The intonation of gapping and coordination in Japanese: Evidence for Intonational Phrase*

Shigeto Kawahara and Takahito Shinya

University of Massachusetts, Amherst

1. Introduction

The theory of prosodic phonology posits a number of levels in prosodic structure. The postulation of such levels is well motivated as they define the domains or loci of a wide variety of phonological processes such as stress assignment, tonal downstep, boundary tone association, assimilation (spreading), dissimilation, and resyllabification, to name a few (Hayes & Lahiri 1991; Jun 1998; Nespor & Vogel 1986; Selkirk 1986, 2001 among many others). Moreover, ample evidence for prosodic layering has been provided from studies on the domain-initial strengthening effect, where articulatory strength gets greater at higher prosodic structure boundaries (Cho & Keating 2001; Fougeron & Keating 1997; Hayashi et al. 1999; Hsu & Jun 1998; Keating et al. 2003; Onaka 2003). One example of prosodic hierarchy is illustrated in (1) (e.g. Selkirk 1986, 1995, 2000, 2001, to appear):

* This work is in part supported by NSF grant BCS-0004038. We thank Kathryn Flack, Mako Hirotani, Sun-Ah Jun, John Kingston, Lisa Selkirk and Hubert Truckenbrodt for their comments. Any remaining errors are ours.

Shigeto Kawahara (ed.), UMOP 30, 91-127.

(1) Utterance
|
Intonational Phrase
|
Major Phrase (a.k.a. Intermediate Phrase/ Phonological Phrase)
|
Minor Phrase (a.k.a. Accentual Phrase)
|
Prosodic Word
|
Foot
|
Syllable
|
Mora

Of particular relevance to our current study is the level *Intonational Phrase* (henceforth IP), a level above Major Phrase (MaP) and below Utterance. Although the intonation of Japanese has been relatively extensively investigated, no evidence has been adduced for the existence of IP. A Minor Phrase (MiP) is signaled by initial lowering caused by a sequences of boundary tones, and it also defines a domain in which maximally one lexical accent is allowed. Major Phrase (MaP) defines a domain of downstep (Poser 1984). However, little has been known about prosodic levels higher than MaP. For instance, Pierrehumbert and Beckman (1988) posit no prosodic level between MaP (= their Intermediate Phrase) and Utterance. In Venditti's (1995, in press) JToBI model, Pierrehumbert and Beckman's Intermediate Phrase and Utterance are merged into a single prosodic level called *Intonation Phrase*, resulting in only one level above Minor Phrase (MiP) (Venditti's *Intonation Phrase* should not be confused with our IP). In short, no evidence has been shown for the existence of a level above MaP and below Utterance – IP in (1) has not been previously motivated in the Japanese prosody.

On the other hand, IP has been shown to play a role in many other languages: Chichewâ (Kanerva 1990: 146-147); English (Beckman & Pierrehumbert 1986; Nespor & Vogel 1986: Chapter 7; Selkirk to appear); German (Baumann et al. 2001; Féry & Hartmann 2004; Truckenbrodt in press); Greek (Arvaniti & Baltazani in press); Hungarian (Vogel & Kenesei 1987); the Tuscan dialect of Italian (Nespor & Vogel 1986); Kinande (Hyman 1990: 112-121); Kinyambo (Bickmore 1990: 8); Luganda (Hyman 1990: 111-112); Spanish (Nespor & Vogel 1986) and others (see also Jun's typology (in press)). One question that immediately arises is whether IP is motivated in the phonology of Japanese. We answer positively to this question.

In many cases cited here, IP usually corresponds to a syntactic clause. Furthermore, Selkirk (to appear) explicitly argues that there is a universal relationship between IP and a syntactic clause. If this is indeed the case, as implied by evidence from a wide variety of languages, then we expect to find evidence for IP in multiple-clause sentences in Japanese. For this reason, gapping and coordination are the targets of the current study. Our investigation is thus informed and driven by a theoretical consideration that a syntactic clause might universally correspond to IP.

With this theoretical question in mind, we show that IP is indeed necessary to account for some aspects of multi-clausal intonational phonology in Japanese. Our empirical evidence primarily comes from the intonation of gapping, and its comparison with coordination. Gapping, as shown in (2) and (3), minimally contrasts with coordination in that the verbs in non-final clauses are unpronounced:[1],[2]

(2) **Gapping** (Subj Obj ~~Verb~~, Subj Obj ~~Verb~~, and Subj Obj Verb)

Murasugi-wa	*namauni-o*	~~*moritsuke*~~*,*	*Munakata-wa*
Murasugi-TOP	sea urchin-ACC		Munakata-TOP

mamemochi-o	~~*moritsuke*~~*,*	*Morimura-wa*	*aemono-o*	*moritsuketa.*
bean rice cake-ACC		Morimura-TOP	aemono-ACC	dished up

'Murasugi dished up sea urchin, Munakara bean rice cake, and Murimura aemono.'

(3) **Coordination** (Subj Obj Verb, Subj Obj Verb, and Subj Obj Verb)

Murasugi-wa	*namauni-o*	*moritsuke,*	*Munakata-wa*
Murasugi-TOP	sea urchin-ACC	dished up	Munakata-TOP

mamemochi-o	*moritsuke,*	*Morimura-wa*	*aemono-o*	*moritsuketa.*
bean rice cake-ACC	dished up	Morimura-TOP	aemono-ACC	dished up

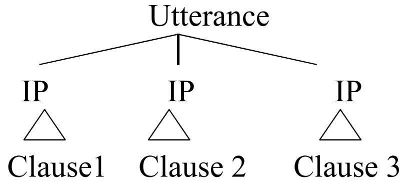
'Murasugi dished up sea urchin, Munakara dished up bean rice cake, and Murimura dished up aemono.'

We argue that to account for the intonation of gapping and coordination, it is necessary to posit IP, a level above MaP and below Utterance. Specifically, we argue that sentences like (2) and (3) are parsed into a prosodic structure where each clause corresponds to IP and the entire sentence is incorporated to Utterance, as depicted in (4a). We point out that IP and Utterance are characterized by different sets of phonetic properties. We further argue that other possible structures such as those depicted in (4b) and (4c) are inadequate as a model of prosodic structure for multiple-clause sentences like gapping and coordination.

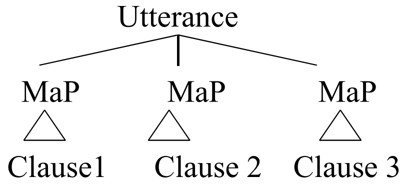
[1] No agreement has been made for the appropriate syntactic account of gapping. Some authors argue that what seems gapping is in fact Right-Node-Raising involving Across-the-Board movement (Kuno 1978; Saito 1987; Kasai & Takahashi 2001), while Abe and Hoshi (1998) argue that gapping is base generated, and gapped verbs are filled in at LF by copying. Several Korean scholars (Sohn 1994; Kim 1997), based on a similar construction in Korean, argue that such constructions involve deletion under identity at PF. We are not concerned with the syntactic aspects of gapping. What is important to us is that verbs in non-final clauses remain unpronounced.

[2] The abbreviations used in this paper are as follows: ACC = accusative, DAT= dative, GEN = genitive, LOC = locative, TOP = topic.

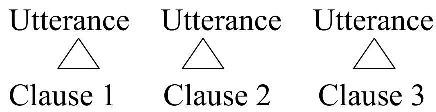
(4a)



(4b)



(4c)



The rest of this paper proceeds as follows. In §2, we lay out the methods for the production experiment which investigated gapping and coordination. §3 compares the intonation of gapping to that of coordination, which shows the existence of domain-final lowering, definable only in terms of IP. Meanwhile we argue against a structure like (4b) for a proper characterization of multi-clause sentences. In §4 we compare the intonation of three clauses within gapping and coordination sentences, and provide several pieces of evidence that there are processes only definable in terms of Utterance, which must be distinct from IP. A structure like (4c) is rejected in this section; at the same time, we identify several properties of Utterance, a level that has hitherto not received much attention in the literature. §5 discusses more general issues within the current context of intonational phonology. The final section concludes the paper.

2. Methods

2.1 Speakers

Four female native speakers (J, N, R, and Y) of Japanese were recruited at the University of Massachusetts Amherst. They were all paid for their participation. They were in their early twenties (N, Y) or in their early thirties (J, R) at the time of recording. All except Speaker R were from the Kanto area where Tokyo Japanese is spoken. Speaker R was from Shizuoka, located at approximately 200 km to the west of Tokyo, but her speech was similar enough to Tokyo Japanese for the purpose of the current experiment.

2.2 Experimental materials

The experimental materials consisted of sets of coordination and gapping sentences, as exemplified by the sentences (2) and (3), in addition to other types of sentences summarized in (5). Words with obstruents (especially voiceless obstruents) were avoided as much as possible because they cause perturbations of F0.

All words in the experimental sentences were accented on the second mora, and they were four moras long (the tonal contour thus being LHLL). The basic syntactic structure for gapping and coordination was “S-O-(V), S-O-(V), S-O-V”. Moreover, since it is known that constituent branching affects intonational patterns (e.g. Bickmore 1990; Kubozono 1993; Selkirk 2000; Shinya to appear), we systematically varied the length of the subjects and objects by changing the number of words that comprised them. The "short" subjects/objects consisted of a single word while the "long" subjects/objects consisted of two words. Three combinations of short/long and subjects/objects were tested: SS (short S and short O), SL (short S and long O), LS (long S and short O).[3] Two versions with different lexical items were created for each of the sentence types. Each clause in gapping and coordination was separated by a comma, as is usually required by the standard orthographic convention.

In addition to these three types of sentences, we included other kinds of sentences. First, in addition to the three conditions (SS, SL, LS), coordination and gapping were paired in the dative condition whose syntactic structure was "S-I(ndirect)O-D(irect)O-(V), IO-DO-(V), IO-DO-V". All constituents consisted of a single word. We furthermore included two kinds of single-clause constructions, which served as fillers. First, we had predicative (copula) sentences which consisted of a subject and either a "short" or "long" predicates followed by *-da* (copula). The short predicate consisted of a single noun followed by *–da,* and the long predicate was comprised of two words (Noun-Gen (*no)* Noun), again followed by *-da*. Schematically, the syntactic structure was thus "S (N-Gen) N-*da*." Finally, we included short and long intransitive sentences whose predicates had either only an intransitive verb or a locative phrase followed by an intransitive verb, which were the short and long conditions, respectively; the sentences had the structure S-(Loc)-V. The six types of experimental sentences are summarized in (5), with their schematic syntactic structures.

(5) Experimental sentence sets

a. SS (coordination vs. gapping)
[[N-Top] [[N-Acc] (V)]], [[N-Top] [[N-Acc] (V)]], [[N-Top] [[N-Acc] V]][4]

b. SL (coordination vs. gapping)
[[N-Top] [[N-Gen N-Acc] (V)]], [[N-Top] [[N-Gen N-Acc] (V)]], [[N-Top] [[N-Gen N-Acc] V]].

c. LS (coordination vs. gapping)
[[N-Gen N-Top] [N-Acc (V)]], [[N-Gen N-Top] [N-Acc (V)]], [[N-Gen N-Top] [N-Acc V]].

d. Dative (coordination vs. gapping)
[[N-Top] [[N-Dat] [N-Acc (V)]]], [[[N-Dat] [N-Acc (V)]]], [[[N-Dat] [N-Acc V]]].

[3] The other possible combination, LL (long subject and long object), was not included in the experimental conditions. We assumed that the intonation pattern in that condition would be analogues to the SS condition because both subject and object are of the same length, and the sentence thus has a symmetrical structure.

[4] The topic marker (*-wa*), not the nominative case marker (*-ga*), was used to mark subjects. Generally, the topic marker sounds more natural than the nominative case marker when the item that the marker attaches to is given in the discourse. Coordination and gapping function in such a way that a subject and an object are selected from given sets of possible subjects and objects, and are matched up within a clause. Thus, the sentences sound more natural with the topic marker *-wa*.

e. Predicative (short vs. long predicates)
[N-Top] [(N-no) N-da]

f. Intransitive (short vs. long predicates)
[N-Top] [(N-Loc) V]

2.3 Procedure

Each speaker had two recording sessions. The minimal pair sentences like (2) and (3) were recorded on different days, lest the speakers notice the contrasts. Predicative and intransitive sentences served as fillers so as not to have the speakers realize that the experiment was about gapping and coordination. Eight additional fillers were added.

Their speech was recorded to CDs in a sound-attenuated booth in the phonetics laboratory at the University of Massachusetts, Amherst. The experimental sentences were written on index cards in the usual Japanese orthography, which is a mixture of the *hiragana* syllabary and Chinese characters. The speakers were first asked to read through all the sentences silently to familiarize themselves with the material. They were then asked to read them aloud at a normal speech rate as naturally as possible. They read through the cards six times. The stimuli order was randomized between repetitions. When the speakers stumbled in the middle of a sentence, they were asked to read it again.

2.4 Data analysis

The recorded materials were transferred to a pc with an 11,025 Hz sampling rate and 16 bit quantization level, and then submitted to F0 measurement using *PitchWorks* (Scicon R&D).

The guidelines for F0 measurement were as follows. A typical F0 contour indicating measurement points is represented in Figure 1, which shows the intonation pattern of one clause of a coordination sentence. The arrows indicate where we measured F0. MiP boundaries and mora boundaries are denoted by solid lines and dashed lines, respectively. Accent location is indicated by an apostrophe on the gloss. The F0 of an accented word is characterized by a L% boundary tone at its left edge and a H*+L tone on the accented mora (Pierrehumbert & Beckman 1988). We measured F0 peaks and valleys that appeared in each MiP, assuming that the peaks represent a H* tone and valleys, the boundary L% (for measurements of initial L%H on verbs in coordination, see §4.2).

