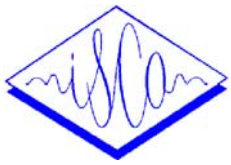


Speech Prosody 2002
Aix-en-Provence, France
April 11-13, 2002

Conditions on Post-FOCUS Dephrasing in Tokyo Japanese

Mariko Sugahara

Department of Linguistics
University of Massachusetts, Amherst
sugahara@linguist.umass.edu

Abstract

This paper investigates whether or not presence or absence of a post-FOCUS Minor Phonological Phrase boundary in Tokyo Japanese is conditioned by grammatical factors such as the presence or absence of a syntactic boundary and a pitch accent. Each boundary of a Minor Phonological Phrase (henceforth Minor Phrase or MiP) coincides with an F0 rise from the initial syllable to the second syllable. The result of our production experiment shows that there is always a significant F0 rise (i.e. a MiP boundary) at the onset of a post-FOC word as long as (i) an XP boundary or (ii) a pitch accent is present. That is to say, only when conditions (i) and (ii) are absent, the MiP boundary disappears at the onset of a post-FOC word. In this way, post-FOC dephrasing in Tokyo Japanese is not arbitrary but governed by grammatical factors.

1. Introduction

The most widely accepted view of phonological phrase formation is that it is affected by multiple factors such as morpho-syntactic structure ([4], [8], [9]), information structure, e.g. presence or absence of a contrastive focus ([3], [5], [6], [10]), purely prosodic requirements on size and length of a phonological unit ([1]), speech rate ([2]), and so on. In languages like Japanese which have two types of words in terms of presence or absence of a pitch accent, whether a word is specified for an accent or not is another factor in determining a prosodic structure ([4]). The main concern of this article is how these multiple factors, especially syntactic structure and information structure as well as presence or absence of a pitch accent, interact to determine where to insert a Minor Phrase boundary in Tokyo Japanese. (A "Minor Phrase" is also known as an "Accentual Phrase").

In Tokyo Japanese, Minor Phrase boundaries may disappear at the left edge of items interpreted as already "given" in a discourse when those items are preceded by a contrastive focus, henceforth FOC ([5], [6]). However, the following questions are not fully answered yet:

- Is the absence of a post-FOC Minor Phrase boundary variable in view of different conditions or obligatory?
- If it is variable, then what are the grammatical conditions on presence or absence of a post-FOC MiP boundary? Do syntactic conditions, e.g. presence or absence of a XP boundary play any role? Do accent conditions, i.e. presence or absence of a pitch accent, matter?

The main goal of this paper is to investigate those questions. The organization of this paper is the following. In Section 2, basic notions and assumptions about Japanese prosodic structures will be introduced. In Section 3, reading materials and the procedure of the production experiment will be presented. Section 4 will discuss the results of the experiment.

2. Basic Notions and Assumptions

Throughout this article, I assume the prosodic hierarchy of Selkirk ([8]) and others such that syllables are built into a PWd, and PWds are grouped into a MiP, and so on.

Prosodic Hierarchy
- Utterance (Utt)
- Intonational Phrase (IP)
- Major Phonological Phrase (MaP)
- Minor Phonological Phrase (MiP)
- Prosodic Word (PWd)
- Foot
- Syllable
- Mora

The left edge of each MiP in Japanese is marked with two edge tones: a L% boundary tone is associated with the initial mora and a H- phrasal tone is associated with the second one.

[(m m m)$_{PWd}$ (m m)$_{PWd}$]$_{MiP}$ [(m m m)$_{PWd}$]$_{MiP}$
L% H- L% H-

Figure 1: Sequence of unaccented MiPs & edge tones.

When a MiP consists of an unaccented PWd only, then those are the only tones associated with the phrase. When it consists of an accented PWd, a bitonal pitch accent, H*+L, appears to the right of those two edge tones.

[(m m m' m)$_{PWd}$]$_{MiP}$ [(m m m' m)$_{PWd}$]$_{MiP}$
L% H- H*+L L% H- H*+L

Figure 2: Accented MiPs, edge tones & a pitch accent.

Due to the presence of these edge tones, the onset of a MiP coincides with a F0 rise from the initial to the second syllable.

3. Experimental Materials and Procedure

To investigate questions raised in the Introduction, a production experiment was carried out. The experimental materials and procedure will be presented in this section. First of all, sentences consisting of accented words and those consisting of unaccented words were used. In this article, I will only focus on the initial word (Word 1) and the second word (Word 2) of those sentences.

