Original Paper

Phonetica 2008;65:62–105
DOI: 10.1159/000130016

Received: January 23, 2005
Accepted: December 15, 2007

The Intonation of Gapping and Coordination in Japanese: Evidence for Intonational Phrase and Utterance

Shigeto Kawahara[a] Takahito Shinya[b]

[a]University of Georgia, Athens, Ga., USA; [b]RIKEN Brain Science Institute, Wako, Saitama, Japan

Abstract

In previous studies of Japanese intonational phonology, levels of prosodic constituents above the Major Phrase have not received much attention. This paper argues that at least two prosodic levels exist above the Major Phrase in Japanese. Through a detailed investigation of the intonation of gapping and coordination in Japanese, we argue that each syntactic clause projects its own Intonational Phrase, while an entire sentence constitutes one Utterance. We show that the Intonational Phrase is characterized by tonal lowering, creakiness and a pause in final position, as well as a distinctive large initial rise and pitch reset at its beginning. The Utterance defines a domain of declination, and it is signaled by an even larger initial rise, as well as a phrasal H tone at its right edge. Building on our empirical findings, we discuss several implications for the theory of intonational phonology.

1 Introduction

The theory of prosodic phonology posits a number of levels in prosodic structure. The prosodic levels define the domains or loci of a wide variety of phonological processes such as stress assignment, tonal downstep, boundary tone association, assimilation, dissimilation, and resyllabification, to name a few [Hayes and Lahiri, 1991; Jun, 1998; Nespor and Vogel, 1986; Pierrehumbert and Beckman, 1988; Selkirk, 1980, 1986, 2001 among many others]. Moreover, previous studies have demonstrated that articulatory strength becomes greater at higher prosodic structure boundaries [Cho and Keating, 2001; Fougeron and Keating 1997; Hayashi et al., 1999; Hsu and Jun, 1998; Keating, 2003; Keating et al., 2003; Onaka 2003]. One example of a prosodic hierarchy is illustrated in (1) [Selkirk, 1986, 1995, 2000, 2001, 2005]:

(1) Utterance
|
Intonational Phrase
|

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0031–8388/08/0652–0062
$24.50/0
Accessible online at:
www.karger.com/pho

Shigeto Kawahara
Department of English
University of Georgia
Park Hall 254, Athens, GA 30602-6205 (USA)
Tel. +1 706 542 2196, Fax +1 706 542 2181,
E-Mail kawahara@uga.edu

Major Phrase
|
Minor Phrase
|
Prosodic Word
|
Foot
|
Syllable
|
Mora

Of particular relevance to our current study is the Intonational Phrase (IP), one level above the Major Phrase (MaP) and below the Utterance. Although the intonation of Japanese has received much attention in the previous literature, no studies have adduced evidence for the existence of the IP. The Minor Phrase (MiP) in Japanese is signaled by initial lowering, and it also defines the domain in which a single lexical accent is allowed [McCawley, 1968]. The MaP domain is characterized by a larger pitch reset and a larger initial rise than the MiP [Selkirk et al., 2003; Selkirk and Tateishi, 1991]. However, little has been said about prosodic levels higher than the MaP. For instance, Pierrehumbert and Beckman [1988] posit no prosodic levels between the MaP (= their Intermediate Phrase) and the Utterance. In the JToBI model [Venditti, 1997, 2005], Pierrehumbert and Beckman's Intermediate Phrase and Utterance are merged into a single prosodic level called the Intonation Phrase, resulting in only one level above the MiP (Venditti's Intonation Phrase differs from our IP). In short, the IP in (1) had no previous motivation in studies of Japanese prosody.

However, the IP has been shown to play a role in many other languages: Chicheŵa [Kanerva, 1990, pp. 146–147]; English [Beckman and Pierrehumbert, 1986; Nespor and Vogel, 1986, chapter 7; Selkirk, 2005]; German [Baumann et al., 2001; Féry and Hartmann, 2005; Truckenbrodt, 2004]; Greek [Arvaniti and Baltazani, 2005]; Hungarian [Vogel and Kenesei, 1987]; the Tuscan dialect of Italian [Nespor and Vogel, 1986]; Kinande [Hyman, 1990, pp. 112–121]; Kinyambo [Bickmore, 1990, p. 8]; LuGanda [Hyman, 1990, pp. 111–112]; Norwegian [Kristoffersen, 2000]; Spanish [Nespor and Vogel, 1986] and others [see the contributions in Jun, 2005]. For example, in Chicheŵa, the IP defines the domain of tonal catathesis, hosts several types of boundary tones, and exhibits final lengthening at its end [Kanerva, 1990]. The IP can also define the domain of segmental phonological processes; for instance, intervocalic spirantization of voiceless stops in Italian applies only within the IP [Nespor and Vogel, 1986, pp. 205–211]. Given the cross-linguistic role of the IP across languages, we raise one question: does the IP exist in Japanese? We intend to show in this paper that it does.

In many of the cases cited above, a syntactic clause projects its own IP. Given the cross-linguistic tendency for a syntactic clause to project an IP, we expect to find evidence for an IP in multi-clause sentences in Japanese. For this reason, gapping and coordination serve as useful targets for the current study, as they constitute typical multi-clause constructions in Japanese.[1] The investigation of the intonation of gapping

[1]This paper focuses on clause boundaries appearing in coordination and gapping sentences. The properties of clause boundaries appearing in other constructions such as embedded clauses should be investigated in a separate study.

and coordination shows that an IP indeed exists in Japanese prosody. Gapping, as shown in (2), minimally contrasts with coordination in (3) in that the verbs in nonfinal clauses go unpronounced:[2, 3]

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

Murasugi-wa	namauni-o	~~moritsuke,~~	Munakata-wa
Murasugi-TOP	sea urchin-ACC	put on dish	Munakata-TOP

mamemochi-o	~~moritsuke,~~	Morimura-wa	aemono-o	moritsuketa.
bean rice cake-ACC	put on dish	Morimura-TOP	mixed salad-ACC	put on dish

'Murasugi put a sea urchin on a dish, Munakara a bean rice cake, and Morimura a mixed salad.'

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

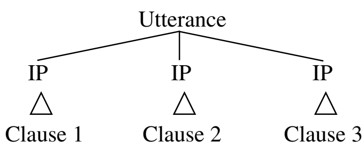
Murasugi-wa	namauni-o	moritsuke,	Munakata-wa
Murasugi-TOP	sea urchin-ACC	put on dish	Munakata-TOP

mamemochi-o	moritsuke,	Morimura-wa	aemono-o	moritsuketa.
bean rice cake-ACC	put on dish	Morimura-TOP	mixed salad-ACC	put on dish

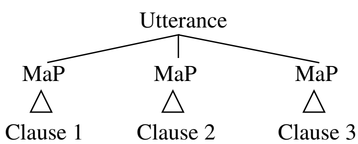
'Murasugi put a sea urchin on a dish, Munakara put bean rice cake on a dish, and Morimura put a mixed salad on a dish.'

We argue that in order to account for the aspects of intonation of gapping and coordination sentences, we must posit an IP, a level above the MaP and below the Utterance. Specifically, we analyze sentences like (2) and (3) as receiving a prosodic parse where each clause corresponds to an IP, and the entire sentence has the status of an Utterance, as depicted in (4a). We show that the IP and the Utterance are characterized by different sets of phonetic properties, and that these properties contrast with the properties of the MaP. We further argue that other possible structures such as those depicted in (4b) and (4c) are inadequate models of the prosodic structure of multi-clause sentences in Japanese.

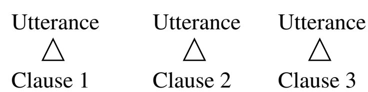
(4a)



(4b)



(4c)



[2] Syntactically, nonfinal verbs remain unpronounced in Japanese gapping sentences whereas noninitial verbs are elided in corresponding English sentences. No agreement has been reached on the syntactic analysis of gapping in Japanese. Some authors argue that what seems to be gapping is in fact Right-Node-Raising involving Across-the-Board movement [Kasai and Takahashi, 2000; Kuno, 1978; Saito, 1987], while Abe and Hoshi [1997] argue that gapping is base-generated, and gapped verbs are filled in at LF by copying. Several Korean scholars [Kim, 1997; Sohn, 1994], based on a similar construction in Korean, argue that such constructions involve deletion under identity at PF. We are not concerned with this debate regarding the syntactic nature of gapping. Of importance to us are the fact that gapping consists of multiple distinct clauses, and the fact that nonfinal verbs in gapping remain unpronounced.

[3] We use the following abbreviations in this paper: ACC = accusative, ADJ = adjectival ending, DAT = dative, GEN = genitive, LOC = locative, TOP = topic.

Previous studies have shown that there is no consistent association across languages between levels in the prosodic hierarchy and the phonological phenomena that characterize them (see the literature cited above). The cross-linguistic inconsistency suggests that the levels of the hierarchy are not defined by these phenomena. We thus instead define prosodic levels by syntactic criteria, following Selkirk [2005], and use syntax as our basis for finding prosodic levels above the MaP (see § 6.1 for further discussion). Specifically, we compare edges of four different syntactic structures illustrated in (5) (NP, VP, clause, sentence), and show that prosodic levels aligned with these boundaries must all be distinctive.

(5)					
	Syntactic boundaries:	$_{Sentence}$[. . .	$_{Clause}$[. . .	$_{VP}$[. . .	$_{NP}$[. . .]. . .]. . .]. . .]
		↓	↓	↓	↓
	Prosodic boundaries:	level 4 (Utterance)	level 3 (IP)	level 2 (MaP)	level 1 (MiP)

The lowest level – level 1 – maximally contains one accented NP, and this level has traditionally been labeled as the MiP [McCawley, 1968]. We argue that three levels exist above the MiP. In § 3 we demonstrate that the prosodic boundary aligned with the left edge of a VP shows properties distinct from the MiP. Following Selkirk and Tateishi [1991], we call this phrase the MaP. In § 4 we provide evidence which shows that clause boundaries in gapping and coordination show properties that differ from the MaP – we call this level the IP. In this section, we reject the structure (4b), which does not recognize a difference between the MaP and the IP. Finally, in § 5 we show that sentence edges exhibit properties distinct from those of IP edges. We call this highest level the Utterance. We also reject the structure in (4c), which does not recognize any distinction between a clause-level and sentence-level prosodic constituent.

2 Methods

2.1 Speakers

We recruited 4 female native speakers of Japanese (J, N, R, and Y) at the University of Massachusetts, Amherst. All received payment 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 Iwata (Shizuoka prefecture), 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 materials consisted of sets of coordination and gapping sentences, in addition to other types of sentences. The sentences used in the experiment are summarized in (6), with a schematic structure and one example for each condition:

(6) Experimental sentences

a. SS (coordination vs. gapping)

[[N-TOP] [[N-ACC] (V)]], [[N-TOP] [[N-ACC] (V)]], [[N-TOP] [[N-ACC] V]].

Murasugi-wa	namauni-o	(moritsuke),	Munakata-wa
Murasugi-TOP	raw sea urchin-ACC	put on dish	Munakata-TOP

mamemochi-o	(moritsuke),	Morimura-wa	aemono-o	moritsuketa.
bean rice cake-ACC	put on dish	Morimura-TOP	mixed salad-ACC	put on dish

'Murasugi put a raw sea urchin on a dish, Munakata put a bean rice cake on a dish and Morimura put a mixed salad on a dish.'

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]].

Yonekura-wa	murasaki-no	aibana-o	(nagame),	Ninomiya-wa
Yonekura-TOP	purple-GEN	dayflower-ACC	stare at	Ninomiya-TOP

yawaraka-na	oniyuri-o	(nagame),	Imamoto-wa	uraraka-na
soft-ADJ	lily-ACC	stare at	Imamoto-TOP	bright-ADJ

yamabuki-o	nagameteiru.
yellow rose-ACC	staring at

'Yonekura is staring at a purple dayflower, Ninomiya is staring at a soft lily and Imamoto is staring at a bright yellow rose.'

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]].

Morioka-no	aniyome-wa	Muramatsu-o	(urayami),	Yamagata-no
Morioka-GEN	sister-in-law-TOP	Muramatsu-ACC	envy	Yamagata-GEN

yamabushi-wa	Yamanishi-o	(urayami),	Aomori-no	omawari-wa
monk-TOP	Yamanishi-ACC	envy	Aomori-GEN	police-TOP

Nonomura-o	urayanda.
Nonomura-ACC	envied

'The sister-in-law in Morioka envied Muramatsu, the monk in Yamagata envied Yamanishi, and the police in Aomori envied Nonomura.'

d. Dative (coordination vs. gapping)

[[N-TOP] [[N-DAT] [N-ACC (V)]]], [[N-DAT] [N-ACC (V)]], [[N-DAT] [N-ACC V]].

Morishita-wa	Yamanashi-ni	yamabuki-o	(hakobi),	Aomori-ni	oniyuri-o
Morishita-TOP	Yamanashi-DA	yellow rose-ACC	bring	Aomori-DAT	lily-ACC

(hakobi),	Nagasaki-ni	aibana-o	hakonda.
bring	Nagasaki-DAT	dayflower-ACC	brought

'Morishita brought a yellow rose to Yamanashi, brought a lily to Aomori, and brought a dayflower to Nagasaki.'

e. Predicative (short vs. long predicates)

[[N-TOP] [(N-GEN) N-da]].

Yamaura-wa	(Umemachi-no)	awauri-da.
Yamaura-TOP	Umemachi-GEN	millet seller-copula

'Yamaura is a millet seller (in Umemachi).'

f. Intransitive (short vs. long predicates)

[[N-NOM] [(N-Loc) V]].

Mari-ga	(nominoichi-de)	nenaoshita.
Mari-NOM	flea market-LOC	fell back to sleep

'Mari fell back to sleep (in a flea market).'

g. Long single-clause sentence

[[N-GEN N-NOM] [[N-GEN N-LOC] [N-DAT [N-ACC V]]]].

Omawari-no	Muramatsu-ga	Urumuchi-no	yamayama-de	naramatsu-ni
police-GEN	Muramatsu-NOM	Ürūmqi-GEN	mountains-LOC	pine trees-DAT

namamizu-o	agemashita.
water-ACC	gave

'Muramatsu, a police officer, watered pine trees in the mountain in Ürūmqi.'

All words in the experimental sentences had accents on the second mora, and were four moras long except in single-clause filler sentences (e–g). The basic syntactic structure for gapping and coordination was 'S-O-(V), S-O-(V), S-O-V'. Since it is known that constituent branching affects intonational patterns [Bickmore, 1990; Kubozono, 1993; Selkirk, 2000; Shinya, 2005], we systematically varied the length of the subjects and objects by changing the number of words they comprised. The 'short' subjects/objects consisted of a single word while the 'long' subjects/objects consisted of two words. We tested three combinations of short/long and subjects/objects: SS (Short subject and Short object, 6a), SL (Short subject and Long object, 6b), LS (Long subject and Short object, 6c).[4] 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 required by Japanese orthographic convention.

In addition to these three types of sentences, we included other kinds of sentences. First, we used coordination and gapping sentences with dative and accusative objects (6d). The dative sentences had the following syntactic structures: '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' predicate followed by *-da* (copula) (6e). The short predicate consisted of a single noun followed by *-da*, and the long predicate consisted of two words [Noun-GEN (*-no)* Noun], followed by *-da*. Schematically, the syntactic structure was thus 'S (N-GEN) N-*da*'. Second, we included short and long intransitive sentences (6f) whose predicates had an intransitive verb. The long sentences had a pre-verbal locative phrase, whereas the short sentences did not; the sentences had the structure 'S-(LOC)-V'. Finally, for each gapping-coordination pair, we added two single-clause sentences with the same number of words as the gapping condition (6g). The appendix provides a list of all of the sentences.