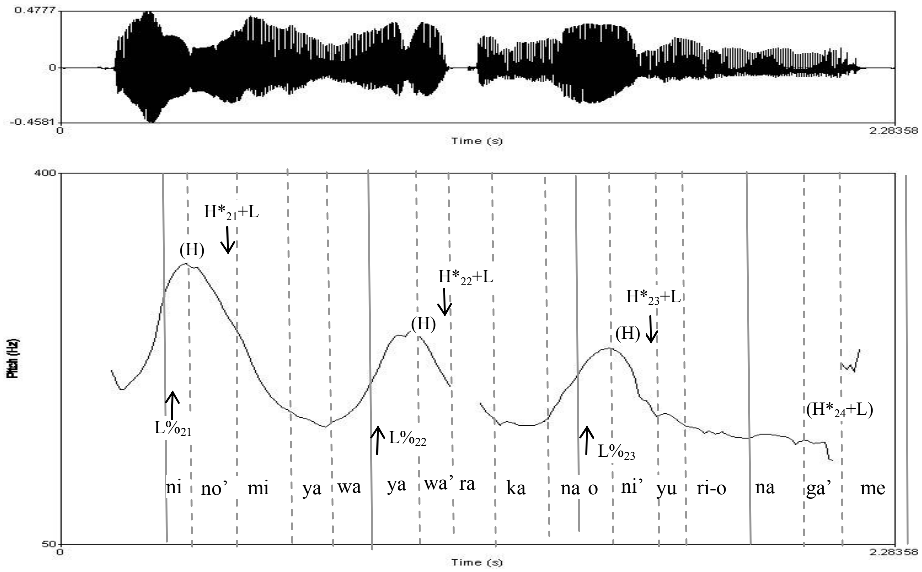


Figure 1. Illustration of the F0 measurement points. The utterance was extracted from the second clause in a coordination sentence spoken by Speaker R, which is *Nino'miya-wa yawa'rakana oni'yuri-o naga'me*... 'Ninomiya gazed at the soft lilies…'

In theory, a phrasal H is present between a L% boundary tone and a H*+L, but it is not clearly seen because in words with accent on the second mora, the phrasal H is realized on the second mora together with the H*+L tone (potential phrasal H tones are represented in brackets).

In what follows, to refer to specific tones we adopt the labeling convention T_{ij}, where *i* stands for a tonal number within a clause, and *j* stands for the clause number in the sentence. For instance, the second H* tone in the first clause is referred to by H_{21}.

For the statistical analyses in the main sections of this paper, we normalized the raw F0 values using the formula in (6) adopted from Truckenbrodt (2004):

(6) Transformed value = (Original value – $Mean_L$)/($Mean_H$-$Mean_L$)

where $Mean_T$ is the speaker-specific mean of tone T.

The mean value of H_{11} for each speaker was used as the Mean H value, and $Mean_L$ is defined by non-final L tone (= L_{23} for the SS and dative conditions and L_{33} for SL and LS conditions). These tones define the speakers' highest and lowest tones. Thus, what this transformation essentially does is to define a pitch range for each speaker by $Mean_H$-$Mean_L$, and relativize each tonal value to the tonal range.

This normalization has the following virtues: since overall little inter-speaker variation was found (with one exception; see §4.4), with normalization it is possible to pool all speakers' data. This considerably simplifies analyses and exposition of the data. Second, by pooling all speakers' data, a less drastic post-hoc α-level adjustment (e.g. Bonferroni adjustment) is necessary for multiple comparisons in statistical analyses. If we were to analyze the data separately for each individual, and if we tried to make a three-way comparison across three clauses (e.g. comparing T_{1j}-T_{2j}, T_{2j}-T_{3j}, and T_{1j}-T_{3j}), for example, then α-level would have needed to be adjusted to 0.05/(3*4)=0.004. With normalization, we can avoid such a drastic adjustment.

For statistical analyses, when we compare data points from a single sentence, a repeated-measures analysis is used. This includes a repeated-measures ANOVA when an independent variable has more than 2 levels and a paired *t*-test when an independent variable has 2 levels. A within-subject comparison like this is made whenever possible, because it reduces variability across each token, and hence it usually has more power. However, an independent samples *t*-test was used when we make comparisons of different sentences (e.g. comparison of gapping and coordination). In the case of *t*-tests, we used two-tailed tests to be conservative.

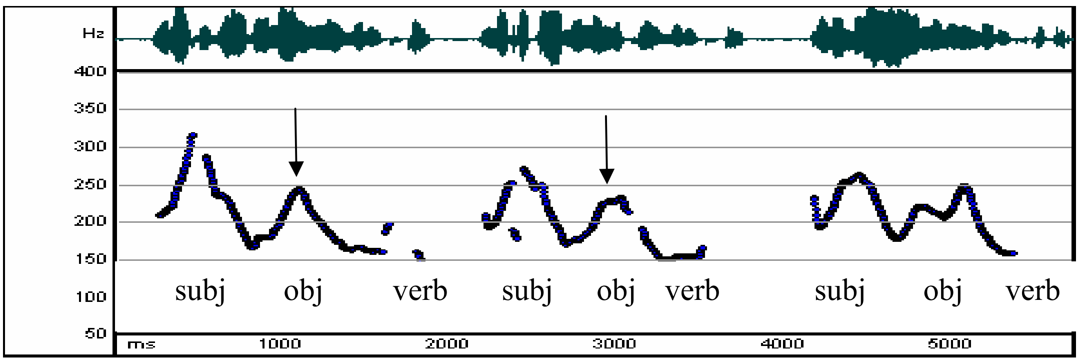
3. Comparison between gapping and coordination

Based on the results of the experiment, this section argues for the existence of IP in the intonational phonology of Japanese, which in the case at hand corresponds to a syntactic clause in gapping and coordination. In §3.1 we report on a lowering phenomenon which takes place at clause-final positions. In §3.2 we show that from the perspective of prosodic phonology, IP is the only level that can adequately characterize the positions in which lowering occurs. In §3.3 we further motivate the difference between IP and MaP.

3.1 Observation

A novel finding in our experiment is that for each type of gapping sentence, the clause-final H* peaks in non-final clauses systematically appear lower than the corresponding H* peaks in the corresponding coordination sentence. An illustrative pair of pitch tracks is given below in Figure 2, where the clause-final accent H* peaks (i.e. the H* in the objects) in the non-final clauses are lower in gapping (shown by the thick arrows) than in coordination (thin arrows), despite that these tones are hosted by the same lexical items.

a. Coordination



b. Gapping

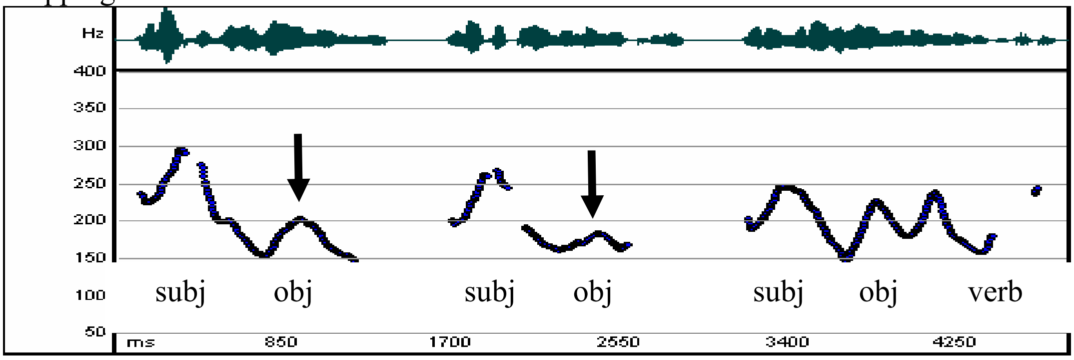


Figure 2. Representative F0 track of coordination(a) and gapping (b) uttered by Speaker Y.

To show that this lowering is not a sporadic phenomenon, Figure 3 plots the mean values of relevant tonal contours for each clause in a simple gapping and coordination sentence from Speaker Y's data. Shown in the graphs are the values of the first three tones for each clause (H_{1j}, L_{1j}, H_{2j}). The crucial observation is that in the non-final gapping clauses, the second accent H* tones (= H_{21} and H_{22}) are realized in a lower range than the corresponding H* tones in the coordination sentences. The difference is neutralized in the final clause where the verb is not elided (H_{23}).

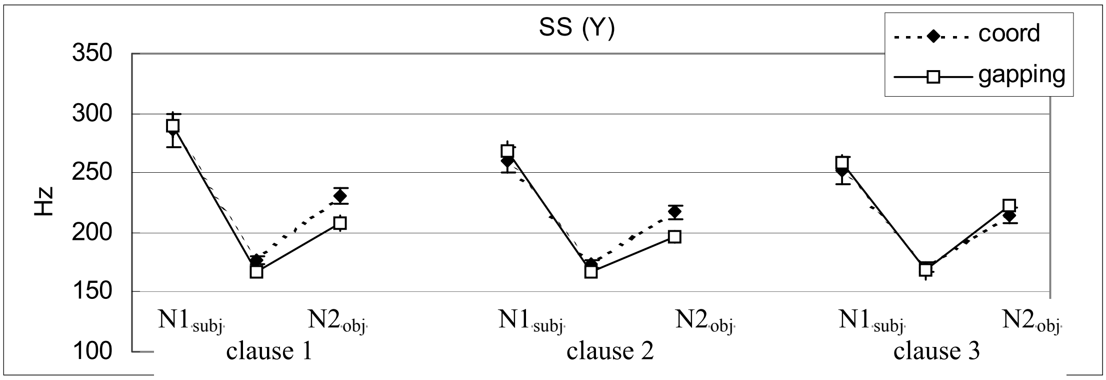
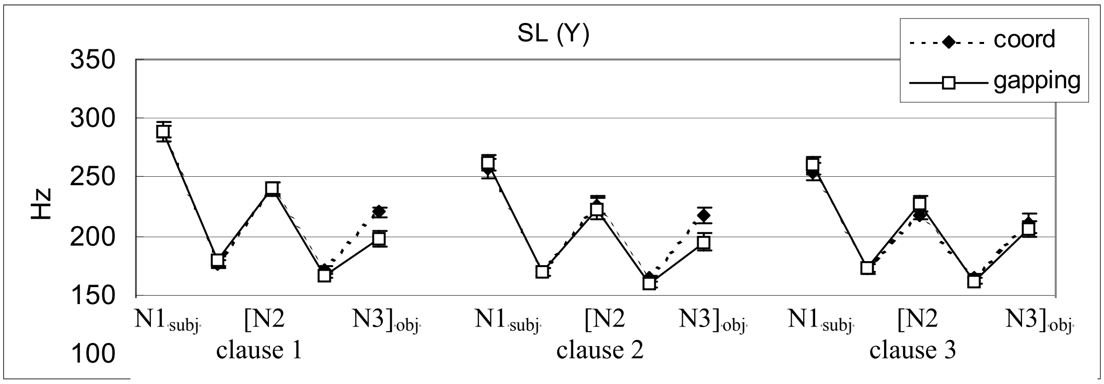
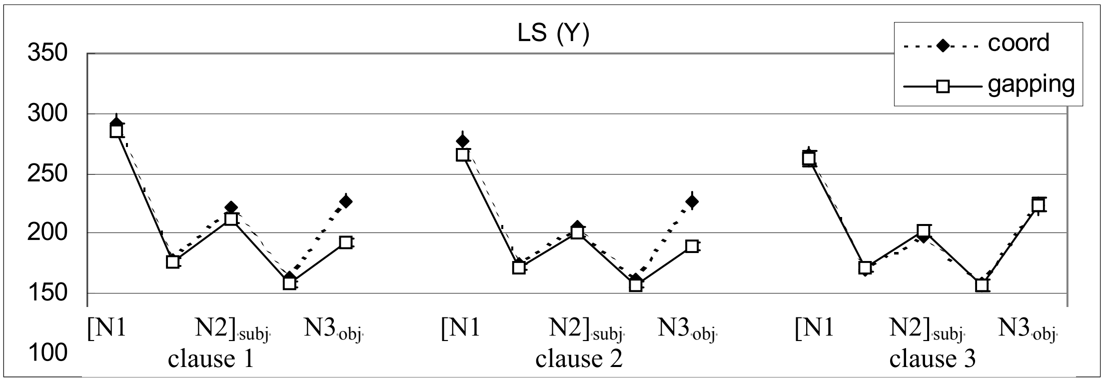


Figure 3. Means of H_1, L_1, H_2 for each clause: data from Speaker Y. Error bars indicate 95% confidence intervals.

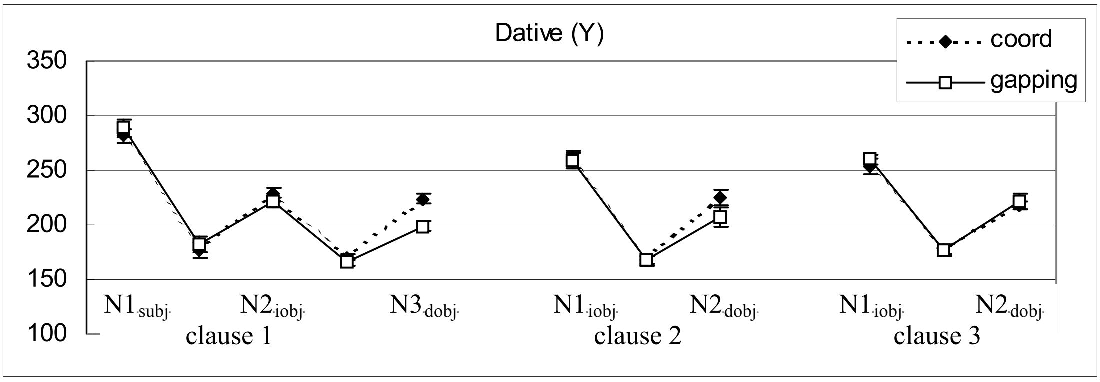
The same tendency is observed in the SL, LS, and dative conditions. The generalization is that *all clause-final* H* tones in gapping are systematically realized lower than those in the corresponding coordination sentences. This is illustrated in Figure 4. This shows that final lowering is observed no matter what the clause-internal structure of gapping is: SS, SL, LS and dative gapping behave all alike in that H*s in clause-final positions appear lower – branching turns out to play no role in this regard.



a



b



c

Figure 4. Means of penultimate and final accent H tones and boundary L tones in each clause in the SL (a), LS (b), and dative (c) conditions: data from Speaker Y.