Sequence of Accented Words: AA

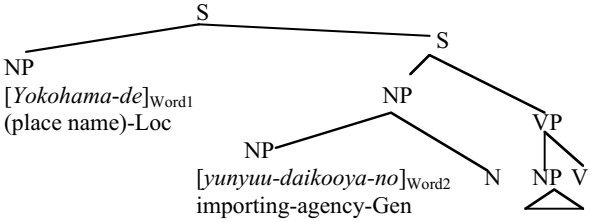
[*Háyama*-Case]$_{Word1}$ [*yunyuu-daikóoten*-Case]$_{Word2}$ …
[Hayama-Case]$_{Word1}$ [importing-agency-Case]$_{Word2}$ …

Sequence of Unaccented Words: UU

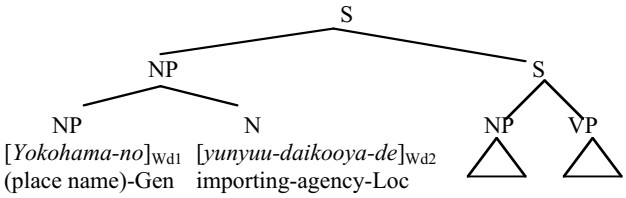
[*Yokohama*-Case]$_{Word1}$ [*yunyuu-daikooya*-Case]$_{Word2}$…
[Yokohama-Case]$_{Word1}$ [importing-agency-Case]$_{Word2}$ …

Each of those word sequences was further divided into two syntactic forms: Structure 1 and Structure 2. In Structure 1, Word 2 is a locative phrase which adjoins to the whole sentence and Word 2 corresponds to an NP that modifies the head noun of a subject NP of that sentence. As a result, there is an XP boundary at the left edge of Word2. In Structure 2, on the other hand, Word 2 is a head noun of a subject NP and Word1 is an NP which modifies the head noun. Consequently, there is no XP boundary at the onset of Word 2 in Structure 2.

Structure 1: XP Boundary between Wd1 and Wd2



Structure 2: No XP Boundary between Wd1 and Wd2



Furthermore, there are two contexts in which those sequences of words were produced. In one context (i.e. Context NN), both Word 1 and Word 2 are interpreted as simply "new". In the other context (i.e. Context FG), Word 1 is interpreted as a contrastive focus (i.e. FOC) and Word 2 is interpreted as already "given".

Context NN: Word 1 = new, Word 2 = new

e.g.
Speaker: *chotto kiiteyo.*
"Hey, just listen to me."
Experimenter: *nani?*
"What?"
Speaker: [*Yokohama-no*]$_{Wd1}$ [*yunyuu-daikooya-de*]$_{Wd2}$ …
"At an importing-agency in Yokohama ... … "

Context FG: Word 1 = FOC, Word 2 = given

e.g.
Experimenter: *Aóyama-de yunyuu-daikooya-no maneejaa-ga ... kiitaga, hontookánee?*
"I heard that, in Aoyama a manager of an importing-agency …, but is it true?"
Speaker: *chigaimásu.*
"No, that's not true."
Aóyama-dewa-naku [*YOKOHAMA-de*]$_{Wd1}$ [*yunyuu-daikooya-no*]$_{Wd2}$ *maneejaa-wa...*
"Not in Aoyama but in Yokohama, the manager of the importing agency …"

Given all of those conditions, there are 2*2*2=8 target forms in total. Each target form is named after the conditions they match. For example, a sequence of unaccented words which has Structure 1 and produced in Context NN is called "UU1NN", a sequence of accented words with Structure 2 in Context FG is called "AA2FG", and so on.

Recordings took place in a sound proof studio in the University of Tokyo, Komaba, between July and August of 2001. Three Tokyo Japanese native female speakers (SK, RO, MR) produced each of those forms in three different sessions. Those forms were embedded in a dialogue as already shown above, and presented to the speakers typed on a card using the Japanese writing system. No commas were used in the written materials so that it is up to the speakers where to insert phrase breaks. In each dialogue, each speaker and the experimenter played a hypothetical role to exchange a conversation. They were asked to give natural renditions as much possible. Those dialogues were recorded using a SONY DAT recorder, and were re-digitized at the sample rate of 22kHz using PitchWorks on a Mac PowerBook G3.

The F0 value (Hz) associated with the right edge of each syllable of Word 1 and Word 2 was measured. We refer to each of those syllables in the following way: the initial syllable of Word 1 is S11, the initial syllable of Word 2 is S21, the second syllable of Word 2 is S22, and so on.

e.g.

