: The F0 level of the second pitch accent is lower than that of the first pitch accent. In the phrase of Figure 2.2 on the other hand, there is no catathesis because the two pitch accents are intervened by an

intonational phrase boundary.

2. Comparison with 'Intonation Unit' in Discourse Analysis

In contrast with the strictly phonological definition of our 'intonational phrase', the concept 'intonation unit' used in various discourse studies is defined on the basis of such characteristics as the coherency of pitch contour, baseline pitch reset, pause, anacrusis[2], and syllable lengthening (Chafe 1987; Du Bois et al. 1992). However, these are characteristics of English phonology, so they cannot be directly applied to Japanese (which differs from English in terms of sound and intonational structures). Anacrusis in particular is a characteristic of stress-timing languages such as English, in contrast to mora-timing languages such as Japanese.

Moreover, there is cause for concern regarding 'transcriber's bias', by which transcribers are influenced by the semantic and functional structure of the very text they try to analyze, in the studies concerned with intonation units for whatever language, be it English or Japanese. That is, we cannot rest assured that a particular segmentation of a text into intonation units was not due to the transcriber being influenced by the semantic and functional structure of the text. For instance, for the utterance of displacement *Ame futteta yo, ano toki wa* 'It was raining, at that time' some transcribers might segment it into two intonation units, *Ame futteta yo* and *ano toki wa* (regardless of the fact that its phonological analysis might indicate there is no intonational break in the utterance) simply because that is exactly the semantic and functional division of the utterance. This is a serious problem because the very idea of intonation units is to correlate the phonological organization with the semantic and functional organization of texts in discourse, and being influenced by the latter to determine the former is circularity in and of itself.

The objective of this study is to determine the degree of agreement between the segmentation of a particular Japanese text into intonation units by the Du Bois transcription system at the University of California, Santa Barbara (which will be

referred to as UCSB transcription hereafter) and the segmentation into intonational phrases by the strictly phonological criterion of catathesis, described in the previous section.

2.1 Method

The sample text used in this experiment is a description by a woman (a native of Tokyo) of her experience with the Tokyo air raid on March 10, 1945, and runs as follows:

2) The text used in the experiment

The equal sign in this transcription represents a pause.

Line breaking is arbitrary and does not have any bearing on intonational phrasing.

See Note 1 for the explanation of the symbols and abbreviations used here.

Lines

```
1   datte = ano: = atashi wa ne  kitasenji       kara = ano: =
    INJ    INJ    I     TP IT (place name)       ABL   INJ

2   kitasenji      ni   ita no   sono  toki wa ne =
    (place name)   LC   was SE   that  time TP  IT

3   sangatsu tooka n   toki wa =
    March    tenth LK  time TP

4   suide = ano: = ane        ga    fukagawa     ni   ite =
    and     INJ    older-sister NM  (place name) LC   be-and

5   kameido      ni   niken   shinseki ga  atte =
    (place name) LC   two     relative NM  be-and

6   soide ano: = kuushuu = tte   no    wa
    and   INJ    air-raid   QT    NML   TP

7   moo     sugokatta  tte  no   wa  shitteru wake =
    EMP     was-severe QT   NML  TP  know     SE

8   m = me  de    wa  minakatta   kedo =
    FRG eye INST  TP  didn't-see  though

9   dakara = soide = uchi: = ga  yakenai    noni
    so       and     home   NM  not-burn   though

10  daremo konai     kara =
    nobody not-come  because

11  shinjatta no kana: to   omotta no
    died      SE  Q    QT   thought SE

12  mikka  tatte     ne
    3-days pass-and  IT
```

English translation

  1 So, from Kitasenju,

  2 I was in Kitasenju at that time

  3 on March 10th.

  4 And my older sister was in Fukagawa,

  5 I had two relatives in Kameido

  6 and, speaking of air raids,

  7 I know how severe they were

  8 although I didn't see them with my own eyes.

  9 So, although my home did not burn down,

10 nobody came, so

11 I thought they had died,

12 (I thought) three days later.

There were 8 transcribers trained in the UCSB transcription system. Five of them were native speakers of Japanese, and the rest nonnative speakers familiar with the Japanese language. They listened to the tape and transcribed it in terms of intonation units, following the guidelines laid out in Du Bois et al. (1992). I myself did the segmentation of the same text into intonational phrases by analyzing the catathesis patterns on the pitch tracks made by the CSpeech software at the University of California, Los Angeles.