The observed difference between gapping and coordination is a quite general one observed across all the speakers. We calculated the differences between the H* of our interest and the immediately preceding one for all conditions. The result shows that the differences are systematically larger in the gapping condition than in the coordination condition, suggesting that the H*s in gapping are indeed lower. The graphs that summarize the results are given in Figure 5.

For a statistical analysis, we conducted an ANOVA on F0 differences between the penultimate and final H with two independent variables, CLAUSE (1st, 2nd, 3rd) and TYPE (coordination and gapping). There were significant main effects for both variables (CLAUSE: $F(2, 392)=50.551$, $p<.0001$, TYPE: $F(1, 196)=179.965$, $p<.0001$) and for the interaction ($F(2,392)=104.814$, $p<.0001$) as well.

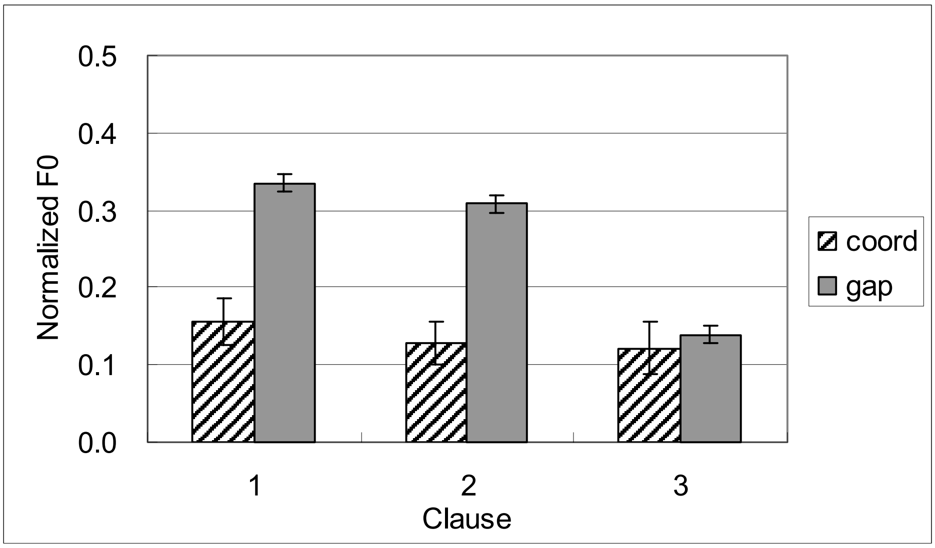
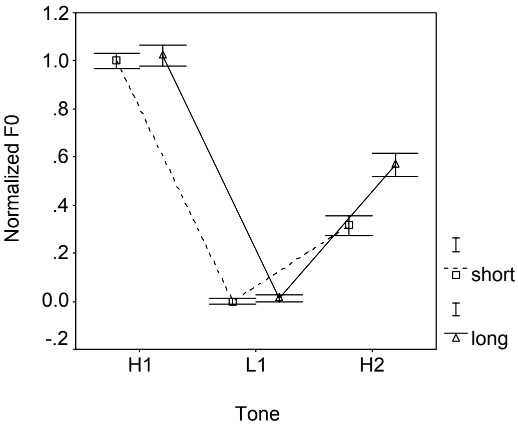


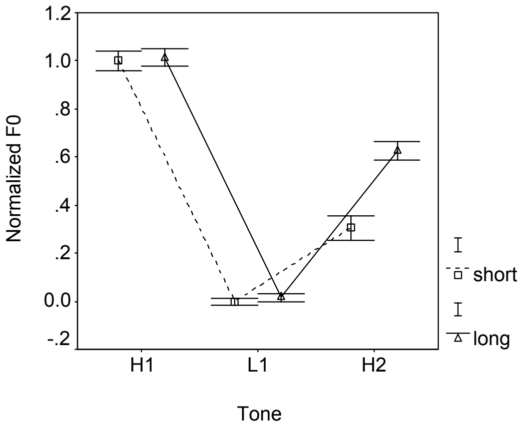
Figure 5. Means of the normalized F0 differences between penultimate and final peak values for coordination and gapping for all speakers. Error bars represent 95 % confidence intervals.

Of the most importance here is that there is a significant effect of TYPE, supporting a categorical difference between the intonations of gapping and coordination. The significance of the interaction suggests that the difference observed in the first and the second clauses is neutralized in the third clause. The results of post-hoc multiple comparison tests confirm this: the peak differences between coordination and gapping are highly reliable in the first and the second clause (C1: $t(393)=8.073$, $p<.0001$, C2: $t(393)=8.610$, $p<.0001$), but the difference is neutralized in the final clause where there are no syntactic differences ($t(393)=.722$, $p=.471$). Further, the interaction effect disappeared when we reran an ANOVA only with the data from the first and second clause ($F(1, 196)<1$. $p=.917$).

The overall results thus suggest that in non-final clauses, the final H* tones in gapping systematically appear in a lower range than the corresponding H* tones in coordination, even though these two H*s are hosted by identical lexical items. Such lowering is not construction-specific. The pattern obtained in the gapping condition also emerges in the predicative sentences and intransitive sentences: when an item hosting H* is located in clause-final positions, the values of H* appear lower compared to those of H* in non-final positions. Figure 6 shows the schematic F0 patterns of the three tones for the predicative and intransitive sentences; namely, H*L of the subject (H_1 and L_1) and the second H (H_2), which is final in short conditions and non-final in long conditions:



a. Predicative



b. Intransitive

Figure 6. Normalized means of the first three tones for the predicative and intransitive constructions. H_1 corresponds to the peak of the subject (both for predicatives and intransitives), L_2 to the initial valley at the predicate word (for predicatives) or the intransitive verb (for intransitives), and H_2 to peak of the predicate word or the intransitive verb.

As is clear from Figure 6, the H_2 values are smaller in the short condition where H_2 is clause-final. A statistical analysis shows that the F0 descent from H_1 to H_2 is significantly larger in the short condition than in the long condition, for both the predicative and intransitive sentences. Independent-sample *t*-tests suggest that for both of the constructions, the difference is quite significant (predicative: $t(49)=9.254$, $p<.0001$, intransitive: $t(97)=11.236$, $p<.0001$). This shows that the final-lowering effect is seen not only in gapping sentences but in clause-final positions in general (see §4.2 for evidence that such lowering manifests itself in coordination sentences via lowering of H+L peaks on verbs).

3.2 Proposal

This section provides an analysis of how the difference between gapping and coordination arises. In §3.1 we have shown that clause-final H*s in gapping and other constructions are lowered. Viewed from the perspective of prosodic phonology, then, there must be a level in which this lowering is definable. In other words, the generalization is that H* which is domain-final at *some* level are lowered. We argue that IP is the only domain which can serve this purpose.

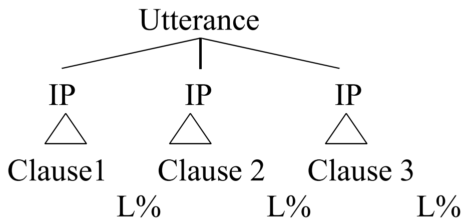
The domain for final lowering cannot be Utterance, because the end of each clause in a gapping sentence does not correspond to the end of a whole utterance. It cannot be MaP either, as such lowering is unmotivated at MaP-final positions (these arguments are developed in more detail in the following subsection). We thus postulate a level above MaP and below Utterance with respect to which final lowering is definable. We call this level *Intonational Phrase* (hereafter IP). IP, in the case at hand, corresponds to a syntactic clause. Further, all IPs are incorporated into a higher prosodic level, Utterance, which corresponds to an entire sentence in syntax (see §4 for more on

Utterances).

This idea is in line with the observations made by several authors that IP corresponds to the so called “comma intonation” in other languages, and is usually followed by a pause (Bing 1979; Nespor & Vogel 1986; Potts 2003; Selkirk to appear). This generalization is also true in the Japanese case at hand: each clause is orthographically separated by a comma and phonologically followed by a pause. In particular, that IP corresponds to each conjunct in gapping and coordination is compatible with Selkirk’s (to appear) recent claim that IP corresponds to a syntactic Comma Phrase, a projection headed by a [+comma] feature; a Comma Phrase can consist of syntactic clauses or parentheticals, non-restrictive clauses, appositives and others (see Potts 2003 for semantic contribution of [+comma] feature).

To account for the lowering observed in gapping, we posit that there is a L% boundary tone associated with the right edge of IP, as in (7). This L% causes lowering of the IP-final Hs by way of tonal coarticulation (cf. Xu 1994, 1997; see §5.2).

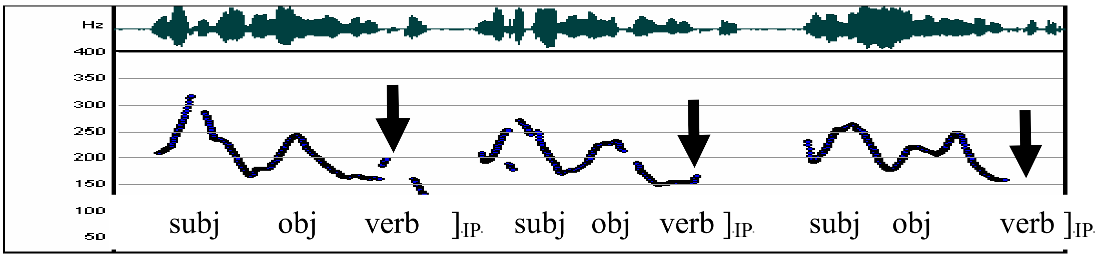
(7)



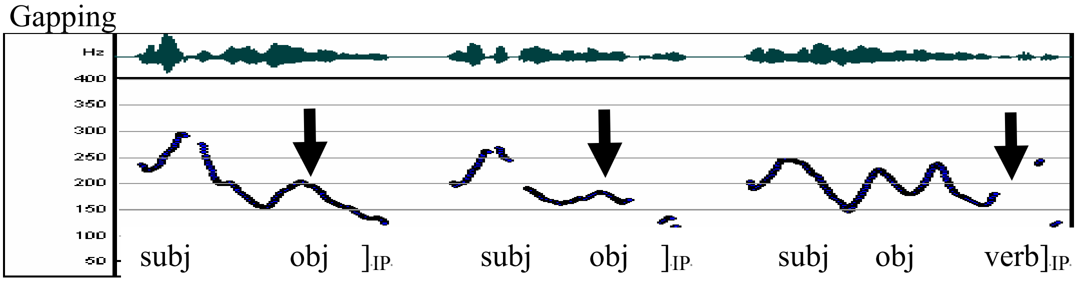
With this, H* on the coordination verbs is subject to lowering in coordination (8a); on the other hand, it is H* on the final objects that is subject to lowering in gapping (8b) (see §4.2 for evidence that H* on verbs are lowered in coordination). This explains why the H* tones of the final objects in gapping appear lowered compared to those of coordination, as the objects in coordination are not subject to this lowering.

(8)

a. Coordination



b. Gapping



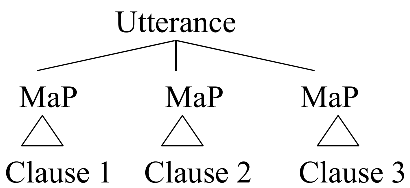
3.3 More on MaP-IP distinction

In the proposed prosodic structure illustrated in (7), each clause in gapping corresponds to IP. One of its phonological cues is a boundary L% tone, which is not motivated MaP-finally. This subsection discusses further differences between MaP and IP.

First, in addition to final lowering, another cue that signals IP is a pause at its end, which is nearly invariably observed at clause-final positions across all of our tokens, as seen, for example, in (8ab). This almost obligatory pause again does not usually appear at the end of MaP.

These observations show that we cannot assume a structure like below where each clause corresponds to MaP:

(9)



In addition to the evidence discussed above, there are at least three cues that distinguish IP from MaP, each of which is discussed below.

3.3.1 Creakiness

First, vowels become creaky IP-finally, as shown in Figure 7. The waveform and spectrogram of the first clause in a coordination sentence in the SS condition (Speaker J).[5] A waveform is also given to show that irregular glottal pulses, which are a characteristic of creaky voiced vowels, are observed on final vowels.

[5] Creaky voice is cross-linguistically often associated with L tones (see e.g. Gordon & Ladefoged 2001). This does not necessarily mean, however, that creaky voice is an automatic correlate of a boundary L%, as other low tones (such as the trailing +L tone in the accent H*+L tone) do not cause creaky voice. Therefore, creaky voice should be considered an independent phonetic cue that signals IP boundaries, rather than an automatic correlate of L%.