[*yo*	*ko*	*ha*	*ma*	*no*]$_{Wrd1}$	[*yu*	*nyuu*	*dai*	*koo*	*ya*	*de*]$_{Wrd2}$
S11	S12	S13	S14	S15	S21	S22	S23	S24	S25	S26

Primarily, we are interested in whether or not there is any F0 rise from S21 to S22 in each of the eight target forms. We interpret an F0 rise from S21 to S22 as an indication that L% and H- are present at the onset of Word 2, which in turn signals the presence of a MiP boundary there.

4. Results and Discussion

As mentioned in the previous section, the main goal of this experiment is to explore whether or not the presence or absence of an F0 rise from S21 to S22 of Word 2 is conditioned by the three factors introduced above: (a) the accent factor, (b) the syntactic factor and (c) the context factor. It turned out that an F0 rise was absent only when (i) no XP boundary was at the onset of Word 2, (ii) both Word 1 and Word 2 were unaccented, and (iii) Word2 was a post-FOC "given" item.

4.1. When There is an XP Boundary at the Onset of Wd2

First, let us consider forms where there is an XP boundary at the left edge of Word 2, i.e., forms that are categorized as Structure 1. There are four of those forms: AA1NN, AA1FG, UU1NN and UU1FG. We are interested in whether or not the edge tones (i.e. L% and H-) appear at the first two initial syllables of Word 2 (i.e. S21 and S22) respectively. If edge tones are present in all of those four kinds of forms, then we expect that the F0 value associated with S22 is significantly higher than that associated with S21 regardless of what the accent conditions and context conditions are.

Table I shows mean F0 values (Hz) of S21 [*yu*] and S22 [*nyuu*] and the result of 1-tailed paired t-tests between those two mean F0 values. According to the t-tests, mean F0 values of S22 turned out to be significantly higher than those of S21 regardless of who the speaker is and whatever the accent and context conditions are.

Table I: Mean F0 (Hz) of S21 & S22, and t-test results.

Spk	*Conditions*	*S21*	*S22*	*df*	*paired t-test p*
SK	AA1NN	200	328	4	< .001*
	AA1FG	191	236	8	< .001*
	UU1NN	236	328	4	< .001*
	UU1FG	276	289	4	< .008!
RO	AA1NN	170	252	8	< .001*
	AA1FG	186	220	8	< .001*
	UU1NN	169	242	8	< .001*
	UU1FG	226	236	7	< .001*
MR	AA1NN	196	291	8	< .001*
	AA1FG	192	216	9	< .001*
	UU1NN	197	271	7	< .001*
	UU1FG	241	258	10	< .001*

"*" means that the difference is significant at $p < .05/12 = .0042$ (Bonferroni adjustment for multiple procedure). "!" means that the difference is significant at $p < .01$ without adjustment.

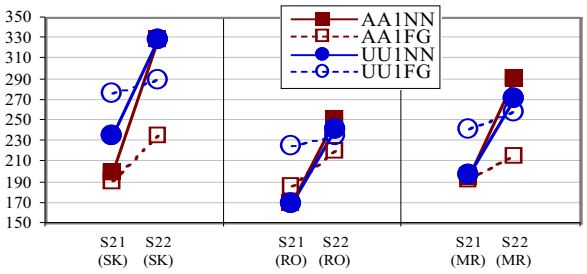


Figure 3: Mean F0 values (Hz) of S21 & S22.

- Fundamental Observation: «when there is an XP boundary»
  An F0 rise from S21 to S22 is present regardless of the accent and the context conditions.

Based on this result, I conclude that when there is an XP boundary at the left edge of Word 2, a L% tone and a H- phrasal tone are always present at the onset of Word 2, i.e. a MiP boundary is present. (In Section 4.2., we will observe cases where there is no XP boundary at the onset of Word 2. We will see in that section that the presence of an XP edge is responsible for a presence of a MiP boundary in a post-FOC part of an utterance.)

In addition to the observation made above, we find that there are differences in the magnitude and the height of the F0 rise depending on what the context conditions and accent conditions are. First of all, the magnitude of the F0 rise from S21 to S22 is far greater in Context NN (i.e., when both Word 1 and Word 2 are interpreted as new: AA1NN & UU1NN) than in Context FG (i.e., when Word 1 is a FOC and Word 2 is interpreted as given: AA1FG & UU1FG). The magnitude of the former is about 80 ~ 130 Hz while that of the latter is only about 10 ~ 45 Hz. (see Figure 3).