2.2 Results

One of the results of the experiment was the degree of agreement among the transcribers (Iwasaki, Nagahara, and Ono 1993), which was very high among the native transcribers, as expected. Our concern in this study, however, is the degree of agreement between their segmentation and the strictly phonological segmentation of the same text into intonational phrases.

By phonological segmentation I detected 21 boundaries of intonational phrases on the pitch tracks. Of the 21 boundaries 18 (or 86% of 21) were exactly the places where the majority (3 or more) of the five native transcribers agreed on. I

would say this was a very high degree of agreement, despite our differences in methodology.

The disagreements with the majority of the native transcribers occurred for two types of structure: displacements and interjection/function-word sequences. The following summarizes the instances of displacement disagreement between the two segmentation methods:

3) Displacement disagreement
   Brackets = intonational phrases; P = phonological analysis.

Lines 2, 3: disagree
kitasenji ni ita no   sono toki wa ne = sangatsu tooka n toki wa =
                                        'I was in Kitasenju at that time, on March 10th.'
[   ] [   ]   P
[   ] [   ] [   ]   Majority

Lines 7, 8: agree
 moo   sugokatta tte no wa shitteru wake = m   me de wa minakatta kedo =
       'I knew how severe they had been although I hadn't see them with my own eyes.'
    ] [   ] [   ] [   ]   P
    ] [   ] [   ]   Majority

Lines 11, 12: agree
  shinjatta no   kana: to omotta no   mikka   tatte   ne
                       'I thought they had died; (I thought) three days later.'
[   ] [   ]   P Majority

The native transcribers consistently separate the main clause from the displaced comments (and the comments from one another) without regard to the presence or absence of a pause. (The nonnative transcribers on the other hand, rely on pause more than the native transcribers do. In fact, they agree with the majority of the native transcribers when there is a pause separating the main clause from the following comments.) It seems clear that the native transcribers are biased by the idea that displaced comments should be separate from the main clause, disregarding their intonational characteristics.

There were a few instances of interjections and function words in sequence in the text. The issue here is whether these sequences make up a single intonational phrase or not and whether a particular interjection is part of the preceding or following intonational phrase. The following summarizes the cases of our disagreement:

4) Disagreement on interjection/function-word sequences

   Numbers = native transcriber 1, 2, 3, etc.

Line 1	[datte] = [ano:]	P 12345		
Line 1	·· · · · kara = ano:]	P	· · · · kara] = [ano:	12345
Line 4	[suide = ano:]	P	[suide] = [ano:]	12345
Line 6	[soide ano]	P 125		

We agreed with each other completely on the interjection sequence on line 1, *datte = ano:* being two separate intonational phrases. In the case of the sequences *kara = ano:* on line 1 and *suide = ano:* on line 4, the phonological segmentation indicates that they are single intonational phrases, whereas all the transcribers judged them to be two phrases. In the case of the last sequence *soide ano* (in which there was no pause and *ano* is short) three of the transcribers agreed with the phonological analysis.

It is clear from this case of sequences of interjections and function words, that the transcribers 3 and 4 always separate the two words, regardless of the presence or absence of a pause and vowel lengthening, whereas the transcribers 1, 2, and 5 are sensitive to the presence of a pause. Either group, however, is not sensitive to the actual intonational characteristics, therefore, biased by their preconceived ideas about pauses and interjection and function words.

2.3 Discussion

There are a few things we can learn from this experiment. First, pauses are more important than pitch characteristics for at least some of the transcribers. Witness the

fact that pause can override the pitch criterion in interjections for those transcribers. Here I am not saying that pauses are not important, but that some transcribers are very sensitive to them despite the fact that they are almost irrelevant to intonational phrasing.