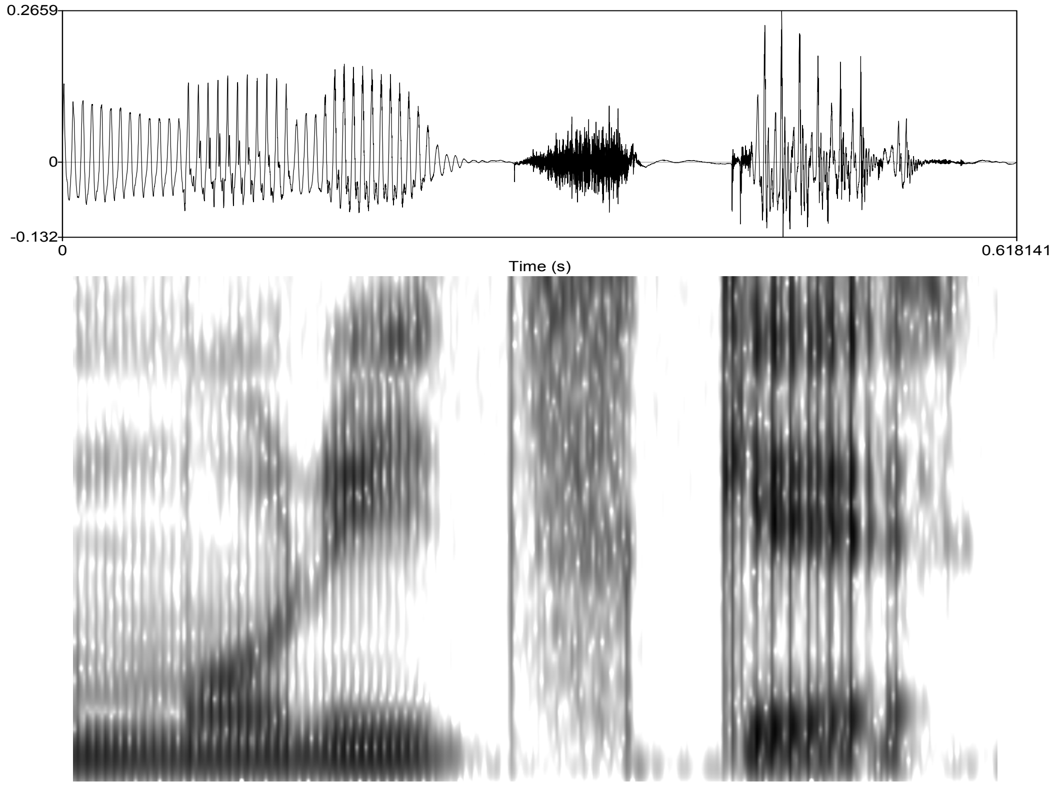


Figure 7. The waveform and spectrogram of the first clause in a coordination sentence in the SS condition (Speaker J).

This provides further evidence that positions where lowering appears cannot be equated with a MaP-boundary because creakiness is not observed at the end of MaP. To confirm this generalization, we counted the frequency of creaky vowels in (i) subject particles (and in the dative condition, in the dative particle *ni*) and (ii) sentence-final particles in gapping and coordination sentences. The first position represents non-clause-final MaP-final position because it is followed by the right edge of XP boundary (Selkirk & Tateishi 1991). The second position is IP-final position. Since we do not control for a vowel quality in these two positions, a quantitative analysis on spectrum slices was impossible. Instead, we relied on auditory impressions with the help of some acoustic cues. Vowels are judged creaky if they show an irregular waveform as well as excitation of energy in higher formants. Sometimes only a later portion of a vowel is creaky, in which case it was judged "semi-creaky."

Speaker J	MaP-final	IP-final	Speaker N	MaP-final	IP-final
Creaky	0	263	Creaky	2	184
Semi-creaky	0	13	Semi-creaky	2	57
Non-creaky	288	12	Non-creaky	308	71
Speaker R	MaP-final	IP-final	**Speaker Y**	MaP-final	IP-final
Creaky	0	151	Creaky	0	279
Semi-creaky	0	117	Semi-creaky	0	8
Non-creaky	306	38	Non-creaky	288	1

Table 1. Distribution of creaky vowels in MaP-final and IP-final position for each speaker.

The results are summarized in Table 1, which shows that creaky vowels rarely or never appear MaP-finally. On the other hand, creaky vowels are very common IP-finally, especially for Speakers J and Y.

To check the reliability of this identification of creaky vowels, four phonetically-trained native speakers of Japanese, who were naïve to the purpose of the experiment, were recruited; they were asked to judge the creakiness of vowels in 40 sentences selected at random, which amounted to one tenth of the whole data. The transcribers were asked to judge creakiness based on their auditory impression with the visual aid of irregularity of wave forms as well as excitation of energy in higher formants. They were told to classify a vowel as "non-creaky" if its entire portion is a modal voice, "creaky" if the entire portion is creaky, and "semi-creaky" if only a later portion is creaky. The results are that there is a fairly reliable consistency in the judgment of creakiness. There was fairly large inter-transcriber variability in the distinction between "creaky" and "semi-creaky vowels," presumably because the transcribers interpreted "only a later portion" differently. However, if we abstract away from the difference between "creaky" and "semi-creaky," essentially treating vowels as creaky if they are at least partially creaky, then the percentage of tokens for which all transcribers (including the four recruited transcribers and the two authors) agreed upon was 94.83% for the vowels in MaP-final positions, and 83% for the vowels in IP-final positions.

3.3.2 Pitch reset and initial rises

The next property that distinguishes IP from MaP is the degree of pitch reset. Given two successive H tones, we compared an F0 difference across a MaP boundary and an F0 difference across a clause boundary. If our hypothesis is correct in positing an IP boundary between clauses, the prediction is that the pitch reset is stronger across a clause boundary compared to a MaP boundary, as a higher prosodic edge induces more robust pitch resetting (e.g. Ladd 1988).[6] To quantitatively test this prediction, a comparison was made using coordination sentences in LS and dative conditions.[7] Specifically, we

[6] Thanks to Hubert Truckenbrodt for pointing this out.
[7] The other conditions were not suitable for this comparison because the Utterance-initial H*, where its F0 is boosted by domain-initial strengthening (§4.1), is involved in the calculation of the between-clause F0 differences.

compared F0 differences in two environments, (i) the difference between H_{21} and H_{31} (within-clause difference across a MaP boundary) and (ii) the difference between H_{31} and H_{12} (between-clause difference across an IP boundary). This is illustrated in Figure 8 below. Note that in the first environment, the two H* tones are separated by a MaP boundary, as they are separated by a VP boundary (Selkirk & Tateishi 1991) – note that no downstep is observed for the second H.

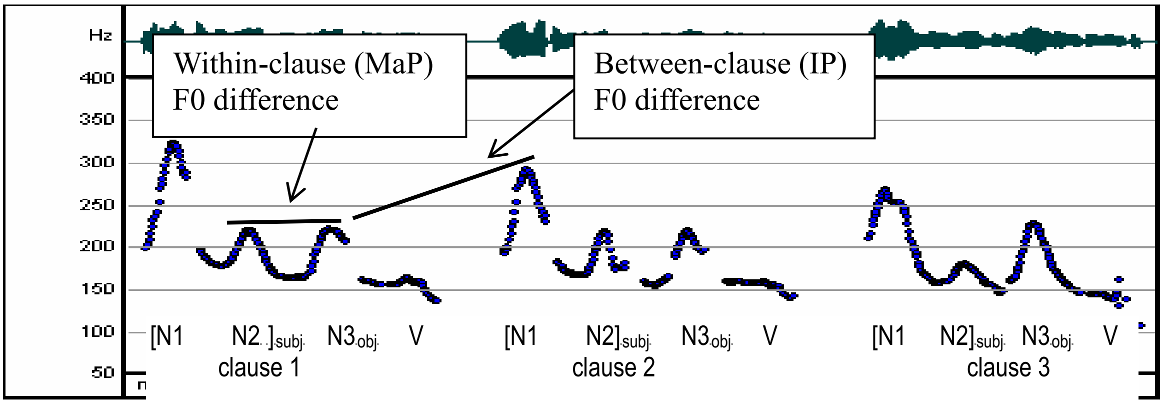


Figure 8. The F0 difference between two H* tones in two conditions: (i) within-clause condition and (ii) between-clause condition. The pitch track is taken from a coordination sentence in the LS condition uttered by Speaker R.

As observed, the between-clause F0 difference is much larger than the within-clause difference. If the former H* tones were separated merely by a MaP boundary, i.e. if the clause-initial rises were due to a pitch range reset at a MaP boundary, then we should expect no differences between the two environments. Figure 9 shows the results of a statistical analysis on all relevant tokens. The difference in normalized F0 is much larger in the between-clause condition than in the within-clause condition. A paired *t*-test shows that the difference is significant ($t(197)=5.363$, $p<.0001$). This provides further evidence that the prosodic boundary appearing between the clauses is not MaP but one which shows a larger pitch reset.

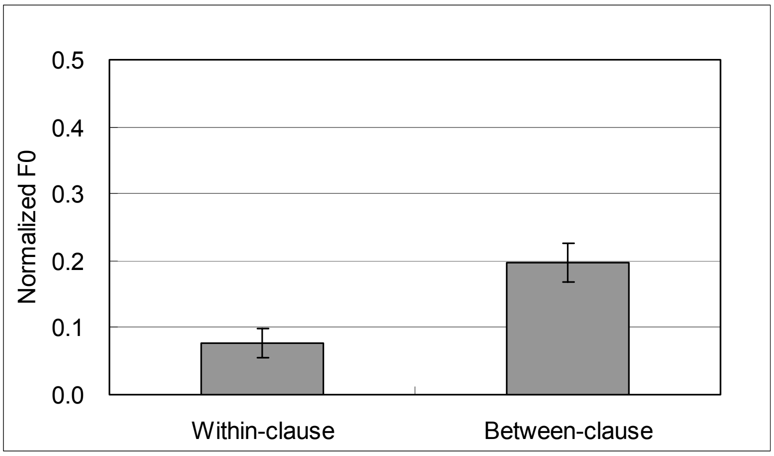


Figure 9. Means of the normalized F0 peak difference in the within-clause

condition and the between-clause condition (see Figure 8 for illustration).

3.3.3 Initial rises

There is yet another measure that distinguishes an IP boundary from a MaP boundary. Several previous works have shown that initial rises are cross-linguistically larger at IP edges than at MaP edges (Ladd 1988, 1990; Selkirk to appear; Truckenbrodt 2002). If a clause boundary is also an IP boundary as we have argued, then clause-initial rises should be larger than VP-initial rises, which coincide with MaP boundaries (Selkirk & Tateishi 1991).

To test this prediction, the F0 rises at the beginning of the VP in the SL and dative conditions (H_{21}–L_{21})[8] were compared with the clause-initial rises of the second clause[9] of the same sentences (H_{12}–L_{12}). Consider Figure 10. The former is a representative of MaP-initial rises, which were compared to the clause-initial rises.

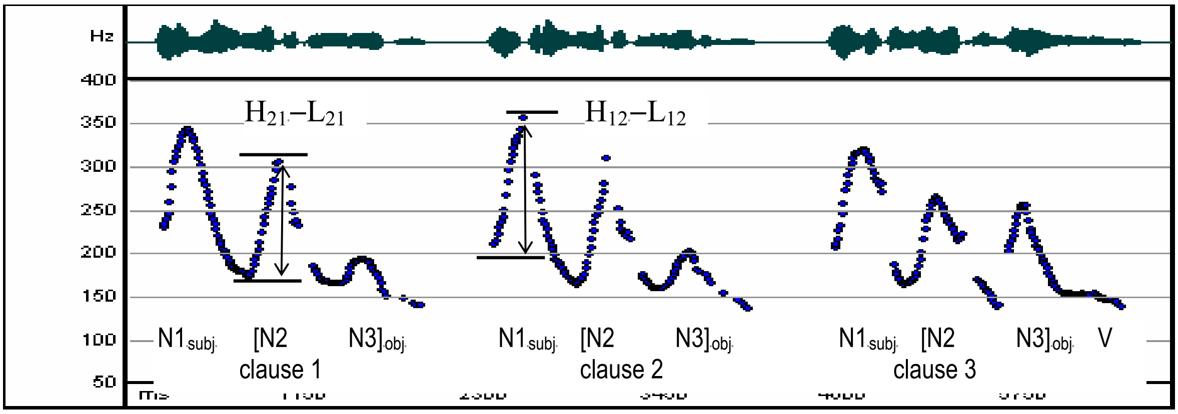


Figure 10. The clause-initial F0 rise (=IP-initial rise) compared to that the VP-initial rise (=MaP-initial rise). The pitch track is of a gapping sentence in the SL condition uttered by speaker R.

The results of the comparison are given in Figure 11. The F0 rises at the left edge of the second clause (=IP-initial rises) are greater than those at the left edge of the VP (=MaP-initial rises). This observation is supported by a paired samples *t*-test ($t(198)=2.890$, $p=.004$).

[8] In the SL and dative conditions, the second words in a clause coincide with the beginning of the VP, and are not in a clause-final position which is subject to the lowering effect in gapping. In the other conditions, the left edge of the VP is clause-final in the gapping, and therefore is not suitable for the comparison.
[9] Utterance-initial rises are avoided as they are subject to Utterance-initial boost (see §4.1).

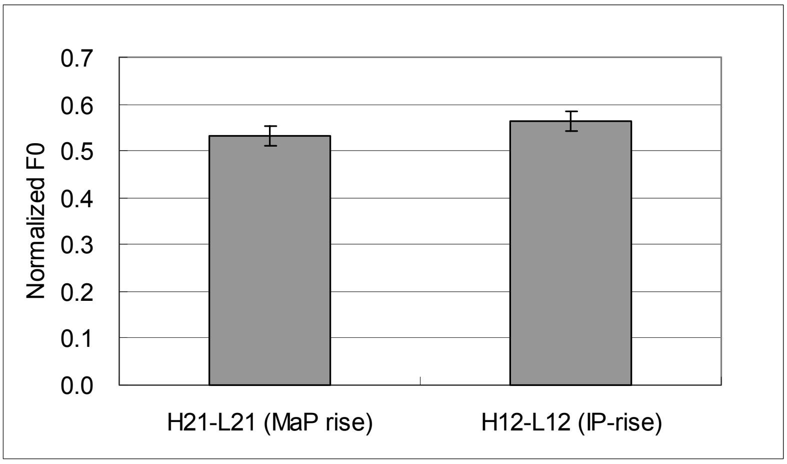


Figure 11. Means of normalized VP-initial F0 rises (H_{21}–L_{21}, MaP-initial rise) compared to the clause-initial rises (H_{12}-L_{12}, IP-initial rise).