- Additional Observation 1: «Context Effect»
  The magnitude of F0 rise from S21 to S22:
  FG < NN

Furthermore, those smaller F0 rises from S21 to S22 in Context FG are realized in two different pitch levels depending on what the accent conditions are: the F0 rise of the accented AA1FG case is realized far lower than that of the unaccented UU1FG case. In contrast, the greater F0 rises observed in Context NN, i.e. the rise of the AA1NN case and that of the UU1NN case, are realized in the same pitch level regardless of the difference in their accent conditions.

- Additional Observation 2: «Accent * Context Interaction»
  The pitch level in which the F0 rise is realized:
  AA1FG < UU1FG,
  AA1NN = UU1NN

Those two additional observations are explained in the following way. In addition to the MiP boundary in all of those cases, there is a higher phrase boundary, i.e., a boundary of a Major Phrase (a.k.a. "Intermediate Phrase") or an Intonational Phrase, at the onset of Word 2 produced in Context NN. The MiP boundary, however, is the only phonological phrase boundary at the onset of Word 2 produced in Context FG.

In Tokyo Japanese, tones which coincide with the onset of a MajP bear more extreme values than those which do not ([6]). As a result, the F0 rise excursion size (i.e. the magnitude of the rise) in the NN cases is far greater than that in the FG cases.

In addition, tones that follow a pitch accent undergo *catathesis,* a post-accent downtrend, unless there is a MajP boundary intervening between the preceding pitch accent and those tones ([6], [7]). The reason that tones associated with Word 2 of the accented AA1FG case are realized lower than those of unaccented UU1FG case is because they follow a pitch accent of the preceding accented word and there is no MajP boundary between the preceding pitch accent and those tones. As a result, they undergo *catathesis*. On the other hand, tones associated with Word 2 of the unaccented UU1FG case do not undergo such a post-accent downtrend despite the lack of a MajP boundary at the onset of Word 2 because they are preceded by no pitch accent.

Such a contrast between the AA case and the UU case is neutralized in Context NN because there is a MajP boundary between Word1 and Word2: the presence of a MajP boundary blocks the application of *catathesis* even when Word 1 carries a pitch accent. That is why both the F0 rise of the accented AA1NN case and that of the unaccented UU1NN case are realized in the same pitch level (or register).

In summary, I conclude that when there is an XP boundary at the left edge of Word2, at least a MiP boundary is present at the onset of Word 2 regardless of what the accent and context conditions. Also, we found that the effect of "post-FOC giveness" was to reduce the size of the order of the phrase break from a higher phrase break (i.e. a MajP or IP break) to a lower MiP break.

4.2. When There is No XP Boundary at the Onset of Wd2

In this section, I will show the results obtained from cases with no XP boundary at the left edge of Word 2, i.e. forms that match Structure 2. Those are AA2NN, AA2FG, UU2NN and UU2FG. Mean F0 values (Hz) of S21 and that of S22 of each of those four forms were again compared by using a paired t-test. The result is shown in Table II and Figure 4.

In Context NN (i.e., AA2NN and UU2NN), where Word 1 and Word 2 are interpreted as new, there is always a significant F0 rise from S21 to S22 regardless of what the accent conditions are. When it comes to forms produced in Context FG (i.e. AA2FG and UU2FG), where Word 1 is a FOC and Word 2 is "given", things are different. For all the

three speakers, there is a small but significant F0 rise (15 Hz ~ 30 Hz) from S21 to S22 when those words are accented (i.e. AA2FG). However, there is no rise (SK and RO) or only a trivial rise of 2 Hz (MR) when those words are unaccented (i.e. UU2FG). This trivial F0 rise of MR turned out to be insignificant.

Table II: Mean F0 (Hz) of S21 & S22, t-test results.

Spk	*Conditions*	*S21*	*S22*	*df*	*paired t-test p*
SK	AA2NN	202	291	4	< .001*
	AA2FG	196	223	9	< .001*
	UU2NN	294	317	9	< .001*
	UU2FG	312	309	9	> .1
RO	AA2NN	163	236	8	< .001*
	AA2FG	185	201	9	< .004*
	UU2NN	183	239	9	< .001*
	UU2FG	255	252	17	< .001*
MR	AA2NN	181	242	10	< .001*
	AA2FG	185	202	10	< .001*
	UU2NN	221	259	11	< .001*
	UU2FG	275	277	13	> .05

"*" means that the difference between S21 and S22 is significant at $p < .05/12 = .0042$ (Bonferroni adjustment for multiple procedure.)