2.3 Recording

Each speaker participated in two recording sessions. We recorded the minimal pairs of gapping and coordination sentences on different days, lest the speakers notice the contrasts. Their speech was recorded to CDs in a sound-attenuated booth in the phonetics laboratory at the University of

[4] The other possible combination, LL (Long subject and Long object), was not included in the experiment. We assumed that the intonation pattern in the LL condition would be analogous to the SS condition because in both combinations, subject and object are of the same length i.e. have a symmetrical structure.

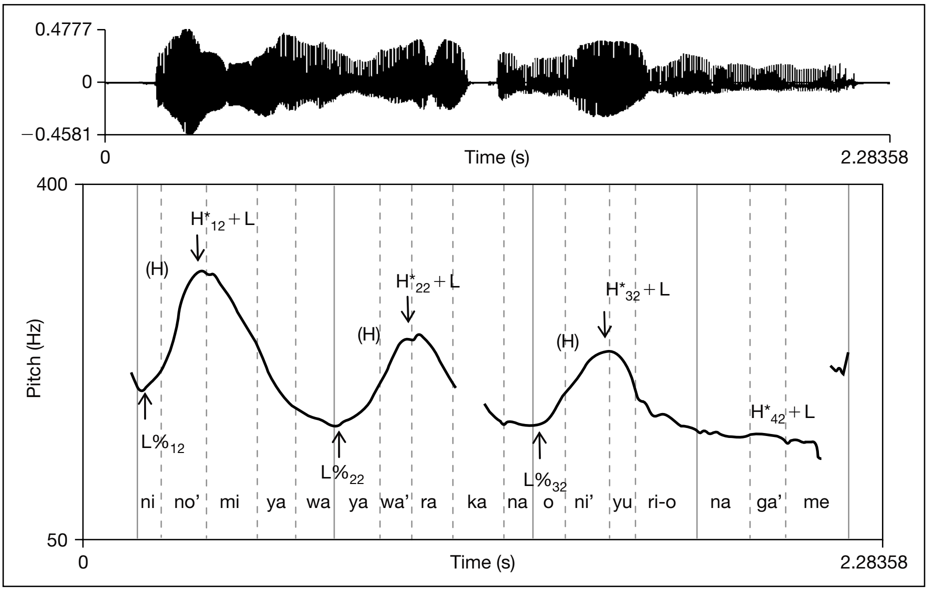


Fig. 1. Illustration of the F_0 measurement points. The pitch track was taken from the second clause in a coordination sentence spoken by speaker R. The clause is *Nino'miya-wa yawa'raka-na oni'yuri-o naga'me*. . . 'Ninomiya gazed at a soft lily. . .'

Massachusetts, Amherst. The experimental sentences were written on index cards in the usual Japanese orthography, which is a mixture of the *hiragana* and *katakana* syllabaries and *kanji* 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 rate of speech as naturally as possible. They read through the stimulus set 6 times. The order of the stimuli 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 digitized at an 11,025 Hz sampling rate and 16 bit quantization level. The recorded sentences were then submitted to F_0 measurement using PitchWorks (Scicon R&D).

The guidelines for F_0 measurement were as follows. Figure 1 shows a typical F_0 contour of one clause of a coordination sentence. The arrows indicate where we measured F_0. The solid lines and dashed lines represent the MiP boundaries and mora boundaries, respectively. Apostrophes on the gloss denote accent locations. The F_0 of an accented word is characterized by a H*+L tone on the accented mora [Pierrehumbert and Beckman, 1988]. We measured the F_0 peaks and valleys that appeared in each MiP, assuming that the peaks represent the H* tones, and valleys represent the boundary L% tones (for measurements of initial L%H on verbs in coordination, see § 5.2).

In theory, a phrasal H is present between a L% and a H*+L, but is not seen clearly in our case because in words with an accent on the second mora, the phrasal H appears 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 as H_{21}.

For the statistical analyses in the main sections of this paper, we normalized the raw F_0 values using the formula in (7), adopted from Truckenbrodt [2004]:

(7) Transformed value = (Original value − $Mean_L$)/($Mean_H$–$Mean_L$)
where $Mean_T$ is the speaker-specific mean of tone T.

The Mean H value was defined as the mean value of H_{11} across all the multi-clause sentences for each speaker, and $Mean_L$ as the mean value of penultimate L tones in the final clauses (= 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, the transformation defines a pitch range for each speaker by $Mean_H$–$Mean_L$, and relativizes each tonal value to the pitch range.

The normalization has the following virtues: since overall we found little interspeaker variation (with one exception; see § 5.3), normalization allows us to pool all speakers' data, which simplifies the analysis and exposition of the data. Second, by pooling all speakers' data, a less drastic post-hoc α-level Bonferroni adjustment was required for multiple comparisons in statistical analyses. For example, if we were to analyze the data separately for each of our 4 speakers, and tried to make the same comparison across three clauses, then the α-level would have needed to be adjusted to $0.05/(3\times4) = 0.004$. With normalization, we can avoid such a drastic adjustment.

In comparing data points from a single sentence, we used a repeated-measures analysis because it has more power. However, we used an independent-sample t test when we made comparisons of different sentences (e.g. comparison of gapping and coordination sentences). When we used t tests, we chose two-tailed tests to be conservative. We adjusted the α-level by the Bonferroni method, when necessary.

Finally, a comment regarding the focus structure of the target sentences is in order: gapping and coordination sentences involve contrastive focus of nonverbal elements, and therefore one might wonder what the effect of such focus is on the intonation of these sentences [see Ishihara, 2003; Ladd, 1996; Rooth, 1996; Selkirk, 2002; Sugahara, 2003; Truckenbrodt, 1995 among others for the effect of contrastive focus on intonation]. In our gapping and coordination sentences, *every* nonverbal element bears new information and stands in contrastive focus, and the focus status of those elements is the same across the three clauses. Therefore, in comparing peaks in different clauses, we assume that the items across the three clauses have equal focus.

3 The Distinction between the MaP and the MiP

The central goal of this paper is to establish prosodic constituent levels above the MaP. Before diving into this discussion, however, we must first establish a prosodic distinction between the MiP and the MaP. This preliminary discussion provides a basis for our argument for the IP in Japanese, developed in § 4. In § 3.1 we review Selkirk and Tateishi's [1991] characterization of the MaP. In § 3.2 we report further evidence from our dataset which reinforces their analysis.

3.1 Selkirk and Tateishi [1991]

A number of previous studies have shown that there exists a prosodic level whose left edge coincides with the left edge of a syntactic VP [Selkirk, 2000; Selkirk et al., 2003; Selkirk et al., 2004; Selkirk and Tateishi, 1991]. The prosodic level aligned with a VP edge is characterized by a pitch reset significantly larger than the one we find at a MiP edge, which can contain at most one lexical accent [McCawley, 1968; Pierrehumbert and Beckman, 1988; Poser, 1984].

Selkirk and Tateishi [1991] found that given two adjacent accented peaks, the second peak appears higher when there is an intervening VP boundary than when there is no such boundary. In other words, the F_0 downtrend – lowering of an H peak after another H peak – is weaker when two F_0 peaks are separated by a VP edge than when they are not. Selkirk and Tateishi's [1991] example sentences appear in (8):

(8) a.	[[Ao'yama-no	Yama'guchi-ga]$_{NP}$	[ani'yome-o	yonda]$_{VP}$]$_{S}$.
	Aoyama-GEN	Yamaguchi-NOM	sister-in-law-ACC	called

Yamaguchi from Aoyama called his/her sister-in-law.'

b.	[[Ao'yama-ga]$_{NP}$	[Yama'guchi-no	ani'yome-o	yonda]$_{VP}$]$_{S}$.
	Aoyama-NOM	Yamaguchi-GEN	sister-in-law-ACC	called

'Aoyama called his/her sister-in-law from Yamaguchi.'

The two sentences in (8) have different syntactic structures: in (8a), the first two nouns form a complex subject NP with no VP boundaries between them, whereas in (8b) the first two nouns are separated by a VP boundary. Selkirk and Tateishi [1991] showed that the F_0 peak of *Yamaguchi* is higher in (8b) than in (8a), relative to the peak of *Aoyama*. The difference between (8a) and (8b) indicates that there exists a prosodic boundary that coincides with a VP edge. Since all accented nouns project their own MiP [i.e. *Yama'guchi* should have a MiP edge to its left in both (8a) and (8b)], the phrase that corresponds to a VP edge must be a level higher than the MiP i.e. the MaP.

One cautionary remark is in order. Selkirk and Tateishi's [1991] MaP might look similar to Pierrehumbert and Beckman's [1988] and Kubozono's [1993] Intermediate Phrase (ip), but the definitions of the MaP and the ip are not exactly the same. The definition of the MaP is based on blocking of downtrend – lowering of an H peak after another H peak – whereas the ip is defined as a domain of downstep – lower realization of an H peak when it is preceded by an accented word than when it is preceded by an unaccented word. If the MaP is identical to the ip, it is predicted that downstep is cancelled at the left edge of a VP. However, since the definitions of the MaP and the ip have different bases, nothing excludes the possibility that the MaP does not coincide with the domain of downstep. We leave this issue – whether MaP and ip are the same or not – for future research.[5]

In the next subsection, we provide evidence from our dataset that reinforces the distinction between the MiP and the MaP. By comparing the F_0 downtrend between two adjacent F_0 peaks, we demonstrate that the H tone sequences with an intervening VP boundary show a smaller downtrend than those without it, confirming that the prosodic boundary that occurs at the left edge of a VP is larger than the MiP boundary.

3.2 Further Data Supporting VP-MaP Alignment

In our dataset, the evidence for the existence of the MaP comes from three comparisons of F_0 peak differences: (a) SS coordination vs. LS gapping, (b) dative gapping vs. SL gapping, and (c) dative coordination vs. LS coordination. These sentences are schematically illustrated in (9). Shown in bold are the target words whose peak differences were

[5] If MaP and ip are the same prosodic level, it is predicted that downstep is blocked across a MaP boundary [see Selkirk and Tateishi, 1991, for a relevant discussion]. This prediction can be tested by comparing the heights of a H-tone in two conditions: when it is preceded by an accented phrase across a VP boundary and when it is preceded by an unaccented phrase.

compared; VP boundaries that separate the target items are shown by ►. We chose these three comparisons because the sentences in each pair had the same number of words per clause and minimally differ from each other in the presence/absence of a syntactic VP boundary. The first type in each pair involves a VP boundary ('across-VP condition') whereas the second type does not ('control'). Comparisons were made in the first and the second clauses in (9a), but only in the first clause in (9b) and (9c), because the second clauses in (9b) and (9c) contain different numbers of words.

(9)

a. Comparison 1
SS coordination (across-VP condition)
$[[$**N-TOP**$]_{NP}$ ► $[$**N-ACC** $V]_{VP}$ $]_{S}$, $[[$**N-TOP**$]_{NP}$ ► $[$**N-ACC** $V]_{VP}]_{S}$, $[[N\text{-}TOP]_{NP}$ $[N\text{-}ACC\ V]_{VP}]_{S}$.

LS gapping (control)
$[[$**N-GEN N-TOP**$]_{NP}$ $[N\text{-}ACC]_{VP}]_{S}$, $[[$**N-GEN N-TOP**$]_{NP}$ $[N\text{-}ACC]_{VP}]_{S}$, $[[N\text{-}GEN$ $N\text{-}TOP]_{NP}$ $[N\text{-}ACC\ V]_{VP}]_{S}$.

b. Comparison 2
Dative gapping (across-VP condition)
$[[N\text{-}TOP]_{NP}$ $[$**N-DAT** ►$[$**N-ACC**$]_{VP}]_{VP}]_{S}$, $[N\text{-}DAT\ N\text{-}ACC]_{S}$, $[N\text{-}DAT\ [N\text{-}ACC$ $V]_{VP}]_{S}$.

SL gapping (control)
$[[N\text{-}TOP]_{NP}$ $[$**N-GEN N-ACC**$]_{VP}]_{S}$, $[[N\text{-}TOP]_{NP}$ $[N\text{-}GEN\ N\text{-}ACC]_{VP}]_{S}$, $[[N\text{-}TOP]_{NP}$ $[N\text{-}GEN\ N\text{-}ACC\ V]_{VP}]_{S}$.

c. Comparison 3
Dative coordination (across-VP condition)
$[[N\text{-}TOP]_{NP}$ $[$**N-DAT** ►$[$**N-ACC** $V]_{VP}]_{S}$, $[N\text{-}Dat\ [N\text{-}ACC\ V]_{VP}]_{S}$, $[N\text{-}Dat\ [N\text{-}ACC$ $V]_{VP}]_{S}$.

SL coordination (control)
$[[N\text{-}TOP]_{NP}$ $[$**N-GEN N-ACC** $V]_{VP}]_{S}$, $[[N\text{-}TOP]_{NP}$ $[N\text{-}GEN\ N\text{-}ACC\ V]_{VP}]_{S}$, $[[N\text{-}TOP]_{NP}$ $[N\text{-}GEN\ N\text{-}ACC\ V]_{VP}]_{S}$.

For each pair in (9), F_0 peak differences between the target words were calculated, and a comparison was made between the across-VP conditions and the control conditions. The graphs in figure 2 show the mean differences in the degree of downtrend for the three comparisons. In these summary figures, and also in those to follow, error bars represent 95% confidence intervals.

In all of the comparisons in figure 2, the mean peak differences are smaller in the across-VP conditions than in the control condition – i.e. downtrend of the second peaks is smaller in the across-VP condition than in the control condition, because the pitch reset is larger when there is an intervening VP edge than when there is not.

An ANOVA was performed on the data for comparison 1 with CLAUSE (1st and 2nd) and TYPE (across-VP condition and within-VP condition) as independent variables. There was a significant main effect for both of the variables [CLAUSE: $F(1, 97) = 87.956$, $p < 0.001$, TYPE: $F(1, 97) = 41.149$, $p < 0.001$], and there was no significant interaction between CLAUSE and TYPE ($F < 1$). The significant main effect of TYPE shows the peak differences are smaller in the across-VP conditions than in the control conditions. The main effect of CLAUSE can be understood as the effect of a narrowed pitch range due to the general declining trend in an utterance (see § 5.1 for more on declination). For comparisons

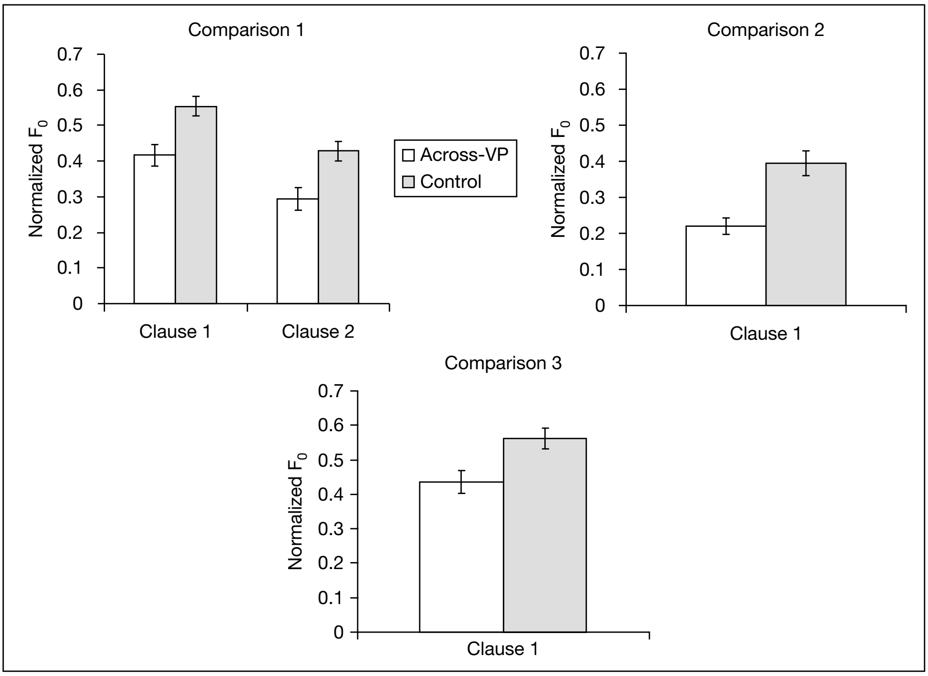


Fig. 2. Mean peak differences between the across-VP conditions and the control conditions: comparison 1 = SS coordination vs. LS gapping; comparison 2 = dative gapping vs. SL gapping; comparison 3 = dative coordination vs. SL coordination. The error bars represent 95% confidence intervals.