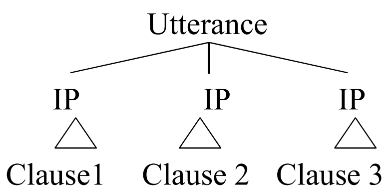
Note that these two kinds of initial rises appear in different positions in an utterance: the VP-initial rises occur in the first clause while the clause-initial rises appear in the second clause, which is later in the utterance. Given the effect of declination, where a speaker's pitch range declines over the course of an utterance (see §4.1), it is expected that the initial rises of the second clause are smaller than VP-initial rises. The fact that the second rises are higher than initial rises even given a declination effect strongly suggests that clause-initial rises are different from MaP-initial rise. Rather they are IP-initial rises which are inherently larger than MaP-initial rises.

4. IP-Utterance distinction

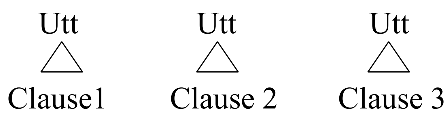
In §3, we have motivated the existence of IP by showing that systematic lowering is observed in domain-final positions definable only in terms of IP. We adduced several additional pieces of evidence for the proposed structure (4a), which is repeated below, where an IP corresponds to a syntactic clause while an Utterance corresponds to an entire syntactic sentence.

This section compares clauses within a sentence, motivating the distinction between IP and Utterance. An alternative analysis that we address in this section is shown below as (10), which considers each clause as an independent and separate Utterance.

(4a) Proposed structure (repeated)



(10) Alternative structure



This alternative deserves serious attention because, at first glance, each gapping clause has a strong pitch reset at its beginning (see Figure 8), and there are fairly long pauses between clauses. For these reasons, it is necessary to entertain (10) as a serious alternative representation for gapping and coordination.

This approach that posits that each clause is an Utterance predicts that all the clauses behave alike in terms of prosodic patterns, because each clause constitutes a separate Utterance. We show that this prediction is not borne out. Initial and final clauses show unique characteristics, and such patterns can be explained only if each clause is further incorporated in a higher prosodic category, as in (4a).[10] While rejecting (10), we also identify several properties of an Utterance, a level above IP in the proposed structure, which is the cause of the non-homogeneity across clauses in gapping and coordination sentences.

4.1 Initial Hs, initial rises, and declination

First, the behavior of clause-initial Hs is of interest. It has been noticed in the previous literature that higher prosodic domains are signaled by higher initial rises (Ladd 1988, 1990; Selkirk to appear; Truckenbrodt 2002). We have seen that IP-initial rises are larger than MaP initial rises in Japanese (see Figure 11). If three clauses in gapping and coordination are incorporated in a higher prosodic level as in (4a), then what we predict is that clause-initial Hs have a higher pitch in the first clause than in the second and third clauses.

This prediction is borne out. Figure 12 shows the comparison of initial rises in three clauses based on the data of the gapping and coordination pooled together.

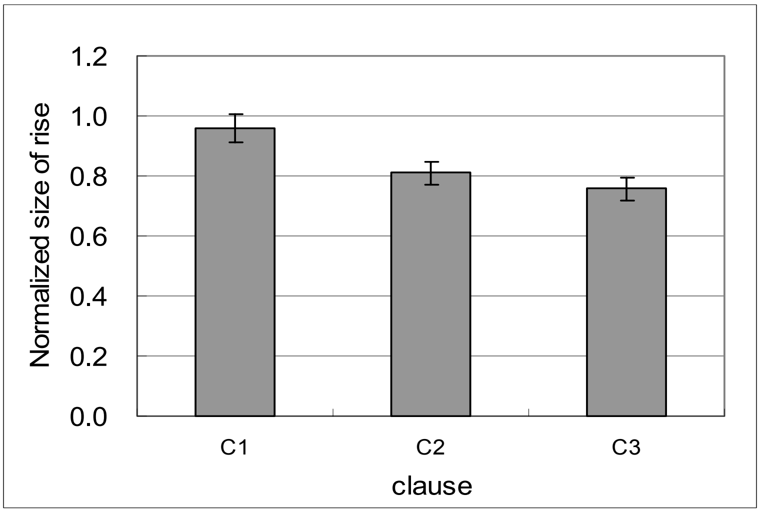


Figure 12. Normalized means of initial rises in each clause.

The size of initial rise is largest in the first clause; the differences between the second and the third clauses are small. A repeated-measures ANOVA on the values of initial rises with CLAUSE as the independent variable reveals that there is a significant main effect

[10] One variant of (10) is a structure in which each Utterance is incorporated into higher recursive Utterance. It is hard to come by evidence that distinguishes such a proposal from ours, as such an approach regards our IP as "Utterance" and our Utterance as "higher Utterance." The crucial point is that there are two levels above MaP, one that incorporates each clause, and the other that incorporates an entire sentence.

($F(2, 788)=189.439$, $p<.0001$). The post-hoc multiple comparison tests show that the initial rises in the three clauses significantly differ from one another (C1 vs. C2: $t(394)=13.731$, $p<.0001$, C1 vs. C3: $t(394)=17.356$. $p<.0001$, C2 vs. C3: $t(394)=4.985$, $p<.0001$). Even though the difference between the second clause and the third clause is statistically significant, it is much smaller than the difference between the first clause and the second or the third clause. We conclude from this that the first clauses have a special status in that they show particularly large initial rises. The difference between the second and the third clause can be attributed to declination, to which we now turn our attention.

Figure 13 plots the mean of clause-initial H peaks as well as clause-initial L peaks. First focusing on the patterns of H*s, two generalizations can be made: (i) the values of H* generally decline from the first clause to the third clause but (ii) the slope between the first clause and the second clause is steeper than the slope between the second and the third clause.

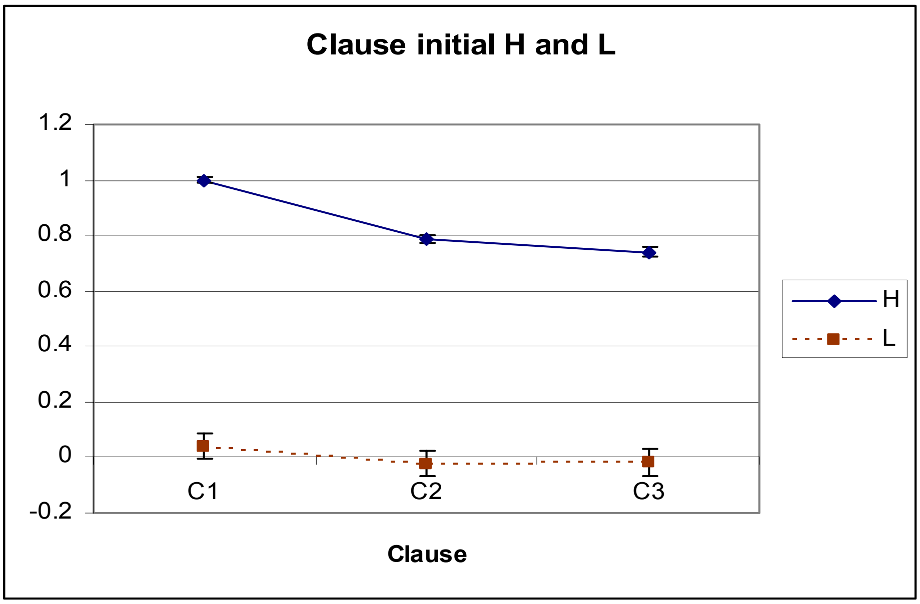


Figure 13. Normalized means of clause-initial H* and L% tones.

A statistical analysis based on paired *t*-tests supports these observations. As columns (a) and (b) in Table 2 show, the H* tones are significantly lower in the second clause (H_{12}) than in the first clause (H_{11}), and H_{13} is similarly lower than H_{12}. To examine whether the F0 slopes from the first to the second H* tones (H_{11}-H_{12}) are steeper than those from the second to the third H* tones (H_{12}-H_{13}), the differences between H_{11} and H_{12} and those between H_{12} and H_{13} were calculated and submitted to paired samples *t*-tests. The results are summarized in Table 2, column (c). The difference between the descent from the H_{11} to H_{12} and that from H_{12} to H_{13} is significant such that H_{11}-H_{12} is categorically larger than H_{12}-H_{13}.

a	b	c
H_{11} vs. H_{12}	H_{12} vs. H_{13}	H_{11}-H_{12} vs. H_{12}-H_{13}
$t(394)=26.667$, $p<.0001$	$t(394)=5.962$, $p<.0001$	$t(394)=13.741$, $p<.0001$

Table 2. Results of paired *t*-tests on F0 values between H_{11} and H_{12}, between H_{12}

and H_{13}, and the difference between H_{11}-H_{12} and H_{12}-H_{13}.

The general pattern of realizing accent H* tones, we argue, follows from two factors. There is a general declination effect, whose domain is Utterance. It is well-known that in many languages, F0 gradually declines over the course of an Utterance (Fujisaki & Hirose 1984; Liberman & Pierrehumbert 1984; Maeda 1976; Pierrehumbert 1980; Poser 1984; 'tHart & Cohen 1973; Thorsen 1980 among others). This explains the decrease in pitch between the second and the third clauses. This alone, however, does not explain why the difference between H_{11} and H_{12} is larger than that between H_{12} and H_{13}. Thus, in addition to declination, we argue that Utterance-initial H*s are boosted due to a domain-edge strengthening effect (see the references cited above), and as a consequence the differences between the first clause and the second clause are significantly larger than the differences between the second and the third clauses. This Utterance-initial boosting is perhaps responsible for larger initial rises in the first clauses that we observed above.

The observations made in this subsection provide further evidence for an Utterance that contains the three clauses in gapping and coordination. First of all, Utterance defines the domain of declination, as the H values steadily decline from H_{11} to H_{13}. Second, its left boundary exhibits a domain-edge strengthening effect, such that H_{11} values result in particularly large initial rises. Note that these effects would not be expected under a representation like (10). In such a model, there is no level that defines a domain of declination; neither can it capture the fact that sentence-initial Hs are boosted.

Turning our attention to the patterns of L tones, as shown above in Figure 13, what should be first noted is that the slope is much less steep than the corresponding H* tones. We can also see that the L tone in the first clause is somewhat higher than the other L% tones.

We carried out paired samples *t*-tests to compare the adjacent L% tones (L_{11} vs. L_{12} and L_{12} vs. L_{13}), and the differences between the first and second L% tones (L_{11}-L_{12}) with the differences between the second and third L% tones (L_{12}-L_{13}). The results show that L_{11} is significantly higher than L_{12} and L_{13}, and no difference was found between L_{12} and L_{13}. The slope between L_{11} and L_{12} is significantly higher than the slope between L_{12} and L_{13}. These suggest that just as in the H tone pattern seen above, L tones show a boosting effect at the beginning of an Utterance. The fact that there is no difference between the second and third clauses suggests that L tones are not subject to declination, confirming the view advanced by Ladd (1990, 1993) and Pierrehumbert & Beckman (1988).

a	b	c
L_{11} vs. L_{12}	L_{12} vs. L_{13}	L_{11}-L_{12} vs. L_{12}-L_{13}
$t(394)=6.928, p<.0001$	$t(394)=-.801, p<.424$	$t(394)=5.446, p<.0001$

Table 3. Results of paired *t*-tests on F0 differences between the first and the second L% tones (L_{11}-L_{12}) and those between the second and the third tones (L_{12}-L_{13}), and the difference between L_{11}-L_{12} vs L_{12}-L_{13}.

4.2 Initial rises on verbs

The next piece of evidence for our proposed structure comes from the size of initial rises on verbs in coordination sentences. Under our account, verbal rises in coordination constructions undergo final lowering. Given this, again, under the model in (10), it is predicted that the amount of final lowering is consistent across all the clauses. On the other hand, in the proposed model in (7), it is possible that final lowering is strongest in Utterance-final position, just as initial boosting of Hs is strongest in Utterance-final position. We show that this is indeed the case.

We measured pitch values of the first and the second mora of the verbs in each clause in coordination. In theory, we should observe a rise in this position due to a LH% sequence associated with a MiP, but because of downstep and domain-final lowering effects, rises in these positions were not necessarily visible. So we did not measure the L and H peaks; what was measured instead was the pitch at the steady state of the first and the second vowels. As noted in §3.3, IP-final vowels are creaky, so it was sometimes impossible to measure the F0; the results in this section are based on those tokens whose pitch contour was visible. Speaker J consistently showed heavy creakiness at final positions, and hence her data were not used here.

The results show that the initial rises on verbs in coordination are smallest in the third clause, and as such are qualitatively different from those found in the first and second clauses. The results are summarized in Figure 14:

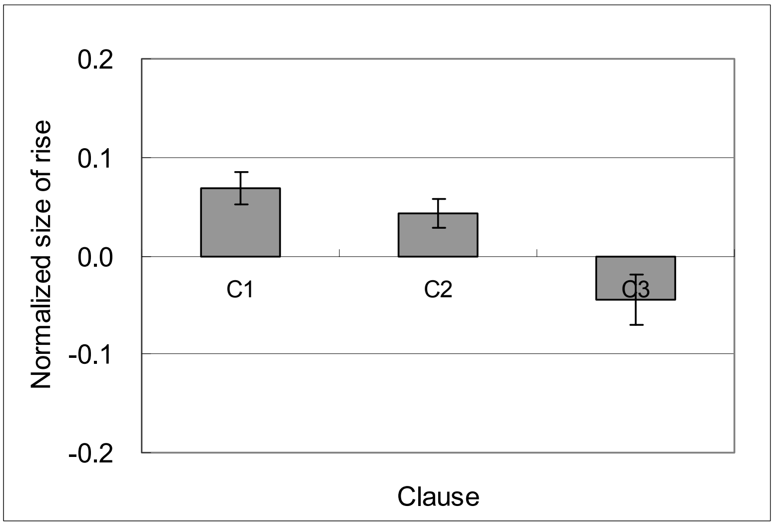


Figure 14. Normalized means of verbal rises in each clause in coordination sentences.