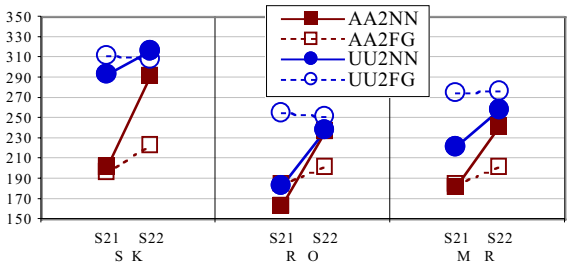


Figure 4: Mean F0 values (Hz) of S21 & S22.

- Observation «when there is no XP boundary»
  No F0 rise from S21 to S22: UU2FG.
  Significant F0 rise: UU2NN, AA2NN, AA2FG.

I interpret this outcome that L% and H- are always present, i.e. there is a MiP boundary, at the onset of Word 2 in the NN cases regardless of their accent conditions. Also, those edge tones and a MiP boundary are present at the onset of Word 2 even in the FG condition as long as both Word 1 and Word 2 are accented (i.e. AA2FG). However, no edge tones and no MiP boundary appear at the left edge of Word 2 in the FG condition when no accent is present (i.e. UU2FG).

Additional observation is that in the NN cases, there are contrasts between the AA case and the UU case despite the fact that both of them are associated with a significant F0 rise from S21 to S22. For example, for SK, the magnitude of the F0 rise in the accented AA2NN case is far greater than that in the unaccented UU2NN case. Also, for both SK and MR, the pitch level in which the F0 rise of the UU2NN case is realized is higher than that in which the rise of the AA2NN case is realized. Those contrasts still need to be explained and I will leave it for future studies.

5. Conclusions

Summarizing the main results obtained in the experiment, we found absence of a MiP boundary at the left edge of Word 2 only when (a) both Word 1 and Word 2 were unaccented, (b) there was no XP boundary at the left edge of Word 2, and (c) Word 2 was a post-FOC "given" item. When (i) both Word 1 and Word 2 are accented or (ii) there is an XP boundary at the onset of Word 2, at least a MiP boundary appears even at the onset of a post-FOC "given" word.

I propose that the reason for the presence of a MiP boundary at the onset of a post-FOC "given" word which coincides with a XP boundary is because there is a strong requirement in the grammar that every XP boundary should coincide with at least a MiP boundary. Satisfying this requirement is more important than satisfying other constraints which require post-FOC "given" items to be in the same MiP as the preceding FOC item. Also, the reason for the presence of a MiP boundary at the onset of a post-FOC "given" word when both the preceding FOC and the post-FOC word are accented is because there is a constraint in the grammar that requires every MiP to dominate at most one pitch accent. Again, satisfying this constraint is more important than satisfying the constraint which requires a FOC and a post-FOC "given" word to be wrapped in the same MiP. Therefore, only when those two important constraints are irrelevant, a MiP boundary is absent at the onset of a post-FOC "given" word.

In this way, absence or presence of a MiP boundary in a post-FOC part of an utterance is not arbitrary but well regulated by the grammar.

[This work was supported by Grant No. BCS-0004038 from the National Science Foundation.]

6. References

[1] Ghini, M., 1993. Phi-Formation in Italian: a New Proposal. *Toronto WPL* 12(2), 41-78.

[2] Hayes, B.; Lahiri, A. The Bengali Intonational Phonology. *Natural Language & Linguistic Theory*.

[3] Jun, S-A. 1996. *The Phonetics and Phonology of Korean Prosody: Intonational Phonology and Prosodic Structure*. New York: Garland..

[4] Kubozono, H. 1993. *The Organization of Japanese Prosody*. Tokyo: Kuroshio.

[5] Nagahara, H., 1994. *Phonological Phrasing in Japanese*. Ph.D. dissertation, UCLA.

[6] Pierrehumbert, J.B.; Beckman, M.E., 1988. *Japanese Tone Structure*. Cambridge: MIT Press.

[7] Poser, W., 1984. *Phonetics and Phonology of Tone and Intonation in Japanese*. Ph.D. dissertation, MIT.

[8] Selkirk, E. O., 1986. On Derived Domains in Sentence Phonology. *Phonology Yearbook* 3, 371-505.

[9] Selkirk, E. O.; Tateishi,K., 1991. Syntax and Downstep in Japanese. In *Interdisciplinary Approaches to Language: Essays in Honor of S.-Y. Kuroda*, Georgopolous, C; Ishihara, R. eds. Dordrecht: Kluwer.

[10] Truckenbrodt, H. 1996. Phonological Phrases: Their Relation to Syntax, Focus, and Prominence. Ph.D. dissertation, MIT.