2 and 3, independent-samples t tests were conducted. The results show that peak differences in the across-VP conditions were significantly smaller than those in the control condition in both comparisons ($\alpha = 0.05/2 = 0.025$; comparison 2: $t(87.38) = -7.129$, $p < 0.001$, comparison 3: $t(96) = -4.658$, $p < 0.001$). These results statistically confirm our previous conclusion that the F_0 downtrend is smaller between two peaks when they are separated by a VP boundary than when they are not. The data reinforce the claim that there must be a level higher than the MiP at the left edge of a VP.

To summarize, we have provided further evidence for Selkirk and Tateishi's [1991] finding that there exists a MaP boundary at the left edge of a VP. In the next section, based on this finding, we show that there exists a prosodic constituent level yet higher than the MaP in Japanese, i.e. the IP.

4 The Distinction between the IP and the MaP

Drawing upon the results of § 3, we now argue for the existence of the IP in Japanese, which in the case at hand corresponds to a syntactic clause in gapping and coordination sentences. We discuss five pieces of evidence for the existence of the IP: F_0 lowering at clause-final positions (§ 4.1), creaky vowels and pauses at clause-final

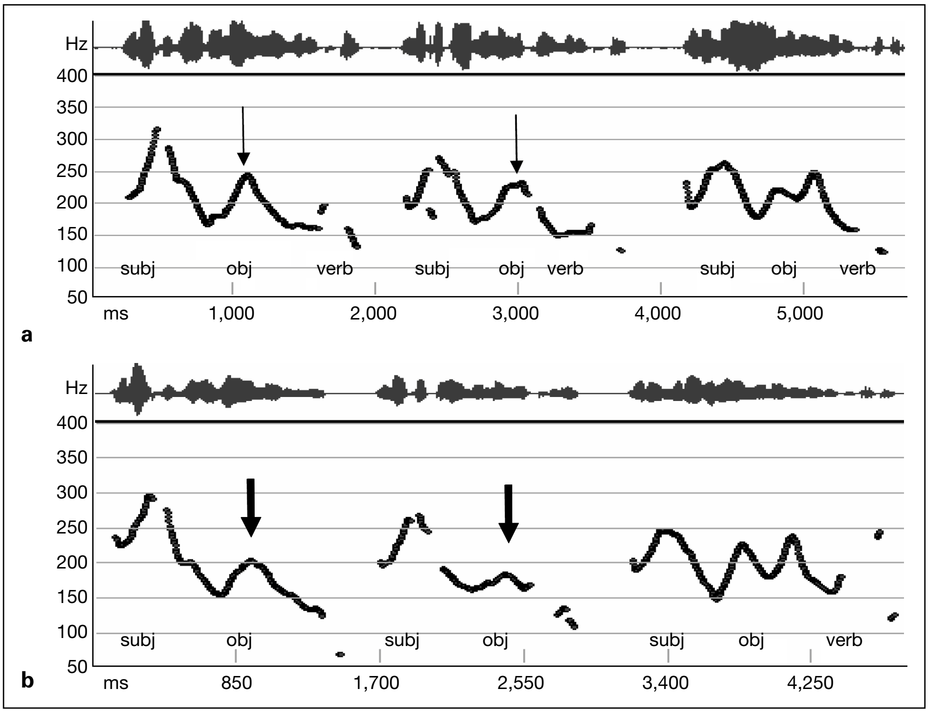


Fig. 3. Representative pitch tracks of coordination (**a**) and gapping (**b**) uttered by speaker Y.

positions (§ 4.2 and § 4.3), pitch resets between clauses (§ 4.4), and clause-initial F_0 rises (§ 4.5).

4.1 Final Lowering

4.1.1 Observation

We start with the final lowering found at the end of each clause. We found that the clause-final H* peaks in nonfinal gapped clauses are systematically lower than the corresponding H* peaks in the corresponding coordination sentence. An illustrative pair of pitch tracks is given in figure 3, where the clause-final accent H* peaks (i.e. the H* in the objects) in the nonfinal clauses are lower in gapping (shown by the thick arrows) than in coordination (thin arrows), despite the fact that these tones are hosted by the same lexical items.

To show that the greater lowering is not a sporadic phenomenon, figure 4 plots the mean values of the first three tones for each clause (H_{1j}, L_{1j}, H_{2j}) in SS gapping and coordination sentences from speaker Y's data. The crucial observation is that in the nonfinal gapped 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 not observed in the final clause where the verb is not elided (= H_{23}).

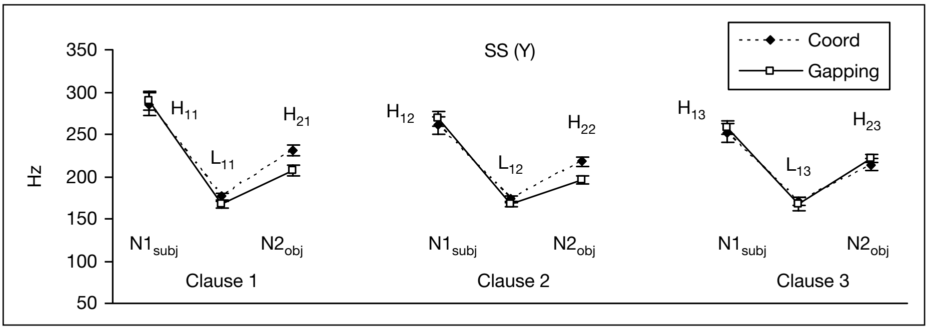


Fig. 4. Means of H_1, L_1, H_2 for each clause. Data from speaker Y.

The same tendency appears in the SL, LS, and dative conditions, as illustrated in figure 5: *all clause-final* H* tones in nonfinal gapped clauses are systematically lower than those in the corresponding coordination sentences. That is, SS, SL, LS and dative gapping all behave alike: branching turns out to play no role in this regard.

The difference between gapping and coordination is quite general, it is observed across all the speakers. We calculated the differences between the H* peak on final objects and the immediately preceding peak for all conditions. The result, shown in figure 6, shows that in the first two clauses, the differences are systematically larger in the gapping condition than in the coordination condition, indicating that the H*s on final objects are indeed lower in gapping than in coordination.

We conducted an ANOVA with two independent variables, CLAUSE (1st, 2nd, 3rd) and TYPE (coordination and gapping). There were significant main effects for both variables [CLAUSE: $F(2, 392) = 50.551$, $p < 0.001$, TYPE: $F(1, 196) = 179.965$, $p < 0.001$) and for the interaction [$F(2, 392) = 104.814$, $p < 0.001$).

Of most importance is the fact that the main effect of TYPE is significant, statistically confirming a difference between gapping and coordination. The interaction between TYPE and CLAUSE is significant because the difference observed in the first and second clauses is not present in the third clause. The results of post-hoc multiple comparison tests confirm this conclusion: the peak differences between coordination and gapping are reliable in the first and the second clause [C1: $t(393) = 8.073$, $p < 0.001$, C2: $t(393) = 8.610$, $p < 0.001$], but the difference is not present in the final clause [$t(393) = 0.722$, $p = 0.471$]. In fact, the interaction effect disappeared when we reran an ANOVA with only the data from the first and second clause ($F < 1$).

4.1.2 Analysis

We have seen that in nonfinal clauses, clause-final H*s in gapping appear lower than corresponding non-clause-final peaks in coordination. To account for this observation, we propose that all clause-final H*s are lowered: this lowering targets the H*s of the final objects in gapping in nonfinal clauses (because the verbs are gapped), while it targets the H*s of the verbs in coordination. Since the final object peaks do not undergo this additional lowering in coordination, they appear higher than the final

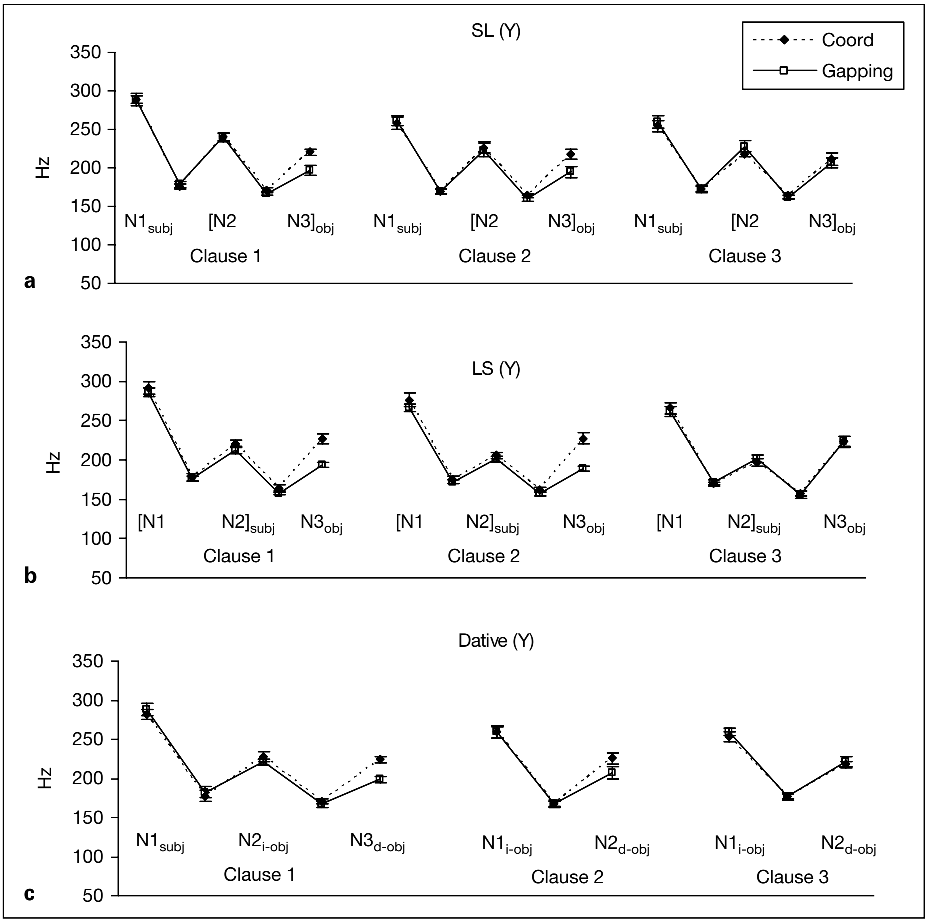


Fig. 5. Means of penultimate and final accent H tones and boundary L tones in each clause in the SL (**a**), LS (**b**), and dative (**c**) conditions. Data from speaker Y.

object peaks in gapping, which do undergo lowering.[6] The analysis is illustrated in figure 7, with lowering shown by thick arrows.

As illustrated in figure 7, postulating clause-final lowering accounts for the difference between gapping and coordination. (The lowering analysis illustrated in figure 7 predicts that initial rises on verbs in coordination sentences undergo final lowering as well because they are clause-final. We show in § 5.2 that in fact they do.) Next, we may ask which prosodic level defines the domain of the final lowering.

[6] We assume that there always exists a VP boundary between the subjects and the objects in gapping sentences [e.g. Abe and Hoshi, 1997] as well as in coordination sentences, and hence the left edge of the objects are always right-aligned with a MaP edge; i.e. the objects in gapping and coordination are prosodically comparable in their left edges.

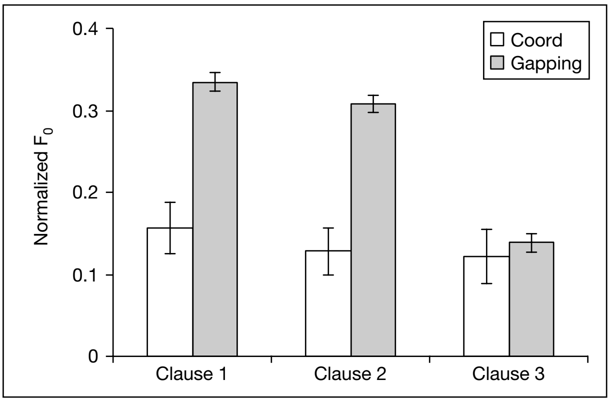


Fig. 6. Means of the normalized F_0 differences between penultimate and final peak values for coordination and gapping. Based on data from all speakers.

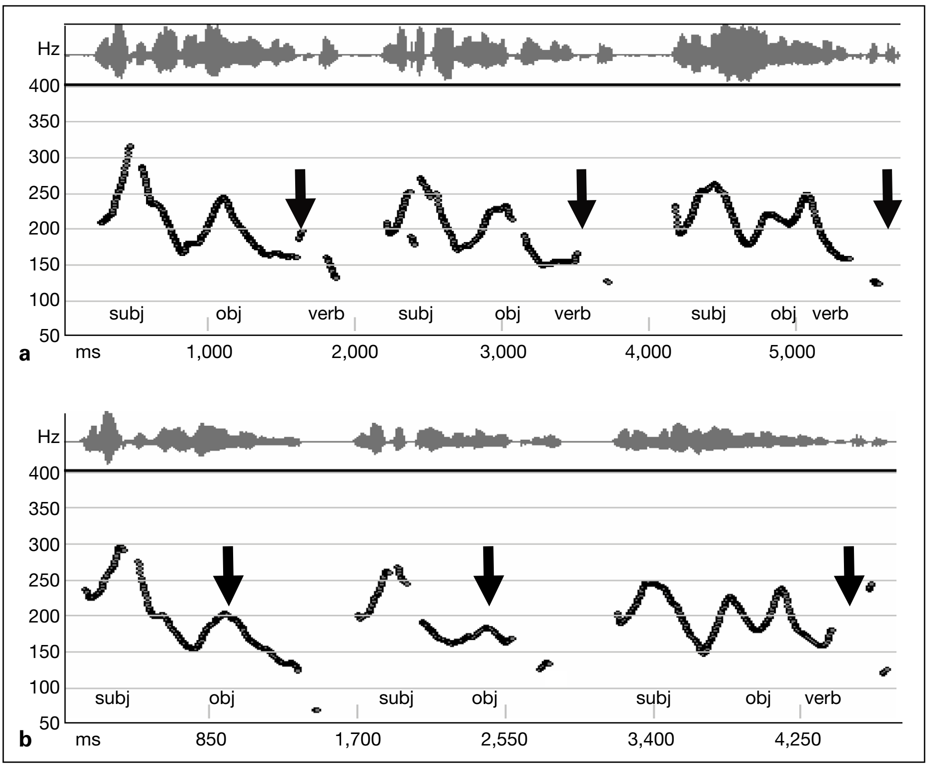


Fig. 7. Illustration of final lowering, indicated by thick arrows. Pitch tracks reproduced from figure 3.