The third clause is markedly different from the first and the second clauses. The values are positive in the first and the second clauses, showing that there are rises in these positions, though they are very small. On the other hand, the values in the third clause are negative: this indicates that there is no rise, and instead the F0 declines from the first mora to the second mora, despite the presence of a phonological LH% tone.

A repeated-measures ANOVA on the normalized values of verbal rise (H-L) with CLAUSE as the independent variable reveals that there is a significant main effect ($F(2, 82)=54.170$, $p<.0001$). Post-hoc multiple comparison tests show that the verbal rises in the three different positions are significantly different from one another (C1 vs. C2: $t(41)=3.833$, $p<.0001$, C2 vs. C3: $t(41)=8.959$, $p<.0001$, C1 vs. C3: $t(41)=6.082$, $p<.0001$).

We maintain that the categorical difference in the final clause as opposed to the first and second clause, is an effect of the final lowering that we identified in §3; this effect is enhanced in the third clauses due to a domain-edge strengthening effect specific to Utterance-final positions. A boundary L% tone associated with IP significantly lowers the pitch of the second mora of the verb in Utterance-final position, practically obliterating the pitch accent. This finding cannot be explained if we assumed that each clause in coordination constitutes separate Utterance, as the final clauses exhibit unique behavior.

4.3 Utterance-final H tone

As final evidence for our structure in (4a), we show that Utterance has a H tone associated with its right edge.[11] This H tone's docking site is subject to inter-speaker variability, but its presence is motivated for all the speakers. The presence of this H tone, which appears Utterance-finally but not IP-finally, again provides evidence for a distinction between Utterance and IP.

The evidence for this H tone is most clearly observed in the behavior of Speakers N and Y, for whom the H tone is associated with the accusative case particle *o* of the preverbal object. This manifests itself as an extra H peak, shown in Figure 15 where there are two H peaks on the final-object. Figure 16 illustrates the alignment of this extra H, which coincides with the accusative case particle -*o*.

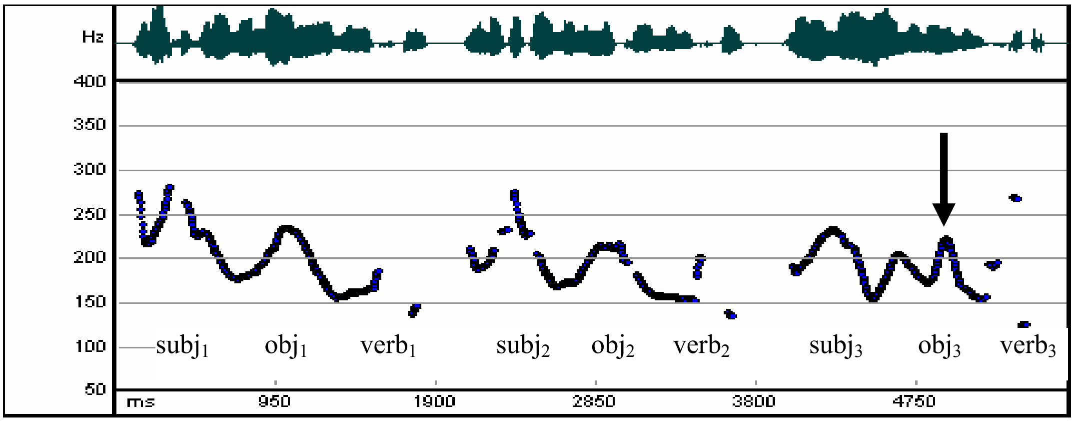


Figure 15. Representative pitch track of a coordination sentence uttered by Speaker Y, illustrating a sentence-final H tone docking onto a case particle on the final object (marked with an arrow). The sentence is *Mura'sugi-wa*$_{\text{subj1}}$ *nama'uni-o*$_{\text{obj1}}$ *mori'tsuke*$_{\text{verb1}}$, *Muna'kata-wa*$_{\text{ubj2}}$ *mame'mochi-o*$_{\text{obj2}}$ *mori'tsuke*$_{\text{verb2}}$, *Mori'mura-wa*$_{\text{subj3}}$ *ae'mono-o*$_{\text{obj3}}$ *mori'tsuketa*$_{\text{verb3}}$. 'Murasugi put raw sea urchin on a dish, Munakata put bean rice cake on a dish and Morimura put mixed salad on a dish.'

[11] This H tone is not specific to gapping and coordination sentences, but is seen in single-clause sentences if the sentence is long. In the experiment we had single-clause sentences with SOOV structure which were long enough to be comparable to the gapping and coordination sentences. The speakers showed this final H tone in such control sentences, although its occurrence was optional, as opposed to obligatory in gapping and coordination. Such optionality of this H in these sentences requires further research.

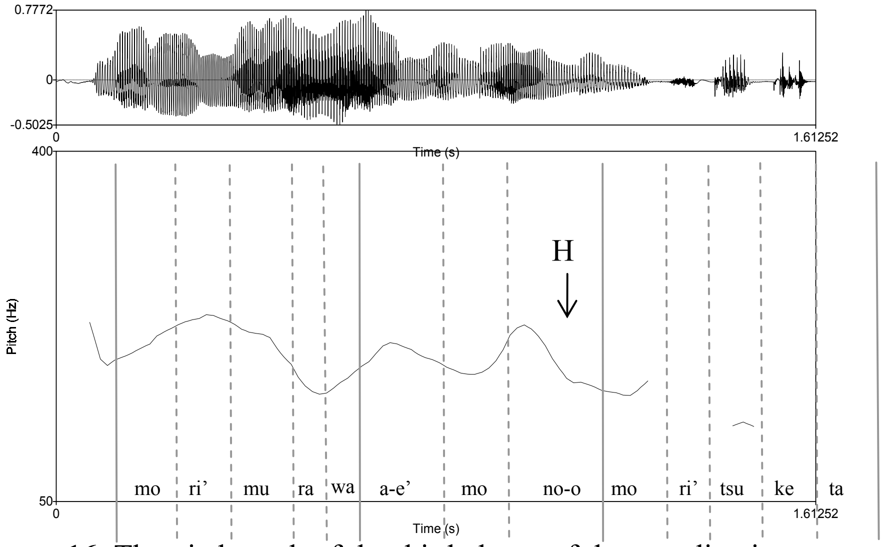


Figure 16. The pitch track of the third clause of the coordination sentence shown in Figure 15, illustrating the utterance-final H tone docking onto the case particle *-o*.

For Speakers J and R on the other hand, the H docks onto the accented mora of the word which immediately precedes the verb, boosting the preverbal rise further. This is evidenced by the fact that the peaks of the final object in the final clause have higher value than the corresponding H tone in the second clause. A representative example, taken from the SS condition of Speaker J, is shown below in Figure 17:

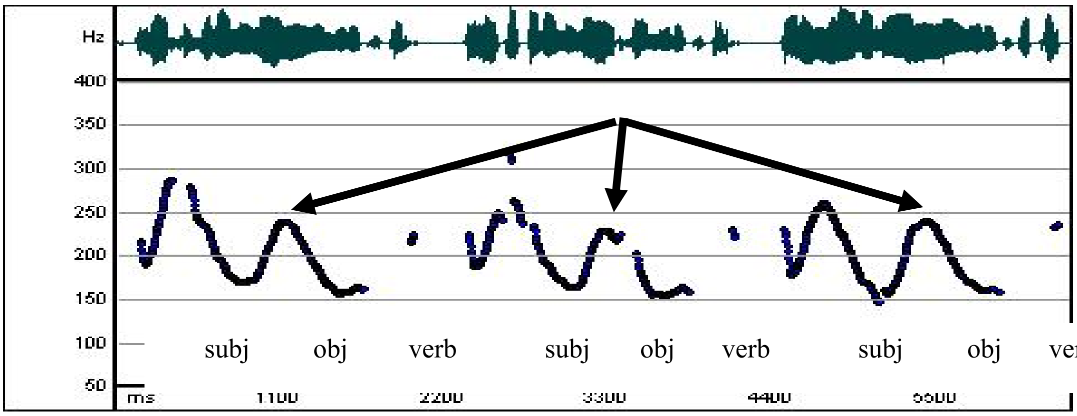


Figure 17. Representative pitch track of a coordination sentence uttered by Speaker J, illustrating boosting of the rise of final objects (marked with arrows). The sentence is *Mura'sugi-wa*$_{\text{subj1}}$ *nama'uni-o*$_{\text{obj1}}$ *mori'tsuke*$_{\text{verb1}}$, *Muna'kata-wa* $_{\text{ubj2}}$ *mame'mochi-o*$_{\text{obj2}}$ *mori'tsuke*$_{\text{verb2}}$, *Mori'mura-wa* $_{\text{subj3}}$ *ae'mono-o*$_{\text{obj3}}$ *mori'tsuketa*$_{\text{verb3}}$. 'Murasugi put raw sea urchin on a dish, Munakata put bean rice cake on a dish and Morimura put mixed salad on a dish.'

The object H peak in the final clause is higher than the corresponding object H peak in the second clause, and it is as high as the object H peak in the first clause. We interpret this high F0 peak in the final clause as having been boosted by an Utterance-final H tone. Recall that there is a general declination effect where H tone values generally decline over the course of an Utterance. The fact that the third H peak is higher than the second H peak is contrary to what is expected given declination, and it thus suggests that an extra boosting mechanism is at work.

To show that, for Speakers J and R, there is indeed a boosting effect on the object peak of the final clause (as opposed to Speakers N and Y who lack this effect), Figure 18 shows each speaker's preverbal H* for the three clauses in coordination. We can see that the third H tone values are larger than the second H tone values for Speaker J and decline to some extent for Speaker R, but less so than Speakers N and Y.

A mixed-design ANOVA is performed with SPEAKER (between-subject) and CLAUSE (within-subject) as the independent variables to assess these observations. A significant effect was found for both SPEAKER ($F(2, 193) = 71.093$, $p<.0001$) and CLAUSE ($F(2, 386) = 59.514$, $p<.0001$) as well as interaction ($F(6, 386) = 23.770$, $p<.0001$).

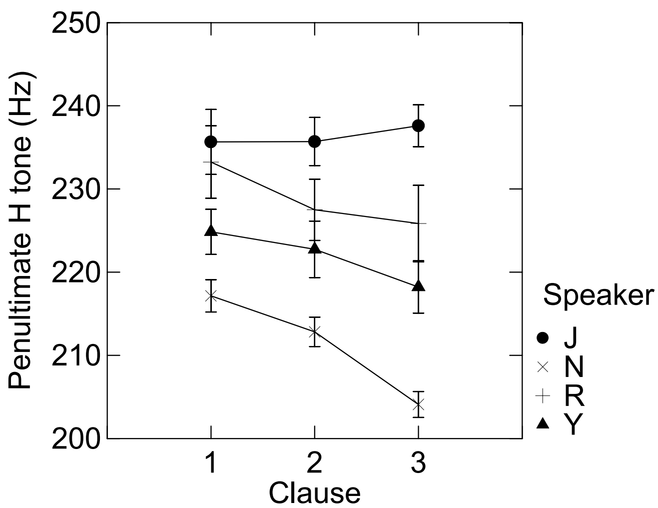


Figure 18. Means of the F0 peaks of the final objects for each clause in coordination.

Speaker	C1 vs. C3	C2 vs. C3
J	$p < .026$	$p < .0001$* (C2 < C3)
R	$p < .0005$*	$p < .012$
N	$p < .0001$*	$p < .0001$* (C2 > C3)
Y	$p < .0001$*	$p < .002$* (C2 > C3)

Table 4. Results of post-hoc multiple comparison tests on the values of Hs in the penultimate MiPs. The alpha level is corrected by a Bonferroni adjustment procedure (.05/8=.0063). Statistical significance is indicated by an asterisk.

The post-hoc comparison tests summarized in Table 4 support our observations: the third H tone is significantly higher than the second H tone for Speaker J. For Speaker R, H in the third clause is not significantly different from the second H tone. This is also an indication that the third Hs are enhanced in their rises. Recall that given declination, we would expect the third Hs appear lower than the second H (see §4.1), but this is not what we observe. Thus, the pattern of Speaker R indicates that the third F0 peaks must be raised by the extra H tone, which cancels out general lowering due to declination.

The assumption that there is a declination effect on preverbal rises is supported by the post-hoc comparison tests for Speakers N and Y in which the third H tone is significantly lower than the second (and the first) H tone. Declination is visible for these speakers because Utterance-final Hs do not boost H* in the penultimate MiP.

To summarize, positional variation of the H tones seen here is schematically represented in (11):

(11) (a) Speakers J and R (cf. Figure 17)

H*+L H

((((NP $_{obj}$ - *o* Verb) $_{MiP}$) $_{MaP}$) $_{IP}$) $_{Utt}$

(b) Speakers N and Y (cf. Figures 15)

H*+L H

((((NP $_{obj}$ - *o* Verb) $_{MiP}$) $_{MaP}$) $_{IP}$) $_{Utt}$

As illustrated in (11a), Speakers J and R associate the H tone with the accented mora of the object, along with the accent H*+L tone, giving an additional boost to the object's F0 peak. For Speaker N and Y, the H tone is associated with a case particle, as illustrated in (11b).

However this H is phonetically realized, it is crucial that it appears only in final clauses. This is inexplicable under the assumption that each clause behaves alike. Rather we need a level which can host this tone; only a level above IP, Utterance, can serve this purpose. See §5.3 for more discussion of this H.