Second, semantic and functional breaks (between the main clause and its displaced comments, for example) are more important than pauses for all the transcribers. Witness the fact that semantic and functional breaks can override even the pause criterion in the displacement cases (where the native transcribers consistently disregarded pauses), whereas only some of the transcribers disregarded pauses for interjections. Another instance that shows the importance of semantic and functional structure is the line 9 in the text, where the speaker says *uchi:* = *ga* with the longest pause in the text. No transcriber (except one nonnative transcriber) separated the subject and the subject marker here, indicating their bias that the subject and the subject marker should be always grouped together. (The nonnative transcribers seem to rely exclusively on pitch and pause, resulting sometimes in a phrasing pattern that no native transcriber would dare to propose. The fact that one nonnative speaker actually separated *uchi:* from *ga* on line 9 indicates the nonnativeness of his command of the language.)

Third, logically speaking, the number of intonational phrases (for a given discourse segment) derived from UCSB transcription should be greater than the number of intonational phrases derived from the pitch-only analysis, since the former makes use of more criteria than just the pitch. In fact, that is the case for English (Schuetze-Coburn, Shapley, and Weber 1991) as well as in this Japanese case. However, there is a difference between English and Japanese. In Japanese, the percentage of agreement between acoustic and transcriptional segmentation is greater than that in English. Perhaps, this is because in Japanese pitch resettings tend to occur at semantic and functional breaks. That is, in Japanese the phonological organization of intonational phrases tends to overlap with the semantic and

functional structure.

3. Conclusion

I have shown that the degree of agreement among our transcribers was high and in the majority of cases, their text segmentation into intonation units correlated well with the strictly phonological segmentation. Yet, they were also influenced by pauses and particularly the semantic and functional structure of the text. That is, they were not really 'listening' to the tape but actually 'reading' the semantic and functional structure of the text, particularly in the cases of displacement phrases and interjection/function-word sequences. This is because the original proposals regarding intonation units are based on English and many studies just attempt to apply the original criteria to Japanese, and also because of the inherent difficulty of native speakers to just 'listen' to phonological characteristics without the intrusion of their native understanding of the semantic and functional structure of texts in discourse. Nonetheless, it is important for us to start listening to what intonation is trying to tell us.

Notes:

1. The following symbols and abbreviations are used in this paper. The equal sign means a pause. The colon indicates vowel lengthening. TP = topic, NM = nominative, ABL = ablative, LC = locational, INST = instrumental, QT = quotation, IT = interactional, SE = sentence-ending, Q = question, LK = linking element, INJ = interjection, NML = nominalizer, EMP = emphasis, FRG = fragment.

2. Anacrusis is a sequence of unstressed syllables (at the beginning of a phrase or sentence)

pronounced faster than a comparable sequence of syllables headed by a stressed syllable.

<div align="center">

References:

</div>

Chafe, Wallace (1987), 'Cognitive Constraints on Information Flow', R. Tomlin (ed.), *Coherence and Grounding in Discourse,* John Benjamins, Amsterdam.

Du Bois, John, Susanna Cumming, Stephan Schuetze-Coburn, and Danae Paolino (1992), *Discourse Transcription,* University of California, Santa Barbara.

Hayes, Bruce (1989), 'The Prosodic Hierarchy in Meter', P. Kiparsky and G. Youmans (eds.), *Rhythm and Meter,* Academic Press, San Diego.

Iwasaki, Shoichi, Hiroyuki Nagahara, and Tsuyoshi Ono (1993), 'The Intermediate Phrase vs. the Intonation Unit', presented at Japanese/Korean Linguistics Conference (October 15, 1993), University of California, Los Angeles.

Nagahara, Hiroyuki (2000), 'Seisa-to Intoneeshon-no Parametorikkusu (The Parametrics of Gender and Intonation)', *Kotoba* (Journal of Modern Japanese Language Association, Tokyo) 21: 37–44.

Nespor, Marina and Irene Vogel (1986), *Prosodic Phonology,* Foris Publications, Dordrecht, The Netherlands.

Pierrehumbert, Janet and Mary Beckman (1988), *Japanese Tone Structure,* MIT Press, Cambridge, Massachusetts.

Schuetze-Coburn, Stephan, Marian Shapley, and Elizabeth Weber (1991), 'Units of Intonation in Discourse: A Comparison of Acoustic and Auditory Analyses', *Language and Speech* 34: 207–234.

Selkirk, Elisabeth (1980), 'Prosodic Domains in Phonology: Sanskrit Revisited', M. Aronoff and M.L. Kean (eds.), *Juncture,* Anma Libri, Saratoga, California.

<div align="right">

(ながはら　ひろゆき)

</div>