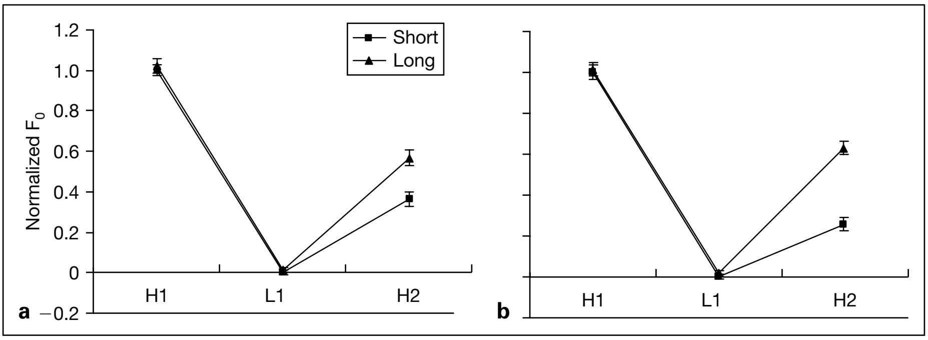
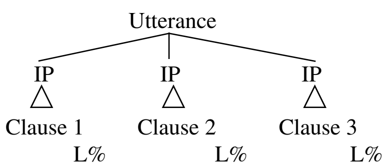


Fig. 8. Normalized means of the first three tones in the predicative (**a**) and intransitive (**b**) constructions. H_1 corresponds to the peak of the subject (both for predicatives and intransitives), L_1 to the initial valley at the predicate word (for predicatives) or the intransitive verb (for intransitives), and H_2 to the peak of the predicate word or the intransitive verb.

The domain for final lowering cannot be the Utterance, because the end of each clause does not necessarily correspond to the end of a whole utterance. It cannot be the MaP either, as the lowering has no motivation at MaP-final positions[7] (these arguments are developed in more detail in the following subsections). We thus postulate a level above the MaP and below the Utterance which can define final lowering. We call this level the *Intonational Phrase* (IP). In the case at hand, each syntactic clause projects its own IP.[8] Furthermore, all IPs are incorporated into a higher prosodic level, the Utterance, which corresponds to an entire sentence in syntax, as illustrated in (10) (see § 5 for more on the Utterance).

To account for the lowering effect, we posit a L% boundary tone associated with the right edge of each IP, as in (10), which causes lowering of the IP-final Hs [cf. Xu, 1994, 1997; see § 6.2 for a formalization of this tonal lowering].

(10)



The lowering is not construction-specific: it emerges in the predicative sentences and intransitive sentences as well. When items hosting a H* are located in clause-final positions, the values of the H*s appear lower compared to those of H*s in nonfinal positions. Figure 8 shows the F_0 targets of the three tones for the predicative and intransitive

[7] It is possible that final lowering does apply MaP-finally, but the degree of lowering is smaller MaP-finally than IP-finally. To test whether final lowering applies MaP-finally, we would need to compare rises in MiP-final positions and those in MaP-final positions, but our experiment was not designed to make this comparison.

[8] The view that a syntactic clause corresponds to an IP accords with the observations in other languages that the IP corresponds to the so-called 'comma intonation', which is usually followed by a pause [Bing, 1979; Nespor and Vogel, 1986; Potts, 2003; Selkirk, 2005]. This generalization also holds in Japanese: each clause is orthographically separated

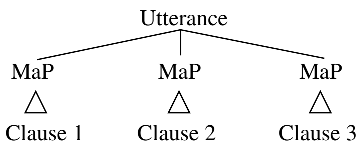
sentences; H_1 (= the H* of the subject), L_1 (= L%) and H_2 (= the H* of the predicate word for predicative sentences and the H* of the verb for intransitive sentences). H_2 is sentence-final in the short conditions and nonfinal in the long conditions.

As observed in figure 8, the H_2 values are lower in the short condition where H_2 is clause-final. A statistical analysis shows that the F_0 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 [predicative: $t(49) = 9.254$, $p < 0.001$, intransitive: $t(97) = 11.236$, $p < 0.001$]. The difference shows that final lowering occurs not only in multi-clause sentences but in clause-final positions in general (see § 5.2 for evidence that such lowering manifests itself via lowering of H*+L peaks on verbs in coordination sentences).

4.1.3 Alternative Analyses

In the proposed prosodic structure formalized in (10), each clause constitutes an IP. Alternative analyses that do not involve an IP fail to properly account for the patterns of the lowering data. One analysis would assume a structure in which each clause constitutes a MaP, as shown below [repeated from (4b)]:

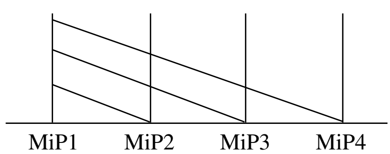
(4b)



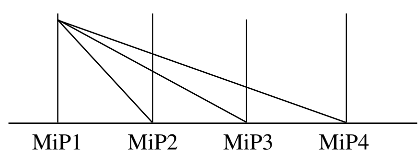
Given the structure in (4b), as an anonymous reviewer pointed out, one could posit that the peak height differences between the coordination and the gapping sentences come about due to the difference in the number of MiPs that each clause contains: the coordination clauses always contain one more MiP than the gapping clauses, and the peak values could be lower in the gapping clauses because they contain one less MiP.

Indeed, some previous studies observe that the heights of F_0 peaks correlate with constituency lengths, and the correlation can manifest itself in two different ways [Grabe, 1998; Prieto et al., 2006; see also Cooper and Sorensen, 1981; Rialland, 2001; Selkirk et al., 2004]. First, the initial F_0 peak can get higher as the number of MiPs increases, as in (11a). Alternatively, the F_0 declination slope can get shallower as the number of MiPs increases, as shown in (11b).

(11a)



(11b)



by a comma and phonologically by a pause (§ 4.3). In particular, that the IP corresponds to each conjunct in gapping and coordination jibes well with Selkirk's [2005] claim that the IP corresponds to a syntactic Comma Phrase, a projection headed by a [+comma] feature; a Comma Phrase can consist of syntactic clauses, parentheticals, nonrestrictive relative clauses, appositives and others [see Potts, 2003, for the semantic contribution of the [+comma] feature].

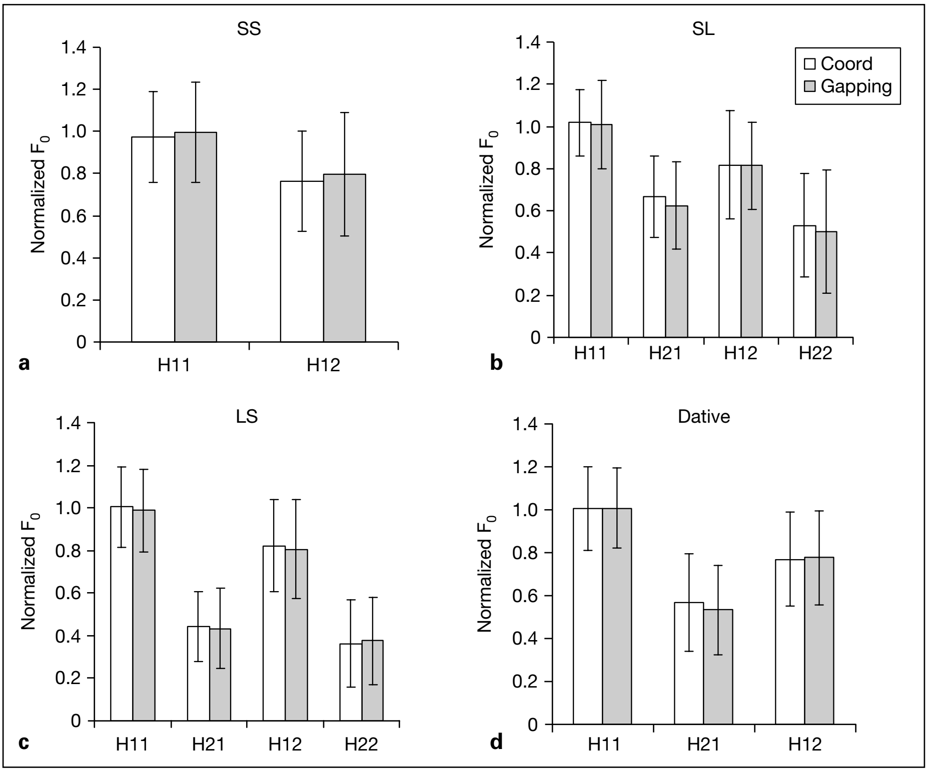


Fig. 9. Means of the non-clause-final peaks in coordination and gapping sentences for the first and the second clauses.

In both of the models in (11), phrases with fewer MiPs are predicted to have lower F_0 peak values on MiPs that are in the same position relative to the beginning of the MaP. For example, let us consider the case of the SL condition where gapping contains three MiPs and coordination contains four MiPs. The final peak in the gapping condition [at MiP3 in (11)] would be realized lower than the corresponding peak in the coordination both in (11a) and (11b).

However, both analyses incorrectly predict that F_0 peak differences should be observed not only at final peaks, but also at nonfinal peaks. Again we can turn to the SL condition to illustrate the prediction. In both scenarios shown in (11), it is predicted that there should be a difference between gapping and coordination in the first and second MiPs and the third MiP. Additionally (11a) predicts a difference in the first MiP. As seen in figure 5, the prediction does not hold – differences emerge only in final peaks.

To confirm that no differences exist in nonfinal peaks, figure 9 shows the non-clause-final F_0 peak values between the coordination and gapping sentences in all four sentence types. In the SS condition, the nonfinal clauses consist of three MiPs in coordination and two MiPs in gapping, and thus the first peak values are presented for the first and second clauses (fig. 9a). In the SL and the LS conditions, each nonfinal clause contains four MiPs in coordination and three MiPs in gapping, and hence the first two

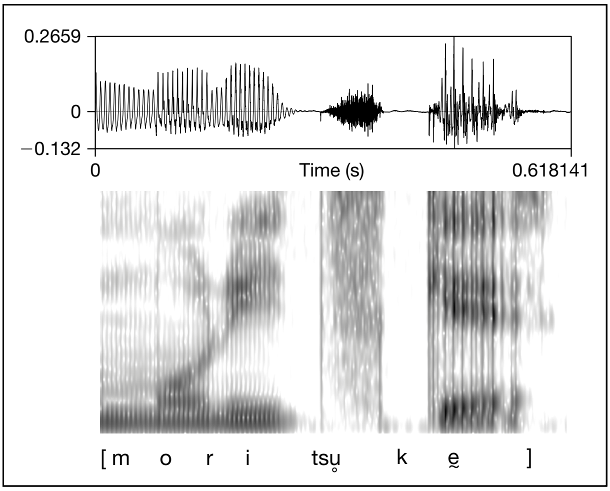


Fig. 10. The waveform and spectrogram of the final verb *moritsuke* 'put on dish' in the first clause of the coordination sentence in the SS condition (speaker J).

peaks from the first two clauses are shown (fig. 9b, c). In the dative condition, the coordination sentences had four MiPs and three MiPs in the first and the second clauses, respectively, while the gapping sentences had three MiPs and two MiPs. Therefore, the two nonfinal peak values of the first clause and the one nonfinal peak value of the second clause are shown for the dative condition (fig. 9d).

As shown in figure 9, there are no substantial differences between the coordination and the gapping sentences in any of the comparisons [SS: $F(1, 97) = 1.274$, $p = 0.262$, SL: $F < 1$, LS: $F < 1$, dative: $F < 1$).[9] These results show that lowering applies only at the end of clauses – i.e. IP-finally. To summarize, a level above the MaP – the IP – gives us the only possible domain for final lowering. In what follows, we provide more evidence for the distinction between the IP and the MaP.

4.2 Distribution of Creaky Vowels

The second characteristic of the IP is that vowels become creaky IP-finally, but not MaP-finally.[10] Figure 10 provides a waveform and spectrogram of the verb *moritsuke* 'to put on dish' in the first clause of the SS coordination sentence pronounced by speaker J. As illustrated in figure 10, the clause-final vowel shows irregular glottal pulses as well as an excitation of energy in the high frequency range. Irregular glottal pulses can be observed in the waveform as well.

[9] The results for the other main effect (DIFFERENT TONES) and its interaction with the coordination-gapping difference are not reported here since they do not relate to the current inquiry.
[10] Cross-linguistically, creaky voice is often associated with L tones [Gordon and Ladefoged, 2001]. The correlation does not necessarily mean, however, that creaky voice is an automatic consequence of a boundary L%, as other low tones in Japanese (such as +L in an accent H*+L tone) do not cause creaky voice. Therefore, creaky voice should be considered as an independent phonetic correlate that signals the IP boundaries, rather than an automatic consequence of L%.

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

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

The fact that creaky vowels systematically appear in clause-final positions provides further evidence that such positions cannot be defined as MaP boundaries, because creakiness is rarely if ever observed at the end of the MaP [see Kim et al., 2006, for a parallel pattern in English]. To confirm this generalization, we counted the frequency of creaky vowels in (i) the subject particle *wa* (and in the dative condition, the dative particle *ni*) (= MaP-final positions) and (ii) clause-final vowels in gapping and coordination sentences (= IP-final positions). Since we did not control for vowel quality in these two positions, a quantitative analysis based on spectral slices was impossible. Instead, we relied on auditory impressions and the known acoustic correlates of creaky vowels. Vowels were judged as creaky if they showed an irregular waveform as well as an excitation of energy in the high frequency range. Sometimes only a later portion of a vowel showed creakiness, in which case we judged it 'semicreaky'.

Table 1 summarizes the results – creaky vowels rarely appear MaP-finally, but they are very common IP-finally, especially for speakers J and Y.

To check the reliability of our identification of creaky vowels, 4 phonetically-trained native speakers of Japanese were recruited, all naïve as to the purpose of the experiment. They were asked to judge the creakiness of vowels in 40 sentences selected at random (one tenth of the whole data set). The transcribers were asked to judge creakiness based on their auditory impression, with the aid of wave forms and spectrograms for spotting the irregularity and excitation of energy in the high frequency range. They were told to classify a vowel as 'noncreaky' if its entire portion was in modal voice, 'creaky' if the entire portion was creaky, and 'semicreaky' if only a later portion was creaky.

Overall, there was a reliable consistency in the judgments of creakiness. The transcribers varied in their distinctions between 'creaky' and 'semicreaky' vowels, presumably because they interpreted 'only a later portion' differently. But, if we abstract away from the difference between 'creaky' and 'semicreaky', treating vowels as creaky if they received judgments of at least partially creaky, then the percentage of tokens for which all transcribers (including the 4 recruited transcribers and the 2 authors) agreed was 94.8% for the vowels in MaP-final positions, and 83% for the vowels in IP-final positions.[11]

[11] The agreement is poorer in the ratings of creakiness in IP-final positions than in MaP-final positions, which indicates that the raters did not consider vowels as creaky unless the vowels are clearly creaky, i.e. they are biased toward judging the vowels as noncreaky. Nevertheless, the extent of agreement remains high in both positions.