4.5. Summary of IP and Utterance

We have identified the following characteristics of IP and Utterance in the prosodic phonology of Japanese:

(12) IP:
- Final lowering (due to L% at the end of IP)
- Final vowel creakiness
- Pause at right edge
- Larger initial rises compared to MaP
- Stronger pitch reset compared to MaP

Utterance:
- Larger initial rise compared to IP
- Stronger final lowering
- Domain of declination
- H associated with some mora in penultimate MiP

5. Discussion

5.1 Some implications for intonational phonology

This paper has investigated the intonational properties of gapping and coordination in Japanese, and adduced several pieces of evidence that IP and Utterance are present in the prosodic organization of Japanese. IP corresponds to a syntactic clause, and is characterized by tonal lowering, creakiness and a pause at its right edge. It also shows a larger initial rise and larger pitch reset compared to MaP. This level is distinct from a higher level of structure, Utterance, which exhibits stronger final lowering and initial rises and defines a domain of declination.

In addition to motivating a new level in the prosodic hierarchy of Japanese, our findings have several implications for the current theories of intonational phonology. First of all, we address the universality of the prosodic level inventory. This paper has shown that IP does play a role in the intonational phonology of Japanese, despite the fact that the past literature has not found evidence for it. This suggests that the set of levels that languages use is universal (cf. Jun in press). It is beyond the scope of this paper to fully defend such a universalist view, but we hope that, by showing that IP, a level hitherto not claimed to play a role in Japanese phonology, is motivated, we contribute toward such a theory.

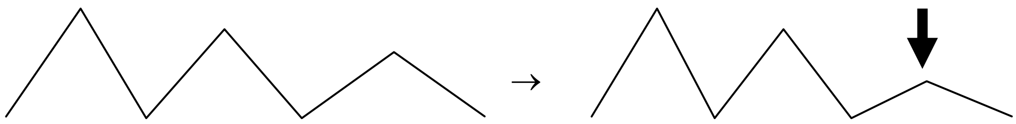
To the extent that prosodic levels are defined in Universal Grammar, what is motivated in one language should be present in every possible grammar; *ceteris paribus*, a theory that does not admit language variation at this fundamental level of linguistic organization is more restrictive. Selkirk (to appear) advances a view that each prosodic level is a phonological reflex of a syntactic category (e.g. MaP is a phonological reflex of XP; IP is a phonological reflex of Comma Phrase, etc.). According to this position, to the extent that syntactic categories are universal, prosodic categories should also be universal. Our study is informed and driven by such a universalist perspective: if IP is found to correspond with a syntactic clause in many languages, then evidence for this should be found in Japanese also. We hope to have shown this theory-driven research is fruitful to the extent that we have found evidence for IP corresponding to a syntactic clause in Japanese. It is of course an empirical issue whether this universal theory of prosodic structure is borne out, and further research is required to further defend this view.

The next issue is the locality of boundary tone influence. We found that IP-final H* tones are realized lower than IP-internal H* tones due to an IP-final L%. Moreover, the influence of this lowering effect is very local, confined to only the last H*+L tone. That is, the value of the L% boundary tones at the beginning of the clause-final MiP, for example, are not lowered in gapping compared to the corresponding coordination (See Figure 3 and Figure 4). This indicates that the MiP-initial L% tone is not subject to the lowering process.

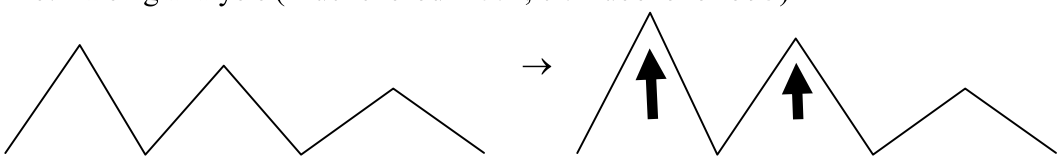
This runs counter to the assumption made by Pierrehumbert & Beckman (1988) that, in prosodic trees, boundary tones defined in terms of a certain prosodic level are associated with the node of that prosodic level (see also Hayes & Lahiri 1991; Pierrehumbert & Hirschberg 1990). For example, the phrasal H tones are directly associated with MiP, and similarly boundary H% and L% tones are directly attached to Utterance; such tones should have an influence over the entire phrase to which they are associated. However, the results obtained here suggest that the L% boundary tone appearing at the end of an IP may not be associated with the IP node, as the effect is quite local: lowering due to the L% boundary tone affects only the adjacent H*+L tone. Therefore, the effect of boundary tones may be much more local than hitherto assumed.

The third implication concerns how we formally capture “lowering.” We postulated that there is a lowering process IP-finally, but an alternative would be that there is a rising process affecting non-final peaks. Truckenbrodt (2004) suggests that the final lowering at prosodic edge may be explained by the tonal raising rule *Raising before Downstep*, whereby the height of a tone is amplified with respect to the following tone by some fixed ratio in positions followed by downstep, as in (13b).[12] With this rule, the final lowering is accounted for by raising all the H tones but the final one. As only the final H tone is exempted from the rule, it looks “lowered” with respect to the preceding H tones. At first glance, the lowering in IP-final positions in our data seems to be accounted for by the Raising-before-Downstep rule as well.

(13) a. Lowering analysis



b. Raising analysis (Truckenbrodt 2004; cf. Kubozono 1993)



However, such a raising analysis is not viable in light of all of our data. We found that the amount of lowering is greater in Utterance-final positions than in non-Utterance-final positions (§4.2). This cannot be explained under any kind of raising analysis because if final rises stay constant, as the raising analysis predicts, then we should not observe a difference between Utterance-final H and non-Utterance-final H, contrary to our observation.

5.2 Lowering process mechanism

In §3, we identified an F0 lowering process at the end of IP. However, we have not considered the precise mechanism behind the process. We address this issue in this section.

Our proposal is based on the idea of phonetic tonal coarticualtion, in which the H* tonal target in IP-final position is undershot because of a following L% boundary tone (cf. Xu 1993, 1997). The lowering process can be expressed as a phonetic implementation rule in which the accent H*+L tone is lowered by a fixed amount before a boundary L% tone at the end of IP, which is formalized below:

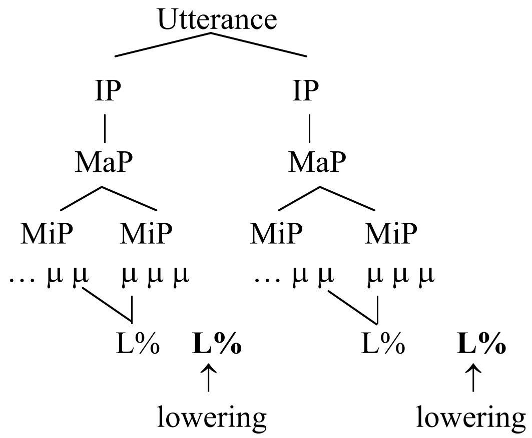
(14) $H^{*}{+}L \rightarrow rH^{*}{+}L \,/\, \_\_\_\; L\%)_{IP}\ (r < 1)$

[12] See also Kubozono (1993) who proposed a raising effect at the left edge of a branching node.

where *r* is a fixed ratio whereby the phonetic scaling of the tone is determined.

Since *r* is less than 1, the H*+L tone is realized lower than the original value. The mechanism in (14) is triggered by only a L% tone at the end of an IP, not by a tone at the end of other prosodic levels, as illustrated in (15).

(15)



One question that arises is how to account for the fact that the amount of lowering is greater at Utterance-final position than at Utterance-internal position. In answer to this question, we draw on Ladd's (1988: 541) proposal that "clause-initial accent peaks are higher following a stronger boundary." In line with this, we can posit that a higher prosodic boundary is signaled by stronger domain-final lowering (i.e. smaller *r*). This explains why the degree of F0 lowering is greater at the end of Utterance than at the end of IP.

An alternative phonological account could postulate a phonological tonal rewriting rule whereby an H* tone is changed into an allophonic downstepped !H* tone before a L% boundary tone (see Herman 1996 for references and counterarguments against such an approach). Such an approach has some undesirable implications for phonological theory, and intonational phonology in particular. In Japanese, there exists a well-known case of downstep whereby an H* tone following another instance of H* becomes !H* (Kubozono 1993; Pierrehumbert & Beckman 1988; Poser 1984 among others). If there were such a rewriting rule, it could apply to these !H*s, yielding doubly-downstepped !!H* tones (in case of the SL condition, for example). The rule can in principle be applied to a !!H* tone to give a !!!H* tone, which can be plugged into the rule to give a !!!!H* tone, etc. Such a rewriting rule is too powerful because it predicts languages with doubly-downstepped !!H or triply-downstepped !!!H* tones in their tonal inventory; these do not seem to be cross-linguistically attested.

Also, assuming multiply-downstepped tones undermines one of the most important advantages of the autosegmental and metrical theory of intonational phonology: various different tonal levels can be represented by only two distinctive tonal levels, H and L (Pierrehumbert 1980). The postulation of multiply-downstepped tones such as !!H and !!!H implies that many different tonal levels can, in principle, be distinguished by the number of downsteps (!), which is equivalent to adopting many different tonal levels.

5.3 The nature of Utterance-final H tones

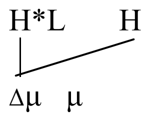
The nature of the extra H tone observed in Utterance-final position, described in §4.3, is worth a special note. This tone exhibits the properties of both a starred tone and a boundary tone. It appears toward the end of an Utterance, suggesting that it is a boundary tone. Its docking site, however, is not the absolute end of it. Instead, it appears in a penultimate MiP. The fact that it can appear on an accented mora also indicates that it might be strange to regard this as a boundary tone.

Grice et al. (2000) report that the phrase accents observed in the question intonation of a number of Eastern European languages show such mixed properties. In fact, they claim that this intermediateness itself is how a phrase accent is defined. For example, they show that the question intonation of Standard Greek is characterized by a H phrase accent followed by a L% boundary tone at the end of a sentence. However, the alignment of the phrase accent varies depending on the nuclear accent position. When the nuclear accent is in the final word, it is associated with its final syllable, which is consistent with the properties of boundary tones. However, when the nuclear accent is not in the final word, then the phrase accent docks onto the stressed syllable of the final word, behaving as if it were a starred tone. The behavior of the H tone found in our study is very similar to that of the phrase accent that Grice et al. (2000) argue for. Though details warrant more empirical investigation, our finding lends support to their idea of phrasal accent from a cross-linguistic point of view.

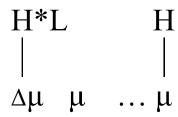
How then would we account for the behavior of this H tone? Related to this question, moreover, where does the inter-speaker variability come from? We claim that Optimality Theory (Prince & Smolensky 1993) sheds light on these issues.

Recall that there are two ways in which the H tone phonetically reveals itself: by boosting the penultimate accent H* or by docking onto the case particle of the preverbal element. For these tonal associations, the following two structures can be posited. In the first pattern, H is associated with the accented mora, shown in (16a) where Δμ represents a head mora in MiP (Δ for Designated Terminal String; Liberman & Prince 1977). The second way to realize the H tone is to associate it with a non-accented mora, depicted in (16b)

(16a)



(16b)



This variation makes sense in terms of cross-linguistic markedness: both ways of pitch docking are cross-linguistically marked. Thus, Optimality Theory (Prince & Smolensky 1993) provides an account of the speaker variation at issue. First, in both patterns, the H tone docks onto the penultimate MiP rather than onto the final MiP. This can be captured as an effect of the well-known tendency against final prominence, also known as NONFINALITY (Prince & Smolensky 1993).

(17) NONFINALITY(MiP): A H tone is not associated with a Utterance-final MiP.

It is also cross-linguistically observed that H is preferentially associated with a head position within a constituent (Bickmore 1995; de Lacy 2002b; Goldsmith 1987; Selkirk 2000).[13] This is the pattern that Speakers J and R conform to. They associate the H tone to the head mora of the penultimate MiP, as in (16a). Let us express this requirement as Δμ-TO-HTONE.

However, this strategy suffers from some markedness problems as well. One is that in this configuration, two distinct H tones are associated with one mora. This multiple linking of autosegments is known to be marked (e.g. Goldsmith 1984; see Kawahara to appear for recent summary and discussion). Another problem, which might be a consequence of multiple linking, is a matter of recoverability: in (16a), the presence of the H tone is signaled only by enhancing the rise of the lexical H*+L tone. It is likely that the presence of the H tone is perceptually hard to detect because listeners expect the presence of a rise there without that H tone anyway due to the lexical H*+L tone. So (16a) suffers from a perceptual problem in that the H tone is poorly signaled. A growing body of literature shows that such perceptual factors directly or indirectly shape phonological patterns (papers in Hume & Johnson 2001; Flemming 1995; Silverman 1997 among others); however, how to formalize such effects of perceptibility is yet to be explored. For the sake of exposition, we use *MULTILINK/RECOV, to express the markedness problems of (16a).

In Optimality Theoretic terms, the inter-speaker variation is obtained if the ranking between Δμ-TO-HTONE and *MULTILINK/RECOV differ between the speakers. For Speakers J and R, the ranking Δμ-TO-HTONE » *MULTILINK/RECOV holds. Therefore, the requirement that the H tone is associated with the head mora takes precedence. The H tone is never associated with the final verb since it would violate NONFINALITY(MiP), which is undominated in the case at hand.[14]

(18) Speakers J, R

	NONFIN(MIP)	Δμ-TO-HTONE	*MULTILINK/RECOV
H*L H │ │ $_{MiP}$(Δμ-o) $_{MiP}$( )	*!		
☞ H*L H │ ／ $_{MiP}$(Δμ-o) $_{MiP}$( )			*
H*L H │ │ $_{MiP}$(Δμ- o) $_{MiP}$( )		*!	

On the other hand, Δμ-TO-HTONE is dominated by *MULTILINK/RECOV for Speakers Y and N. Therefore, it is more important to avoid multiple-association of H to the accented mora. The interaction of these constraints for Speakers Y and N is illustrated

[13] More generally, prominent elements are attracted to prominent positions (see Aissen 1999; de Lacy 2002ab; Prince & Smolensky 1993).

[14] Note however that L% can be and is associated with the final MiP, suggesting that NONFINALITY(MIP) only targets H, which is prominent.

in the following tableau:

(19) Speakers Y, N

	NonFin(MiP)	Recov/*MultiLink	Δμ-to-HTone
H*L H | | $_{MiP}$(Δμ-o) $_{MiP}$( )	*!		
H*L H | / $_{MiP}$(Δμ-o) $_{MiP}$( )		*!	
☞ H*L H | | $_{MiP}$(Δμ - o) $_{MiP}$( )			*

One might wonder why, given that the H tone cannot be associated with the accented mora, it is the mora of the case particle that hosts the H. This is presumably an effect of well-documented preference for aligning tonal elements (and other linguistic elements) with domain edges (McCarthy & Prince 1993 and references cited therein).

We can thus derive the mixed characteristics of the H tone from constraint interaction. More study is warranted to see whether the differences between accentual Hs and boundary Hs are absolute and categorical, or whether they are simply tendencies arising from different rankings of violable constraints.

6. Conclusion

6.1 Summary

In this paper, we have shown that IP, hitherto unmotivated in the studies of Japanese intonation, is indeed motivated in the intonational phonology of Japanese. We have argued that Japanese IP is characterized by an F0 lowering process in its final position. We have proposed that the lowering is accounted for by assuming a L% boundary tone at the end of IP and by positing a phonetic implementation rule that lowers an IP-final H* tone before a L% tone that is at the end of an IP. We also showed that IPs are characterized by a pause and creaky vowels in final position, as well as large pitch reset and initial rises.

We have also provided evidence that multiple-clause constructions like gapping cannot be parsed into a set of homogeneous Utterances by showing that (i) the amount of clause-initial rise is larger in the first clause than in the second and third clauses, (ii) the H* tones generally decline over the course of the whole sentences (iii) verbal initial rises are smaller in the final clauses than in the first and second clauses, and (iv) there is a H tone found only in utterance-final positions.

The locus of IP corresponds to a syntactic clause, as predicted by the universalist view of the prosodic hierarchy advanced in Selkirk (to appear). We want to stress here that our research is much informed by such a universalist view because it is an expectation of such a stance that directed us to look at multiple-clause sentences in Japanese. We hope that this work shows the fruitfulness of such a line of research.

6.2 Remaining issues

We have seen that Utterance-initial H*s appear higher than Utterance-internal H*s and we attributed this observation to domain-initial strengthening effects. It is known that the degree of domain-initial strengthening in articulation is also greater at the boundary of higher-level prosodic constituents (see the references cited above). So a prediction of our proposal is that in gapping such articulatory strengthening at the segmental level is stronger at the beginning of initial clauses than at the beginning of non-initial clauses. Our data set did not control for segmental composition, so this topic is left for future research.

Another prediction of our proposal regards domain-final lengthening. It is known that domain-final segments undergo lengthening (e.g. Wightman 1992), and its extent is stronger at higher prosodic edges. We therefore predict that in gapping and coordination sentences, such lengthening is strongest in final clauses, compared to non-final ones. Again, since we did not control for the segmental contents in our experimental design, a test for this prediction needs to be left for further research.

References

Abe, J. and Hoshi, H. 1998. "Gapping and p-stranding," *Journal of East Asian Linguistics* 6, 101-136.

Aissen, J. 1999. "Markedness and subject choice in Optimality Theory," *Natural Language and Linguistic Theory* 17, 673-711.

Arvaniti, A. and Baltazani, M. in press. "Intonational analysis and prosodic annotation of Greek spoken corpora," In S.-A. Jun ed. *Prosodic Typology: The Phonology of Intonation and Phrasing***.** Oxford University Press.

Baumann, S., Grice, M., and Benzmuller, R. 2001. "GToBI – A phonological system for the transcription of German intonation," In P. Stanislaw and G. Demenko eds. *Prosody 2000: Speech Recognition and Synthesis*. Poznan: Adam Mickiewicz University, Faculty of Modern Languages and Literature, 21-28.

Bickmore, L. 1990. "Branching node and prosodic categories: Evidence from Kinyambo," In S. Inkelas and D. Zec eds. 1-17.

Bickmore, L. 1995. "Tone and stress in Lamba," *Phonology* 12, 307-341.

Bing, J. 1979. *Aspects of English Prosody.* Doctoral dissertation, University of Massachusetts Amherst.

Cho, T. and Keating, P. 2001. "Articulatory and acoustic studies of domain-initial strengthening in Korean," *Journal of Phonetics* 29, 155-190.

de Lacy, P. 2002a. *The Formal Expression of Markedness.* Doctoral dissertation, University of Massachusetts Amherst.

de Lacy, P. 2002b. "The interaction of tone and stress in Optimality Theory," *Phonology* 19, 1-32.

Féry, C. and Hartmann, K. to appear. "The focus and prosodic structure of German Right Node Raising and Gapping," *The Linguistic Review*.

Flemming, E. 1995. *Auditory Representations in Phonology*. Doctoral dissertation, UCLA.

Fougeron, C. and Keating, P. 1997. "Articulatory strengthening at edges of prosodic domains," *Journal of the Acoustical Society of America* 101, 3728-3740.

Fujisaki, H. and Hirose, K. 1984. "Analysis of voice fundamental frequency contours for declarative sentences of Japanese," *Journal of the Acoustical Society of Japan* 5, 233-242.

Goldsmith, J. 1984. Meeussen's Rule. In M. Aronoff and R. Oehrle eds. *Language Sound Structure*. Cambridge: MIT Press, 245-259.

Goldsmith, J. 1987. "Tone and accent, and getting the two together," *Proceedings of BLS* 13, 88-104.

Gordon, M. and Ladefoged, P. 2001. "Phonation types: A cross-linguistic overview," *Journal of Phonetics* 29, 383-406.

Grice, M., Ladd, R. and Arvaniti A. 2000. "On the place of phrasal accents in intonational phonology," *Phonology* 17, 143-185.

Hayashi, W., Hsu, C-S, and Keating, P. 1999 "Domain-initial strengthening in Taiwanese: A follow-up study," *UCLA Working Papers in Phonetics* 97, 152-156.

Hayes, B., and Lahiri, A. 1991. "Bengali intonational phonology," *Natural Language and Linguistic Theory* 9, 47-96

Herman, R., 1996. "Final lowering in Kipare," *Phonology* 13, 171-196.

Hsu, C-S and Jun, S-A. 1998. "Prosodic strengthening in Taiwanese: syntagmatic or paradigmatic?," *UCLA Working Papers in Phonetics* 96, 69-89

Hume, E. and Johnson, K. 2001. *The Role of Speech Perception in Phonology*. New York: Academic Press.

Hyman, L. 1990. "Boundary tonology and prosodic hierarchy," In S. Inkelas and D. Zec eds. 109-125.

Inkelas, S. and Zec, D. eds. 1990. *The Syntax-Phonology Connection*. CSLI Publication: Chicago.

Jun, S.-A. 1998. The accentual phrase in the Korean prosodic hierarchy. *Phonology* 15, 189-226.

Jun, S.-A. in press. *Prosodic Typology.* In S.-A. Jun ed. *Prosodic Typology: The Phonology of Intonation and Phrasing.* Oxford: Oxford University Press.

Kanerva, J. 1990. Focusing on phonological phrases in Chichewâ. In S. Inkelas and D. Zec eds. 145-161.

Kasai, H. and Takahashi S. 2000. "An ATB movement approach to gapping," *Proceedings of the 20000 International Workshop on Generative Grammar.*

Kawahara, S. to appear. "On the proper treatment of non-crisp edges," *Japanese/Korean Linguistics 13.*

Keating, P., Cho, T., Fougeron, C., and Hsu, C. 2003. "Domain-initial strengthening in four languages," *Papers in Laboratory Phonology 6: Phonetic interpretation.* Cambridge: Cambridge University Press, 145-163.

Kim, J.-S. 1997. *Syntactic Focus Movement and Ellipsis: A Minimalist Approach*. Doctoral dissertation, University of Connecticut.

Kubozono, H. 1993. *The Organization of Japanese Prosody.* Tokyo: Kuroshio Publishers.

Kuno, S. 1978. Japanese: A characteristic OV language. In W. Lehmann, ed. *Syntactic Typology.* Austin: University of Texas Press.

Ladd, D. R. 1988. "Declination 'reset' and the hierarchical organization of utterances," *Journal of the Acoustical Society of America* 84, 530-544.

Ladd, D. R. 1990. "Metrical representation of pitch register," In J. Kingston and M. E. Beckman eds. *Papers in Laboratory Phonology: Between the grammar and physics of speech.* Cambridge: Cambridge University Press, 35-57.

Ladd, D. R. 1993. "On the theoretical status of 'the baseline' in modelling intonation," *Language and Speech* 36, 435-451.

Liberman, M. and Prince, A. "On stress and linguistic rhyme," *Linguistic Inquiry* 8, 249-336.

Liberman, M. and Pierrehumbert, J. 1984. "Intonational invariance under changes in pitch range and length," In M. Aronoff and R. Oehrle eds. *Language Sound and Structure.* Cambridge: MIT Press, 157-233.

Maeda, S. 1976. *A Characterization of American English Intonation*. Doctoral dissertation, MIT.

McCarthy, J. and A. Prince. 1993. "Generalized alignment," *Yearbook of Morphology,* 79-153.

Nespor, M. and Vogel, I. 1986. *Prosodic Phonology*. Dordrecht: Foris.

Onaka, A. 2003. "Domain-initial strengthening in Japanese: an acoustic and articulatory study," *Proceedings of the 15th International Congress of Phonetic Sciences*, 2091-2094.

Pierrehumbert, J. 1979 "The perception of fundamental frequency declination," *Journal of the Acoustical Society of America* 66, 363-369.

Pierrehumbert, J. 1980. *The Phonology and Phonetics of English Intonation.* Doctoral dissertation, MIT. Published 1987 by Indiana University Linguistics Club.

Pierrehumbert, J. and Beckman, M. 1988. *Japanese Tone Structure.* Cambridge: MIT Press.

Pierrehumbert, J. and Hirschberg J. 1990. The meaning of intonational contours in the interpretation of discourse. In P. Kohen, J. Morgan & M. Pollock eds. *Intentions in Communications*. Cambridge: MIT Press, 271-311.

Poser, W. 1984. *The Phonetics and Phonology of Tone and Intonation in Japanese*. Doctoral dissertation, MIT.

Potts, C. 2003. *The Logic of Conventional Implicatures*. Doctoral dissertation, University of California at Santa Cruz.

Prince, A. and Smolensky P. 1993. *Optimality Theory: Constraint interaction in generative grammar*. ms. University of Colorado, Boulder and Rutgers University.

Saito, M. 1987. "Three notes on syntactic movement in Japanese," In T. Imai and M. Saito eds. *Issues in Japanese Linguistics*. Dordrecht: Foris.

Selkirk, E. O. 1986. "On derived domains in sentence phonology," *Phonology Yearbook* 3, 371-505.

Selkirk, E. O. 1995. "Sentence prosody: Intonation, stress, and phrasing," In J. Goldsmith ed. *The Handbook of Phonological Theory.* Blackwell, 550-69.

Selkirk, E. O. 2000. "The interaction of constraints on prosodic phrasing," In M. Horne ed. *Prosody: Theory and Experiment*. Dordrecht: Kluwer Academic Publishers, 231-271.

Selkirk, E. O. 2001. "Testing the prosodic structure hypothesis," *Journal of the Phonetic Society of Japan* 5, 53-56.

Selkirk, E. O. to appear. "Comments on intonational phrasing in English". *Prosodies (Phonetics and Phonology Series)*. Mouton de Gruyter.

Selkirk, E. O. and Tateishi, K. 1991. "Syntax and downstep in Japanese," In C. Georgopoulos and R. Ishihara eds. *Interdisciplinary Approaches to Language: Essays in Honor of S.-Y. Kuroda*. Dordrecht: Kluwer Academic Publishers, 519-43.

Shinya, T. to appear. The intonational asymmetry between argument and adjunct in Japanese. *University of Massachusetts Occasional Papers in Linguistics 30: Papers on Prosody*.

Silverman, D. 1997. *Phrasing and recoverability*. Doctoral dissertation, UCLA.

Sohn, K.-W. 1994. On gapping and right node raising. In *Explorations in Generative Grammar: A Festschrift for Dong-Whee Yang*. Hankuk Publishing, 589-611.

'tHart, J. and Cohen, A. 1973. "Intonation by rule: A perceptual quest," *Journal of Phonetics* 1, 309-327.

Thorsen, N. 1980. "A study of the perception of sentence intonation: Evidence from Danish," *The Journal of the Acoustical Society of America* 67, 1014-1030.

Truckenbrodt, H. 2002. "Upstep and embedded register levels," *Phonology* 19, 77-120.

Truckenbrodt, H. to appear. "Final lowering in non-final position," *Journal of Phonetics.*

Venditti, J. J. 1995. 'Japanese ToBI Labelling Guidelines." Ms, Ohio State University. Also printed in K. Ainsworth-Darnell and M. D'Imperio eds. *Ohio State University Working Papers in Linguistics* 50, 127-62.

Venditti, J. J. in press. "The ToBI model of Japanese intonation," In S.-A. Jun ed. *Prosodic Typology: The Phonology of Intonation and Phrasing.* Oxford: Oxford University Press.

Vogel, I. and I. Kenesei. 1987. "The interface between phonology and other components of grammar: The case of Hungarian," *Phonology Yearbook* 4, 243-263.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M. and Price, P. J. 1992. "Segmental durations in the vicinity of prosodic phrase boundaries," *Journal of the Acoustical Society of America* 91, 1707-1717.

Xu, Y. 1994. "Production and perception of coarticulated tones," *Journal of the Acoustical Society of America* 95, 2240-2253.

Xu, Y. 1997. "Contextual tonal variations in Mandarin," *Journal of Phonetics* 25, 62-83.

Yufang, Y. and W. Bei 2002. "Acoustic correlates of hierarchical prosodic boundary in Mandarin," *Proceedings of Speech and Prosody 2002*.