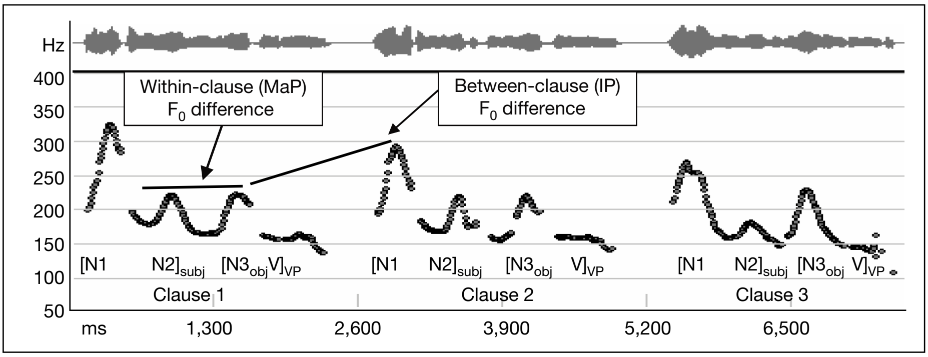


Fig. 11. The F_0 difference between two H* tones in two conditions: (i) the within-clause condition and (ii) the between-clause condition. The pitch track is taken from a coordination sentence in the LS condition uttered by speaker R.

4.3 Obligatory Pauses at IP-Final Positions

The third piece of evidence for the distinction between the IP and the MaP comes from the presence/absence of a pause in IP-final positions. As seen in the pitch tracks in figure 7, a pause is obligatory in clause-final positions: our speakers always inserted a pause after each clause (= IP-final positions). On the other hand, as seen in figure 7, a pause is never obligatory after subject-NPs, i.e. in MaP-final positions (recall from § 3 that the left edge of a VP coincides with a left edge of MaP). In careful speech, it might be possible to insert a pause in MaP-final positions, but a pause is never obligatory in these positions. At the very least, the distinction between the MaP and the IP is clear to the extent that our speakers rarely inserted a pause MaP-finally, but always inserted a pause IP-finally.

4.4 Pitch Reset within and between Clauses

The fourth property that distinguishes the IP from the MaP is the degree of pitch reset. Given two successive H tones, we compared the F_0 difference across a MaP boundary and the F_0 difference across a clause boundary. Our hypothesized IP boundaries should cause stronger pitch reset across a clause boundary compared to a MaP boundary, as a higher prosodic edge induces more robust pitch resetting [Ladd, 1988; see also § 3].[12] To quantitatively test the prediction, we compared F_0 differences in two environments: (i) the difference between H_{21} and H_{31} (a within-clause difference across a MaP boundary) and (ii) the difference between H_{31} and H_{12} (a between-clause difference across an IP boundary), as illustrated in figure 11.[13]

[12] Thanks to Hubert Truckenbrodt for suggesting this analysis to us.
[13] The other conditions were not used to avoid the Utterance-initial H*, which is boosted by domain-initial strengthening (see § 5.1).

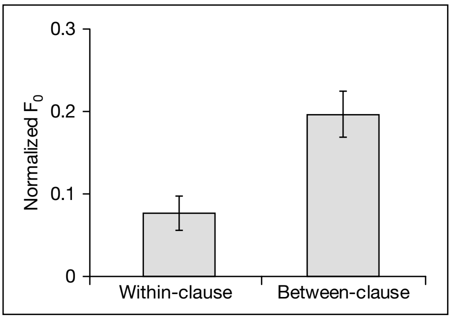


Fig. 12. Means of the normalized F_0 peak differences in the within-clause condition and the between-clause condition.

As observed in figure 11, the extent of pitch reset is larger in the between-clause condition than in the within-clause condition. Figure 12 shows the mean F_0 difference in the two conditions: the difference in F_0 is indeed larger in the between-clause condition than in the within-clause condition, and the difference is statistically significant [$t(197) = 5.363$, $p < 0.001$]. The difference in pitch reset supports the hypothesis that the between-clause condition involves an IP boundary whereas the within-clause condition involves only a MaP boundary.

4.5 Initial Rises in IP-Initial and MaP-Initial Positions

Yet another phenomenon distinguishes an IP boundary from a MaP boundary. A number of previous works have shown that constituent initial rises – rises from L% to the following H – are cross-linguistically larger at IP edges than at MaP edges [Ladd, 1988, 1990; Selkirk, 2005; Truckenbrodt, 2002]. If clause boundaries involve IP boundaries, then clause-initial rises should be larger than VP-initial rises, which coincide with MaP boundaries.

To test this prediction, we compared the F_0 rises at the beginning of the VP in the SL and dative conditions ($H_{21}-L_{21}$)[14] and the clause-initial rises of the second clause[15] of the same sentences ($H_{12}-L_{12}$), as illustrated in figure 13. The first arrow represents a MaP-initial rise, while the second arrow represents an IP-initial rise.

Figure 14 illustrates the results of the comparison. The F_0 rises are greater at the left edge of the second clause (= IP-initial rises) than at the left edge of the VP (= MaP-initial rises) [$t(198) = 2.890$, $p < 0.01$].

The 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 narrows over the course of an utterance (see § 5.1), everything else being equal, it is expected that rises would be smaller in the second clause than in the first clause. The fact that the clause-initial rises in the second clauses are larger than the VP-initial rises in

[14] In the other conditions, the left edge of the VP is clause-final in the gapping condition, and the rises therefore undergo final lowering. For this reason, we excluded these cases from the comparison.
[15] Utterance-initial rises were not used for comparison, as they show effects of Utterance-initial boosting (see § 5.1).

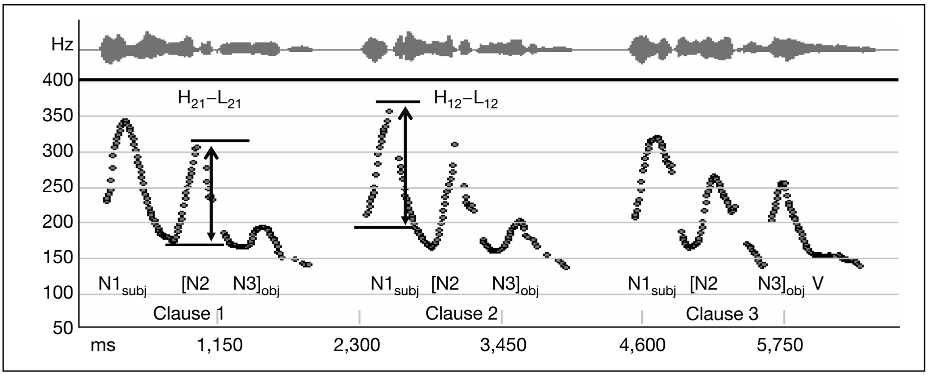


Fig. 13. A VP-initial rise (= MaP-initial rise) compared to a clause-initial rise (= IP-initial rise). The pitch track is taken from the SL condition uttered by speaker R.

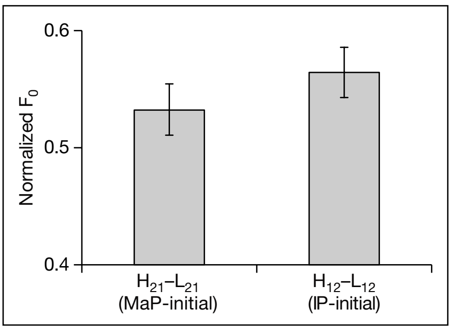


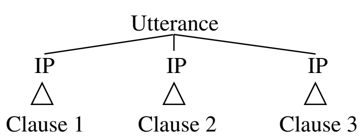
Fig. 14. Means of the normalized VP-initial rises ($H_{21}-L_{21}$, MaP-initial rise) and clause-initial rises (H_{12}–L_{12}, IP-initial rise).

the first clauses, even given the effect of declination, indicates that clause-initial rises are inherently larger than MaP-initial rise.

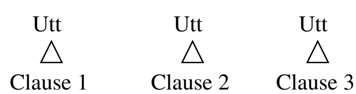
5 The Distinction between the Utterance and the IP

In § 4, we proposed the structure (4a), which we repeat below. In (4a), a syntactic clause projects its own IP, while an entire sentence projects an Utterance. We now show that we must distinguish between the IP and the Utterance. The alternative analysis that we address in this section is shown in (4c) where each clause is an independent and separate Utterance.

(4a) Proposed structure (repeated)



(4c) Alternative structure (repeated)



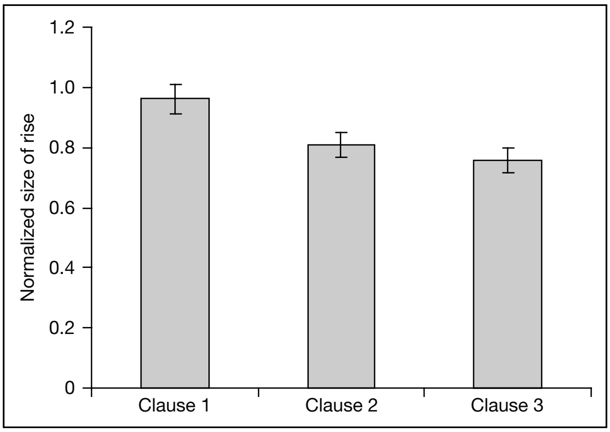


Fig. 15. Normalized means of initial rises in each clause.

The alternative deserves serious attention because at first glance, in gapping and coordination sentences, fairly long pauses separate each clause (§ 4.3), and each clause has a strong pitch reset at its beginning (§ 4.4).

Positing each clause as projecting its own Utterance predicts that all the clauses would show the same prosodic pattern. We show that, in fact, initial and final clauses exhibit unique characteristics, and their patterns can only be explained if each clause is incorporated within a higher prosodic category, as in (4a).[16] Evidence for the distinction between the IP and the Utterance in (4a) comes from the comparison of clause-initial rises (§ 5.1), the patterns of clause-final verbal rises (§ 5.2), and the appearance of a H tone in Utterance-final positions (§ 5.3). Through the discussion of these properties, we also identify several characteristics of the Utterance.

5.1 Initial Hs, Initial Rises, and Declination

First, the behavior of clause-initial Hs supports the structure (4a). Previous researchers have observed that higher prosodic domains are signaled by higher initial rises [Cooper and Sorensen, 1981; Ladd, 1988, 1990; Selkirk, 2005; Selkirk et al., 2003; Truckenbrodt, 2002]. We have seen that IP-initial rises are larger than MaP-initial rises in Japanese (§ 4.5; fig. 14). If the three clauses in multi-clause constructions are incorporated within a higher prosodic level as in (4a), then we expect clause-initial Hs to have a higher pitch in the first clause than in the second and third clauses. Figure 15 compares initial rises in the three clauses based on data pooled together from gapping and coordination sentences.

The size of the 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

[16] One variant of (4c) analyzes each Utterance as part of a higher recursive Utterance. But doing so results in a structure congruent to (4a), because the IP would equate to the Utterance and the Utterance to 'the higher Utterance'. Crucially, two levels exist above the MaP – one that incorporates each clause, and another that incorporates an entire sentence.

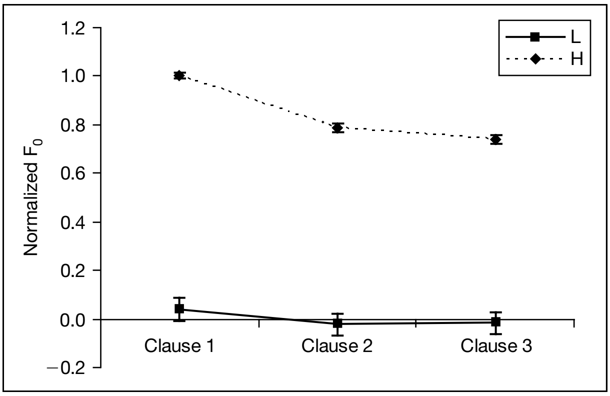


Fig. 16. Normalized means of clause-initial H* and L% tones.

initial rises with CLAUSE as the independent variable reveals a significant main effect [$F(2, 788) = 189.439$, $p < 0.001$]. The post-hoc multiple comparison tests show that all the initial rises in the three clauses differ significantly from one another [C1 vs. C2: $t(394) = 13.731$, $p < 0.001$, C1 vs. C3: $t(394) = 17.356$, $p < 0.001$, C2 vs. C3: $t(394) = 4.985$, $p < 0.001$]. Even though the difference between the second and third clause is statistically significant, it is markedly smaller than the difference between the first clause and the second clause [$t(394) = 5.857$, $p < 0.001$]. We conclude from these results that the first clauses have some special status in that they show particularly large initial rises. The difference between the second and third clause can be attributed to declination, to which we now turn.

Figure 16 plots the mean of clause-initial H peaks and clause-initial L valleys. Focusing first on the patterns of H*s, we can draw two generalizations: (i) the values of H* decline from the first clause to the third clause, but (ii) the slope between the first clause and the second clause is greater than the slope between the second and the third clause.

An analysis based on paired t tests statistically supports these generalizations: the H* tones appear significantly higher in the first clause (H_{11}) than in the second clause (H_{12}) [$t(394) = 26.667$, $p < 0.001$], and H_{12} similarly appears higher than H_{13} [$t(394) = 5.962$, $p < 0.001$]. Further, the F_0 slope differences are greater between the first and second H* (H_{11}–H_{12}) than between the second and third H* (H_{12}–H_{13}) [$t(394) = 13.741$, $p < 0.001$].

The behavior of the H* tones follows from two factors. First, as observed in many languages, F_0 gradually declines over the course of an utterance [Cooper and Sorensen, 1981; Fujisaki and Hirose, 1984; Liberman and Pierrehumbert, 1984; Maeda, 1976; Pierrehumbert, 1980; Poser, 1984; 'tHart and Cohen, 1973; Thorsen, 1980]. The declination effect explains the decrease in pitch between the second and third clauses, but does not explain why the difference between H_{11} and H_{12} is larger than the difference between H_{12} and H_{13}. Therefore, in addition to declination, we argue that Utterance-initial H*s are boosted by domain-edge strengthening [Cho and Keating, 2001; Fougeron and Keating, 1997; Hayashi et al., 1999; Hsu and Jun, 1998; Keating, 2003; Keating et al., 2003; Onaka, 2003]. Consequently, the F_0 differences are larger between the first and second clauses than between the second and third clauses. Additionally, the

Utterance-initial boosting explains the earlier observation that the first clauses exhibit the largest initial rises.

The above observations provide evidence for the Utterance that contains the three clauses in the multi-clause constructions. First, the Utterance defines the domain of declination, as the H values steadily decline from H_{11} to H_{13}. Second, the Utterance shows distinctively large initial rises at its left edge. These effects cannot be explained by a structure like (4c), because it has no level that defines a domain of declination, nor can it capture the boosted sentence-initial Hs.

Turning our attention to the patterns of L tones, as shown in figure 16, their slopes are less steep than the slopes of corresponding H*s. We can nonetheless see that the L tone in the first clause is higher than the following L tones. We carried out t tests to compare the L tones in the three clauses. The results show that L_{11} is significantly higher than L_{12} [$t(394) = 6.928$, $p < 0.001$], but show no significant difference between L_{12} and L_{13} [$t(394) = 0.801$, $p = 0.424$]. Evidently, L tones are boosted at the beginning of the Utterance just like H tones, but they are not subject to declination.

5.2 Verbal Rises

We find yet more evidence for the difference between the IP and the Utterance in the size of initial rises on verbs in coordination sentences. We argued in § 4.1 that clause-final rises undergo final lowering – in coordination, therefore, verbal rises in each clause undergo lowering. (4c) predicts that the amount of final lowering would be equal across all the clauses. On the other hand, our proposed model in (4a) allows for the possibility that final lowering is strongest in Utterance-final positions, just as initial boosting of H*s is strongest in Utterance-initial positions. Our model (4a) makes the right prediction.

To test the predictions of the two models, we measured pitch values of the first and the second mora of the verbs in each clause in coordination. We usually observe a rise in MiP-initial positions, but downtrend and domain-final lowering effects can obscure those rises. Thus we measured pitch during the steady state of the first and the second vowels. As noted in § 4.2, IP-final vowels are creaky, occasionally leaving us with no way to measure F_0; therefore we only accepted into comparisons those tokens whose pitch contour was measurable. Speaker J consistently showed heavy creakiness at Utterance-final positions, which often extended all the way to the beginning of the last word in Utterance, and hence none of her data figured into the analysis.

Figure 17 summarizes the results. The third clause noticeably differs from the first and second clauses. The values are positive in the first and second clauses, representing rises in these positions, though they are very small. On the other hand, the third clause has negative values.

A repeated-measures ANOVA on the normalized values of verbal rises with CLAUSE as the independent variable reveals a significant main effect ($F(2, 82) = 54.170$, $p < 0.001$]. Post-hoc multiple comparison tests show that the verbal rises in the three positions significantly differ from one another [C1 vs. C2: $t(41) = 3.833$, $p < 0.001$, C2 vs. C3: $t(41) = 8.959$, $p < 0.001$].

We maintain that the categorical difference between the final clause on the one hand and the first and second clause on the other is due to the final lowering that we

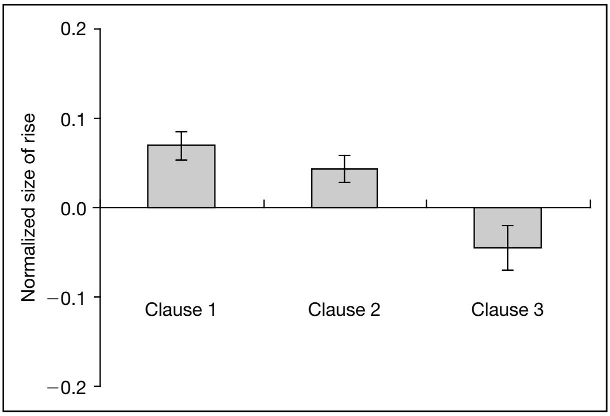


Fig. 17. Normalized means of verbal rises in each clause in coordination sentences.

identified in § 4.1[17]: the lowering effect, generally observed in IP-final positions, is further enhanced in the third clauses by a domain-edge strengthening effect specific to Utterance-final positions. A boundary L% tone associated with the IP substantially lowers the pitch of the second mora of the verb in Utterance-final positions, practically obliterating the pitch accent. Analyzing each clause in multi-clause constructions as constituting a separate Utterance fails to account for the unique behavior of the final clauses.

5.3 Utterance-Final H Tone

As final evidence for the structure in (4a), we show that the Utterance has a H tone associated with its penultimate MiP. The H tone's exact docking site varies between speakers, but all the speakers exhibit it. The presence of the H tone, which appears Utterance-finally but not IP-finally, provides further grounds for a distinction between the Utterance and the IP.[18]

The H tone emerges most visibly in the speech of speakers N and Y, who associate the H tone with the accusative case particle -*o* for the preverbal object. For these speakers, the H tone manifests itself as an extra H peak, shown in figure 18 where the final object shows two clear H peaks. Figure 19 illustrates the alignment of this extra H, which coincides with the accusative case particle-*o*.

[17] An anonymous reviewer pointed out that the repetition of the same verb in coordination sentences might explain the lack of rises in the final clauses – the peaks of informationally given items are lower than those of informationally new items. This explanation for the lack of rises in the final clauses however does not account for the fact that the peaks in the third clauses show a behavior different from the peaks in the first and second clauses; it instead predicts that the peaks in the first clauses (interpreted as new information) would behave differently from the peaks in the second and third clauses (interpreted as old information).

[18] The H tone in question is not specific to gapping and coordination sentences, but appears also in single-clause sentences of exceptional length. In filler single-clause sentences which had constituency lengths comparable to those of the gapping and coordination sentences (5 g), the speakers show the final H tone optionally (cf. the H tone consistently appeared in gapping and coordination). The distribution of the H tone in long single-clause sentences requires further research.

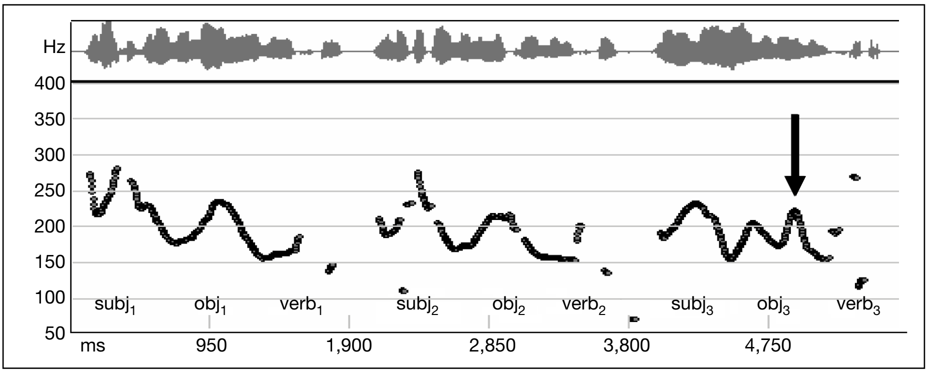


Fig. 18. A representative pitch track of a coordination sentence uttered by speaker Y, showing a sentence-final H tone docking onto a case particle of the final object (marked with arrow).

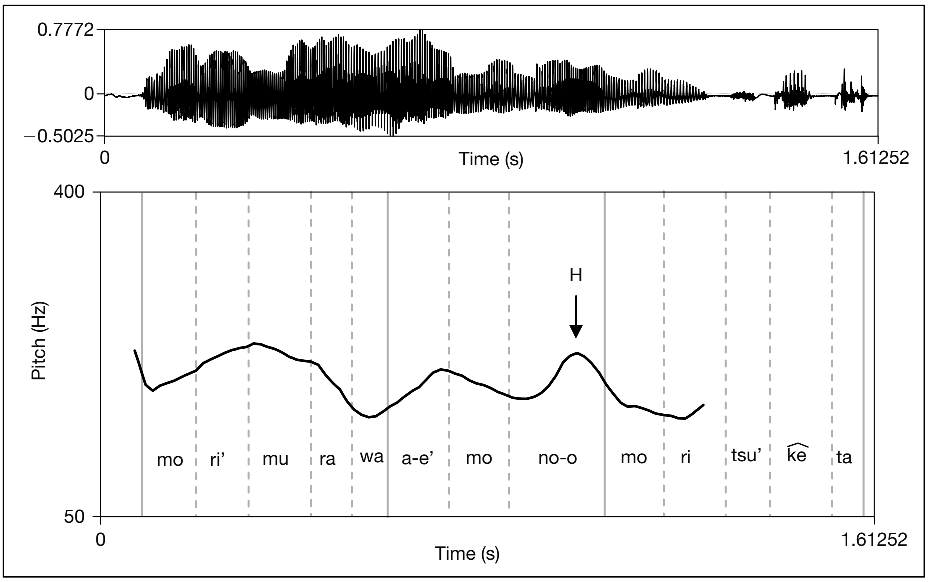


Fig. 19. Pitch track for the third clause of the coordination sentence shown in figure 18 with the utterance-final H tone docking onto the case particle *-o*.

For speakers J and R, the H tone docks onto the accented mora of the word that immediately precedes the verb, boosting the preverbal rise. We can tell that the speakers do so because the peaks of the final object in the final clause have higher values than the corresponding peaks in the second clause. Figure 20 shows a representative example, taken from the SS condition of speaker J.

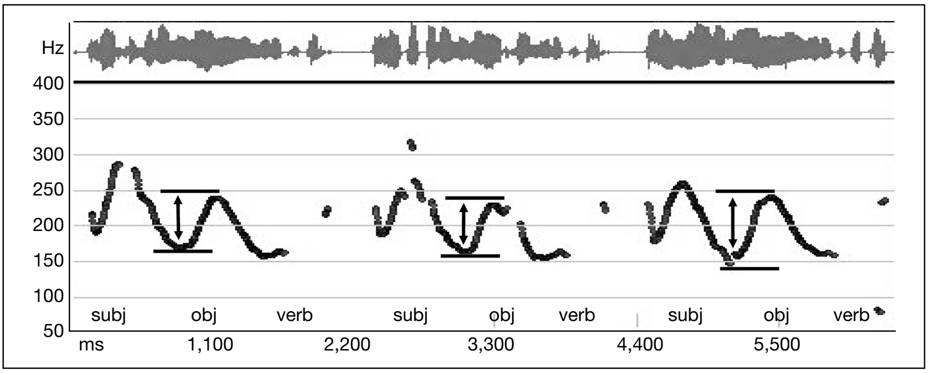


Fig. 20. A representative pitch track of a SS coordination sentence uttered by speaker J, illustrating boosting of the peak of the object in the final clause.

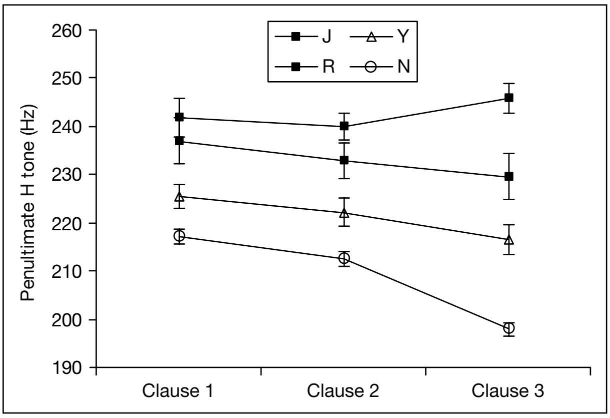


Fig. 21. Means of the F_0 peaks of the final objects for each clause in coordination.

The H peak for the object is higher in the final clause than in the second clause. We interpret the high F_0 peak in the final clause as having been boosted by an Utterance-final H tone. Recall the general declination effect where the H tone values decline over the course of an utterance (§ 5.1). The fact that the third H peak appears higher than the second H peak conflicts with what we would expect given only declination, and thus indicates the work of an extra boosting mechanism.

To establish that speakers J and R boost the object peaks of the final clause, while speakers N and Y do not, Figure 21 shows each speaker's preverbal H*s for the three clauses in coordination. Here we present values for each individual speaker, rather than pooled normalized values, because the individual differences concern us. We can see that the third H tone values are higher than the second H tone values for speaker J; F_0 declines from the second H to the third H to some extent for speaker R, but the extent of the declination is weaker than what we observe for speakers Y and N.

We used a mixed-design ANOVA with SPEAKER (between-subject) and CLAUSE (within-subject) as the independent variables to assess these patterns. A significant effect was found for both SPEAKER ($F(3, 193) = 71.093$, $p < 0.001$] and CLAUSE [$F(2, 386) = 55.514$, $p < 0.001$] as well as for the interaction [$F(6, 386) = 23.770$, $p < 0.001$]. The post-hoc tests comparing the second H and third H support our observations: the third H tone is significantly higher than the second H tone for speaker J [$t(46) = -4.372$, $p < 0.001$]. For speaker R, H in the third clause does not differ significantly from the second H tone [$t(49) = 1.784$, $p = 0.081$], which indicates that the third Hs are enhanced in their rises. Recall that given declination, we expect the third Hs to appear lower than the second H, but in reality the third Hs are no lower than the second Hs. Thus, the pattern of speakers J and R shows that the third F_0 peaks must be raised by the extra H tone, which cancels out the general effect of declination.

The assumption that a declination effect acts on preverbal rises is supported by the post-hoc comparison tests for speakers N and Y in which the third H tones are significantly lower than the second H tones [speaker N: $t(51) = 17.595$, $p < 0.001$, speaker Y: $t(47) = 3.467$, $p = 0.001$]. Declination is visible for these speakers because Utterance-final Hs do not boost the H* tones in the third clauses but instead occur later, on the case particle.

The variation of the H tones can be phonologically represented as in (12):

(12) (a) Speakers J and R (cf. fig. 20) (b) Speakers N and Y (cf. figs 18, 19)

(a)
H*+L H
NP_{obj} -*o*

(b)
H*+L H
NP_{obj} -*o*

As illustrated in (12a), 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 F_0 peak. Speakers N and Y on the other hand associate the H tone with a case particle, as illustrated in (12b). The patterning of the H tone is consistent within each speaker; the two different phonological structures in (12) account for the intraspeaker consistency as well as the interspeaker variability.

Regardless of how this H tone is phonetically realized, it appears only in final clauses.[19] The distribution of the H tone is inexplicable under the assumption that all clauses behave alike; only a level above the IP, the Utterance, can host the final tone. See § 6.3 for more discussion of this H tone.

5.4 Summary

The characteristics of the MaP, the IP and the Utterance identified throughout this paper are summarized in table 2.

[19] An anonymous reviewer pointed out that in spontaneous speech the distribution of the H tone might be freer. A rise-fall pattern is commonly observed on case particles [Maeda and Venditti, 1998; Maeda et al., 1998], especially in the speech style popular among young female speakers, and that these patterns are seen even non-sentence-finally. [This pattern is transcribed as L%HL% in the x-JToBI system in Maekawa et al., 2002.] The issue of whether this H in L%HL% is identical to what we observe here, and if so, why it shows different behaviors in lab speech (obligatory in long sentences, limited distribution) and spontaneous speech (optional, freer distribution) are interesting topics for future research.

Table 2. Phonetic correlates of each prosodic level in Japanese.

	MaP	IP	Utterance
(a) Final lowering	no	observed (§ 4.1)	stronger than IP (§ 5.2)
(b) Final pause	no	yes (§ 4.3)	(yes)
(c) Final creakiness	no	yes (§ 4.2)	yes (§ 4.2)
(d) Initial rises	larger than MiP[1]	larger than MaP (§ 4.4)	larger than IP (§ 5.1)
(e) Pitch reset	larger than MiP (§ 3)	larger than MaP (§ 4.5)	–
(f) Domain of declination	no	no	yes (§ 5.1)
(g) Final H	no	no	yes (§ 5.3)

[1]Though not covered in our study, Selkirk et al. [2003] show that initial rises are larger in MaP-initial positions than in MiP-initial positions [see also Selkirk and Tateishi, 1991].

6 Discussion

6.1 Implications for Intonational Phonology

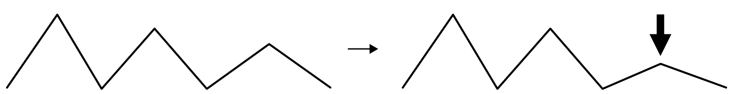
6.1.1 Summary: the Properties of the IP and the Utterance

We have so far investigated the intonational properties of gapping and coordination in Japanese, and cited several pieces of evidence for the existence of the IP and the Utterance in the prosodic organization of Japanese. Each syntactic clause projects an IP, which is characterized by tonal lowering, creakiness and a pause at its right edge. The IP also shows an initial rise and pitch reset distinctively larger than those found at MaP edges. The Utterance has properties distinct from the IP: it exhibits stronger final lowering and stronger initial rises than the IP, defines the domain of declination, and has a H tone which docks onto the penultimate MiP. In addition to motivating new levels in the prosodic hierarchy of Japanese, our findings have several implications for current theories of intonational phonology.

6.1.2 Lowering or Raising?

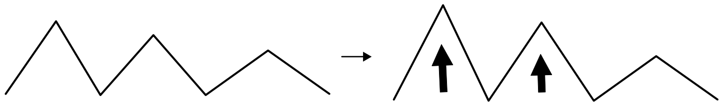
The first issue concerns how we formally capture 'lowering'. We postulated that Japanese has a lowering process IP-finally, but alternatively we could have postulated a raising process affecting nonfinal peaks. Truckenbrodt [2004] proposes that final lowering at a prosodic edge may be explained by the tonal raising rule *Raising before Downstep*, whereby a tone is raised with respect to the following tone by some fixed ratio in positions followed by downstep, as illustrated in (13b).[20]

(13) a. Lowering analysis (our proposal)



[20]Kubozono [1993] also proposed a raising effect at the left edge of a branching node. Our argument against a raising analysis applies to this raising mechanism as well.

b. Raising analysis [Truckenbrodt, 2004; see also Kubozono, 1993]



The raising rule accounts for apparent final lowering by raising all but the final H tones. As only the final H tone is exempted from the rule, it ends up at a lower value relative to preceding H tones. At first glance, IP-final lowering in our data seems to be accounted for equally well by the Raising-before-Downstep rule.

However, the raising analysis is not viable in light of all of our data. Specifically, we found that the absolute heights of verb-initial rises are lower in the final clause than in the nonfinal clauses, i.e., the degree of lowering is greater in Utterance-final positions than in Utterance-internal positions (§ 5.2). The pattern cannot be explained by a raising analysis because if final rises stay constant, as the raising analysis in (13b) predicts, we should find no difference between the absolute heights of Utterance-final H tones and those of non-Utterance-final H tones. Yet we find precisely that.

6.1.2 Universality of Prosodic Hierarchy

Second, the presence of the IP in Japanese has implications for the typology and universality of prosodic level inventories across languages. Our study has shown that the IP evidently plays a role in the intonational phonology of Japanese, though the previous literature found no evidence for it. The discovery of the IP in Japanese suggests that the prosodic hierarchy may be universal [cf. Jun, 2005]. It is beyond the scope of this paper to fully defend such a universalist view, but we hope that motivating the IP in Japanese phonology contributes toward confirming such a theory. To the extent that prosodic levels are defined in Universal Grammar, structures in one language should be present in every possible grammar; ceteris paribus, a theory that does not admit variation at the most fundamental level of linguistic organization is more restrictive.

6.1.3 Phonological Theories Can Inform Phonetic Research

Related to the preceding discussion, Selkirk [2005] advances a view that each prosodic level is a phonological reflex of a syntactic category (e.g. the MaP is a phonological reflex of XP; the IP is a phonological reflex of a syntactic clause, etc.). According to this position, if syntactic categories are universal, prosodic categories should also be universal. This phonological hypothesis informed and drove the current study: if the IP is found to correspond with a syntactic clause in many languages, then we should also find evidence for this level in Japanese. This sort of theory-driven research is fruitful to the extent that we have found evidence for the IP corresponding to a syntactic clause in Japanese. Theoretical considerations can provide a guideline as to what to look for in phonetic studies, and in turn phonetic researches can contribute to the development of phonological theory as well.

6.2 Lowering Mechanism

In § 4.1, we identified an F_0 lowering process at the end of the IP. However, we have not considered the precise mechanism behind the process. We turn to it now. Our proposal draws on the idea of phonetic tonal coarticulation, in which the H* tonal target in IP-final positions is undershot because of a following L% boundary tone [Xu, 1994, 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 in front of a boundary L% tone at the end of the IP, which is formalized in (14):

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

where r is a fixed ratio that determines the phonetic scaling of the tone. Since r is smaller than 1, the H*+L tone is realized lower than its original value. The mechanism in (14) is triggered only by a L% tone at the end of the IP, not by a tone at the end of any other prosodic levels, reflecting the fact that lowering occurs only IP-finally.

We must now account for the fact that the amount of lowering is greater at Utterance-final positions than at Utterance-internal positions (§ 5.2). Extending Ladd's [1988] proposal that clause-initial peaks are realized higher following a stronger boundary, we posit that a higher prosodic boundary is signaled by stronger domain-final lowering (i.e. smaller r). This postulation explains why the degree of F_0 lowering is greater at the end of the Utterance than at the end of the IP.

6.3 The Nature of Utterance-Final H Tones

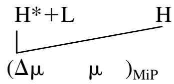
The nature of the extra H tone observed in Utterance-final positions (§ 5.3) deserves special attention. The tone exhibits the properties of both a starred tone and a boundary tone. As with other boundary tones, it shows an edge-oriented property (i.e. it associates with penultimate MiPs). On the other hand, the fact that it can appear on an accented mora indicates that it also shows properties of a starred tone.

The phrase accents observed in question intonation across a number of Eastern European languages show such mixed properties [Grice, et al., 2000]. In fact, Grice et al. [2000] claim that this ambivalence itself characterizes a phrase accent. For example, 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. The alignment of the phrase accent varies depending on the nuclear accent position. When the nuclear accent is on the final word, the phrasal tone associates with the final syllable, which is consistent with the properties of boundary tones. However, when the nuclear accent is not on the final word, the phrase accent docks onto the stressed syllable of the final word, behaving like a starred tone. The Utterance-final H tone in Japanese discussed in § 5.3 resembles this phrase accent – our finding lends support to Grice et al.'s [2000] account of phrasal accents.

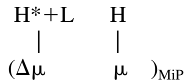
The next question is how to account for the distribution of the Utterance-final H tone in Japanese. A related question is how the interspeaker variability arises. Optimality Theory [Prince and Smolensky, 2004] can shed light on these issues.

Recall the two ways in which the H tone reveals itself phonetically: by boosting the penultimate accent H* or by docking onto the case particle of the preverbal element. Since the realization of the H tone differs between speakers, we posit two distinct phonological structures, depicted in (15a) and (15b) (see § 5.3). In the first pattern, H lodges onto the accented mora, as shown in (15a) where $\Delta\mu$ represents the head mora in a MiP [Δ for Designated Terminal Element, Liberman and Prince, 1977]. The second realization associates the H tone with a nonaccented mora, as illustrated in (15b):

(15a)



(15b)



The interspeaker variation of the H's realization makes sense from the perspective of cross-linguistic markedness: both ways of pitch-docking have structural configurations which are frequently avoided cross-linguistically. Optimality Theory [Prince and Smolensky, 2004] can account for the speaker variation at issue: the variable pattern of the H tone can be captured as the interactions of several conflicting structural (markedness) requirements.

First, in both patterns, the H tone docks onto the penultimate MiP, but never onto the final MiP. We can interpret the avoidance of associating the H tone to final positions as an effect of the well-known prohibition against final prominence, known as NONFINALITY [Hyde, 2003; Prince and Smolensky, 2004]:

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

Next, cross-linguistic studies show a preference for the H tone to associate with a head position within a constituent [Bickmore, 1995; de Lacy, 2002; Goldsmith, 1987; Selkirk, 2000].[21] Speakers J and R conform to this requirement: they associate the H tone to the head mora of the penultimate MiP, as in (15a). Let us express this requirement as Δμ-TO-HTONE:

(17) Δμ-TO-HTONE: A H tone is associated with the head mora of a MiP.

However, associating the H tone with the head of a MiP as in (15a) creates markedness problems. First, in this configuration, two distinct H tones are associated with one mora, but previous literature has shown that multiple linking of autosegments is structurally marked [Goldsmith, 1984; see Kawahara, 2007, for a recent summary]. The second problem is one of recoverability, which might be a consequence of multiple linking: in (15a), 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 already expect a rise in that position due to the lexical H*+L tone. A growing body of literature shows that such perceptual factors directly or indirectly shape phonological patterns [papers in Hume and Johnson, 2001, as well as Chitoran et al., 2002; Flemming, 1995; Silverman, 1995; Steriade, 1997]. How to formalize the effects of perceptibility has yet to be explored; for the sake of exposition, we use *MULTILINK/RECOV to express the markedness problems of (15a):

(18) *MULTILINK/RECOV: At most one H tone can be associated with a mora.

Optimality Theory [Prince and Smolensky, 2004] can model the complex behavior of the H tone with the interaction of the constraints above. Optimality Theory captures

[21] More generally, prominent elements are attracted to prominent positions [Aissen, 1999; de Lacy, 2002; Gouskova, 2003; Prince and Smolensky, 2004].

phonological patterns as arising from interactions of conflicting constraints, ranked in order of priority – an output does not usually satisfy all markedness requirements, but it instead best satisfies the requirements of highest constraint given a particular ranking hierarchy.

In Optimality Theoretic terms, the interspeaker variation results from the differences in the ranking of Δμ-TO-HTONE and *MULTILINK/RECOV. Speakers J and R conform to the ranking Δμ-TO-HTONE » *MULTILINK/RECOV, and therefore give precedence to the requirement that the H tone is associated with the head mora. Consider an OT tableau (19):[22]

(19) Speakers J, R

	NONFIN(MiP)	Δμ-TO-HTONE	*MULTILINK/RECOV
a. H*L H | | (Δμ-o)$_{MiP}$ ( )$_{MiP}$	*!		
b. ☞ H*L H | / (Δμ-o)$_{MiP}$ ( )$_{MiP}$			*
c. H*L H | | (Δμ- o)$_{MiP}$ ( )$_{MiP}$		*!	

Constraint violations are shown by asterisks. The constraints are ranked in order of importance from left to right (solid lines represent crucial rankings whereas dashed lines represent noncrucial ones). Candidate (a) violates NONFIN(MiP) because its H tone is associated with the final MiP, and is therefore ruled out – a fatal violation is shown by an exclamation mark.[23] Candidate (c) violates Δμ-TO-HTONE because the H tone lodges on something other than the MiP-head, and hence is also ruled out. The winning candidate is therefore candidate (b) (indicated by a pointing hand), which violates *MULTILINK/RECOV, but survives because *MULTILINK/RECOV is the least important constraint within the given constraint hierarchy.

On the other hand, speakers N and Y exhibit a hierarchy in which *MULTI LINK/RECOV dominates Δμ-TO-HTONE. Therefore, they avoid associating more than one H to a mora. The interaction of these constraints for speakers N and Y is illustrated in the tableau in (20). Here, the demand of Δμ-TO-HTONE is deprioritized, and hence candidate (c) is selected as the winner.[24]

[22] See Kager [1999], McCarthy [2002], and Prince and Smolensky [2004] for more thorough introductory discussions on Optimality Theory. See also Yip [2002] for general, extensive OT analyses on tonal phenomena, including tone-intonation interactions.
[23] L% can be associated with the final MiP, suggesting that NONFINALITY(MiP) only targets H, which is prominent.
[24] One might wonder why, given that the H tone cannot be associated with the accented mora, the H tone docks onto the mora of the case particle. The orientation toward a right MiP edge is presumably due to the well-documented preference for aligning tonal elements (and other linguistic elements) with domain edges [McCarthy and Prince, 1993, and references cited therein].

(20) Speakers Y, N

<table>
<tr><th></th><th>NonFin(MiP)</th><th>*MultiLink/Recov</th><th>Δμ-to-HTone</th></tr>
<tr><td>a. H*L H
| |
$(\Delta\mu\text{-o})_{MiP}\ (\quad)_{MiP}$</td><td>*!</td><td></td><td></td></tr>
<tr><td>b. H*L H
|/
$(\Delta\mu\text{-o})_{MiP}\ (\quad)_{MiP}$</td><td></td><td>*!</td><td></td></tr>
<tr><td>c. ☞ H*L H
| |
$(\Delta\mu\text{-}\quad\text{o})_{MiP}\ (\quad)_{MiP}$</td><td></td><td></td><td>*</td></tr>
</table>

The Optimality Theoretic analysis described above has at least two virtues. First, the analysis models the observed patterns as arising from interactions of several markedness constraints, each of which has independent cross-linguistic motivation. Second, it can model interspeaker variation in terms of minimal constraint reranking between the speakers. Furthermore, within Optimality Theory, we can derive the mixed characteristics of the H tone from constraint interactions, as illustrated in (19) and (20). The H tone behaves like an accentual H when Δμ-to-HTone dominates MultiLink/Recov, but behaves like a boundary H if subjected to the opposite ranking. More study is warranted to investigate whether the distinction between accentual Hs and boundary Hs are absolute and categorical, or whether the properties previously associated with these two kinds of tones are simply tendencies arising from different rankings of violable constraints.

7 Conclusion

7.1 Summary

Although hitherto unmotivated in studies of Japanese intonation, the IP exists within the prosodic structure of Japanese. The IP in Japanese is characterized by an F_0 lowering process, a pause and a creaky vowel in final position, as well as a distinctively large pitch reset and initial rise.

Each clause in multi-clause constructions is parsed into an IP, and the entire sentence is parsed into one level higher than the IP, an Utterance. Four pieces of evidence support the analysis of the Utterance as a grouping of IPs: (i) the amount of clause-initial rise is larger in the first clause than in the second and third clauses; (ii) the domain of general declination is defined over the course of the entire sentences, rather than the individual clauses; (iii) verbal initial rises are smaller in the final clauses than in the first and second clauses, and (iv) the typical sentence-ending H tone appears only at the end of the final clause.

The locus of the IP corresponds to a syntactic clause, as predicted by the universalist view of the prosodic hierarchy advanced in Selkirk [2005]. The universalist theory of prosodic structure informed the current study: it is an expectation of the theory that directed us to look at multi-clause sentences in Japanese. It is our hope that this work demonstrates the fruitfulness of phonetic researches informed by phonological hypotheses.

7.2 Remaining Issues

Before closing, it behooves us to discuss a number of issues for future research. First, our research made exclusive use of lab-read speech in order to control for constituency length and accent patterns. We assumed as a null hypothesis that laboratory speech and spontaneous utterances exhibit very similar if not identical patterns. However, it is important to verify that spontaneous utterances show the properties of the prosodic phrases we have identified.

Second, we have not addressed the relationship between prosodic levels and processing. Prosody affects sentence processing in nontrivial ways [see Beckman, 1996; Warren, 1999, for reviews]. In particular, Schafer [1997] argues that the English MaP is a domain that affects syntactic processing (e.g. resolving ambiguous syntactic attachments), whereas the IP marks semantically and pragmatically relevant information. Hirotani [2005] claims that in Japanese, the MaP is a domain within which a syntactic binding/licensing relation must hold (e.g. licensing of scope-taking items, anaphoric binding). In light of these previous studies, it is an interesting topic for future research to investigate and determine the role of the IP in sentence processing in Japanese.

Third, we have seen that domain-initial H tones appear higher at the edge of higher prosodic boundaries, which we attributed to domain-initial strengthening effects. Previous work has found that domain-initial strengthening in articulation is also greater at the boundary of higher prosodic constituents [Cho and Keating, 2001; Fougeron and Keating, 1997; Hayashi et al., 1999; Hsu and Jun, 1998; Keating, 2003; Keating et al., 2003; Onaka, 2003]. Therefore, our proposal predicts that in multi-clause sentences, articulatory strengthening at the segmental level would be larger at the beginning of initial clauses than at the beginning of noninitial clauses. Our dataset did not control for segmental composition; we leave the testing of this prediction for future research.

Finally, our proposal makes another prediction. It is known that domain-final segments undergo lengthening [Sugahara and Turk, 2006; Wightman et al., 1992], and that the lengthening is larger at higher prosodic edges. We therefore predict that gapping and coordination sentences will show stronger lengthening in final clauses than in non-final ones. Again, since we did not control for segmental content in our experimental design, we leave the fruit of the prediction for future research.

Acknowledgments

This work is in part supported by NSF grant BCS-0004038 to Lisa Selkirk. We thank Kathryn Flack, Mako Hirotani, Lyn Frazier, Sun-Ah Jun, Daniel Mash, Lisa Selkirk, Hubert Truckenbrodt, and two anonymous reviewers of *Phonetica* for their comments. Also thanks to Kanae Amino, Miki Ikoma, Masahiko Komatsu, and Kiyokaze Sugai for their help with the creakiness judgments reported in § 4.2. Any remaining errors are ours. Portions of this work were presented at Japanese/Korean Conference 14 [2004] at Arizona University. An earlier version of this paper appeared as Kawahara and Shinya [2005].

Appendix: List of All Sentences Used in the Experiment

Coordination

SS1

Murasugi-wa	namauni-o	moritsuke,	Munakata-wa	mamemochi-o
Murasugi-TOP	raw sea urchin-ACC	put on dish	Munakata-TOP	bean rice-cake-ACC

moritsuke,	Morimura-wa	aemono-o	moritsuketa.
put on dish	Morimura-TOP	mixed salad-ACC	put on dish

'Murasugi put a raw sea urchin on a dish, Munakata put a bean rice cake on a dish, and Morimura put a mixed salad on a dish.'

SS2

Muramatsu-wa	Aomori-o	aruki,	Iwasaki-wa	Yamanashi-o	aruki,
Muramatsu-TOP	Aomori-ACC	walk	Iwasaki-TOP	Yamanashi-ACC	walk,

Morishita-wa	Miyazaki-o	aruita.
Morishita-TOP	Miyazaki-ACC	walked

'Murasugi walked in Aomori, Iwasaki walked in Yamanashi, and Morishita walked in Miyazaki.'

SL1

Ninomiya-wa	Nirasaki-no	orenji-o	erabi,	Yonekura-wa	Nagasaki-no
Ninomiya-TOP	Nirasaki-GEN	orange-ACC	choose	Yonekura-TOP	Nagasaki-GEN

amaguri-o	erabi,	Imamoto-wa	Onomichi-no	aemono-o	eranda.
chestnut-ACC	choose,	Imamoto-TOP	Onomichi-GEN	mixed salad-ACC	chose

'Ninomiya chose an orange from Nirasaki, Yonekura chose a chestnut from Nagasaki, and Imamoto chose a mixed salad from Onomichi.'

SL2

Yonekura-wa	murasaki-no	aibana-o	nagame,	Ninomiya-wa	yawaraka-na
Yonekura-TOP	purple-GEN	dayflower-ACC	stare at	Ninomiya-TOP	soft-ADJ

oniyuri-o	nagame,	Imamoto-wa	uraraka-na	yamabuki-o	nagameteiru.
lily-ACC	stare at	Imamoto-TOP	bright-ADJ	yellow rose-ACC	is staring at

'Yonekura is staring at a purple dayflower, Ninomiya is staring at a soft lily, and Imamoto is staring at a bright yellow rose.'

LS1

Morioka-no	aniyome-wa	Muramatsu-o	urayami,	Yamagata-no
Morioka-GEN	sister-in-law-TOP	Muramatsu-ACC	envy	Yamagata-GEN

yamabushi-wa	Yamanishi-o	urayami,	Aomori-no	omawari-wa	Nonomura-o
monk-TOP	Yamanishi-ACC	envied	Aomori-GEN	police-TOP	Nonomura-ACC

urayanda.
envied

'The sister in law in Morioka envied Muramatsu, the monk in Yamagata envied Yamanishi, and the police in Aomori envied Nonomura.'

LS2

Yamadera-no	yamamori-wa	imomushi-o	aishi,	Muroran-no
mountain-temple-GEN	guard-TOP	caterpillar-ACC	love	Muroran-GEN

anemuko-wa	minomushi-o	aishi,	Yamazaki-no	nirauri-wa
sister's husband-TOP	bagworm-ACC	love	Yamazki-GEN	leek seller-TOP

aomushi-o	aishiteiru.
greenworm-ACC	loves

'The guard at a mountain temple loves a caterpillar, the sister's brother in Muroran loves a bagworm, and the leek-seller at Yamazaki loves a greenworm.'

DA 1

Morishita-wa	Yamanashi-ni	yamabuki-o	hakobi,	Aomori-ni	oniyuri-o	hakobi,
Morishita-TOP	Yamanashi-DAT	yellow rose-ACC	bring	Aomori-DAT	lily-ACC	bring

Nagasaki-ni	aibana-o	hakonda.
Nagasaki-DAT	dayflower-ACC	brought

'Morishita brought a yellow rose to Yamanashi, brought a lily to Aomori, and brought a dayflower to Nagasaki.'

DA 2

Morimura-wa	Iraiza-ni	maiwashi-o	nagashi,	Amanda-ni	amaguri-o	nagashi,
Morimura-TOP	Eliza-DAT	sardine-ACC	smuggle	Amanda-DAT	chestnut-ACC	smuggle

norimaki-o	Amaria-ni	nagashita.
sushi-ACC	Amelia-DAT	smuggled

'Morimura smuggled a sardine to Eliza, smuggled a chestnut to Amanda, and smuggled a norimaki-sushi to Amelia.'

Gapping

SS1

Murasugi-wa	namauni-o,	Munakata-wa	mamemochi-o,	Morimura-wa
Murasugi-TOP	raw sea urchin-ACC	Munakata-TOP	bean rice cake-ACC	Morimura-TOP

aemono-o	moritsuketa.
mixed salad-ACC	put on dish

'Murasugi put a raw sea urchin on a dish, Munakata a bean rice cake, and Morimura a mixed salad.'

SS2

Muramatsu-wa	Aomori-o,	Iwasaki-wa	Yamanashi-o,	Morishita-wa
Muramatsu-TOP	Aomori-ACC	Iwasaki-TOP	Yamanashi-ACC,	Morishita-TOP

Miyazaki-o	aruita.
Miyazaki-ACC	walked

'Murasugi walked in Aomori, Iwasaki in Yamanashi, and Morishita in Miyazaki.'

SL1

Ninomiya-wa	Nirasaki-no	orenji-o,	Yonekura-wa	Nagasaki-no	amaguri-o,
Ninomiya-TOP	Nirasaki-GEN	orange-ACC	Yonekura-TOP	Nagasaki-GEN	chestnut-ACC

Imamoto-wa	Onomichi-no	aemono-o	eranda.
Imamoto-TOP	Onomichi-GEN	mixed salad-ACC	chose

'Ninomiya chose an orange from Nirasaki, Yonekura a chestnut from Nagasaki, and Imamoto a mixed salad from Onomichi.'

SL2

Yonekura-wa	murasaki-no	aibana-o,	Ninomiya-wa	yawaraka-na	oniyuri-o,
Yonekura-TOP	purple-GEN	dayflower-ACC	Ninomiya-TOP	soft-ADJ	lily-ACC

Imamoto-wa	uraraka-na	yamabuki-o	nagameteiru.
Imamoto-TOP	bright-ADJ	yellow rose-ACC	is staring at

'Yonekura is staring at a purple dayflower, Ninomiya a soft lily, and Imamoto a mild yellow rose.'

LS1
Morioka-no aniyome-wa Muramatsu-o, Yamagata-no yamabushi-wa
Morioka-GEN sister-in-law-TOP Muramatsu-ACC Yamagata-GEN monk-TOP

Yamanishi-o, Aomori-no omawari-wa Nonomura-o urayanda.
Yamanishi-ACC Aomori-GEN police-TOP Nonomura-ACC envied
'The sister in law in Morioka envied Muramatsu, a monk in Yamagata Yamanishi, and a police officer in Aomori Nonomura.'

LS2
Yamadera-no yamamori-wa imomushi-o, Muroran-no
mountain-temple-GEN guard-TOP caterpillar-ACC Muroran-GEN

anemuko-wa minomushi-o, Yamazaki-no nirauri-wa aomushi-o
sister's husband-TOP bagworm-ACC Yamazki-GEN leek seller-TOP greenworm-ACC

aishiteiru.
loves
'A guard at a mountain temple loves a caterpillar, the sister's husband in Muroran a bagworm, and a leek-seller at Yamazaki a greenworm.'

DA 1
Morishita-wa Yamanashi-ni yamabuki-o, Aomori-ni oniyuri-o, Nagasaki-ni
Morishita-TOP Yamanashi-DAT yellow rose-ACC Aomori-DAT lily-ACC Nagasaki-DAT

aibana-o hakonda.
dayflower-ACC brought
'Morishita brought a yellow rose to Yamanashi, a lily to Aomori, and a dayflower to Nagasaki.'

DA 2
Morimura-wa Iraiza-ni maiwashi-o, Amanda-ni amaguri-o, norimaki-o
Morimura-TOP Eliza-DAT sardain-ACC Amanda-DAT chestnut-ACC sushi-ACC

Amaria-ni nagashita.
Amelia-DAT smuggled
'Morimura smuggled a sardine to Eliza, a chestnut to Amanda, and a norimaki-sushi to Amelia.'

Predective
Yonemura-wa omawari-da.
Yonemura-TOP police-copula
'Yonemura is a police officer.'

Yonemura-wa Umemachi-no omawari-da.
Yonemura-TOP Umemachi-GEN police-copula
'Yonemura is a police officer in Umemachi.'

Yamaura-wa awauri-da.
Yamaura-TOP millet seller-copula
'Yonemura is a millet seller

Yamaura-wa yamiichi-no yakuurida.
Yamaura-TOP black market-GEN drug seller
'Yamaura is a drug seller at a black market.'

Intransitive
Arimoto-ga nomiaruita.
Arimoto-NOM drank around
'Arimoto drank around.'

Arimoto-ga	Nominaga-machi-de	nomiaruita.
Arimoto-NOM	Nominaga-machi-LOC	drank around

‘Arimoto drank around in Nominagamachi.’

Mari-ga	nenaoshita.
Mari-NOM	fell back to sleep

‘Mari fell back to sleep.’

Mari-ga	nominoichi-de	nenaoshita.
Mari-NOM	flea market-LOC	fell back to sleep

‘Mari fell back to sleep at a flea-market.’

Long Single-Clause Sentence

Omawari-no	Muramatsu-ga	Urumuchi-no	yamayama-de	naramatsu-ni
police-GEN	Muramatsu-NOM	Ürūmqi-GEN	mountains-LOC	pine trees-DAT

namamizu-o	agemashita.
water-ACC	gave

‘Muramatsu, a police officer, watered a pine tree in the mountains in Ürūmqi.’

Morioka-no	aniyome-ga	maminami-no	yamadera-de
Morioka-GEN	sister-in-law-NOM	south-GEN	mountain-temple-LOC

aemono-ni	oniyuri-o	soemashita.
mixed salad-DAT	lily-ACC	put next to

‘The sister-in-law in Morioka put a lily next to a mixed salad at a southern mountain temple’.

Murasugi-wa	Aomori-no	yamadera-de	niraurini-ni
Murasugi-TOP	Aomori-GEN	mountain-temple-GEN	leek-seller-DAT

yawaraka-na	erimaki-o	urikonda
soft-ADJ	scarf-ACC	sold

‘Murasugi sold a soft scarf to a leek-seller at a mountain temple in Aomori’

Nagahara-wa	Amaura-no	Nogemachi-de	umidori-ni	yawaraka-na
Nagahara-TOP	Amaura-GEN	Nogemachi-GEN	seagull-DAT	soft-ADJ

aomushi-o	agemashita.
caterpillar-ACC	gave

‘Nagahara gave a soft caterpillar to a seagull in Nogemachi in Amaura.’

References

Abe, J.; Hoshi, H.: Gapping and p-stranding. J. East Asian Linguist. *6*: 101–136 (1997).

Aissen, J.: Markedness and subject choice in Optimality Theory. Nat. Lang. Linguist. Th. *17*: 673–711 (1999).

Arvaniti, A.; Baltazani, M: Intonational analysis and prosodic annotation of Greek spoken corpora; in Jun, Prosodic typology: the phonology of intonation and phrasing, pp. 84–177 (Oxford University Press, Oxford 2005).

Baumann, S.; Grice, M.; Benzmüller, R.: GToBI – a phonological system for the transcription of German intonation; in Stanislaw, Demenko, Prosody 2000: speech recognition and synthesis, pp. 21–28 (Adam Mickiewicz University, Faculty of Modern Languages and Literature, Poznan 2001).

Beckman, M.: The parsing of prosody. Lang. cognitive Proc. *11*: 17–67 (1996).

Beckman, M.; Pierrehumbert, J.: Intonational structure in English and Japanese. Phonology *3*: 255–309 (1986).

Bickmore, L.: Branching nodes and prosodic categories; in Inkelas, Zec, The phonology-syntax connection, pp. 1–17 (University of Chicago Press, Chicago 1990).

Bickmore, L.: Tone and stress in Lamba. Phonology *12*: 307–341 (1995).

Bing, J.: Aspects of English prosody; PhD diss. University of Massachusetts, Amherst (1979).

Chitoran, I.; Goldstein, L.; Byrd, D.: Gestural overlap and recoverability: articulatory evidence from Georgian; in Gussenhoven, Warner, Papers in Laboratory Phonology 7, pp. 419–447 (Mouton de Gruyter, Berlin 2002).

Cho, T.: Keating, P.: Articulatory and acoustic studies of domain-initial strengthening in Korean. J Phonet. *29*: 155–190 (2001).
Cooper, W.E.; Sorensen, J.M.: Fundamental frequency in sentence production (Springer, New York 1981).
de Lacy, P.: The interaction of tone and stress in Optimality Theory. Phonology *19*: 1–32 (2002).
Féry, C.; Hartmann, K.: The focus and prosodic structure of German Right Node Raising and Gapping. Linguist. Rev. *22*: 69–116 (2005).
Flemming, E.: Auditory representations in phonology; PhD diss. UCLA (1995).
Fougeron, C.; Keating, P.: Articulatory strengthening at edges of prosodic domain. J. acoust. Soc. Am. *106*: 3728–3740 (1997).
Fujisaki, H.; Hirose, K.: Analysis of voice fundamental frequency contours for declarative sentences of Japanese. J. acoust. Soc. Japan *5*: 233–242 (1984).
Goldsmith, J.: Meeussen's rule; in Aronoff, Oehrle, Language sound structure, pp. 245–259 (MIT Press, Cambridge 1984).
Goldsmith, J.: Tone and accent, and getting the two together; in Aske, Beery, Michaelis, Filip, Papers from BLS 13, pp. 88–104 (Berkeley Linguistic Society, Berkeley 1987).
Gordon, M.; Ladefoged, P.: Phonation types: a cross-linguistic overview. J. Phonet. *29*: 383–406 (2001).
Gouskova, M.: Deriving economy: syncope in Optimality Theory; PhD diss. University of Massachusetts, Amherst (2003).
Grabe, E.L: Pitch accent realisation in English and German. J. Phonet. *26*: 129–144 (1998).
Grice, M.; Ladd, R.; Arvaniti, A.: On the place of phrasal accents in intonational phonology. Phonology *17*: 143–185 (2000).
Hayashi, W.; Hsu, C.-S.; Keating, P.: Domain-initial strengthening in Taiwanese: a follow-up study. UCLA Working Papers Phonet. *97*: 152–156 (1999).
Hayes, B.; Lahiri, A.; Bengali intonational phonology. Nat. Lang. Linguist. Th. *9*: 47–96 (1991).
Hirotani, M.: Prosody and LF interpretation: processing Japanese Wh-questions; PhD diss. University of Massachusetts, Amherst (2005).
Hsu, C.-S.; Jun, S.-A.: Prosodic strengthening in Taiwanese: syntagmatic or paradigmatic? UCLA Working Papers Phonet. *96*: 69–89 (1998).
Hume, E.; Johnson, K.: The role of perceptual phenomena in phonology (Academic Press, San Diego 2001).
Hyde, B.: Nonfinality; Washington University (unpublished, 2003).
Hyman, L.: Boundary tonology and the prosodic hierarchy; in Inkelas, Zec, The phonology-syntax connection, pp. 109–125 (University of Chicago Press, Chicago 1990).
Ishihara, S.-I.: Intonation and interface conditions; PhD diss. MIT (2003).
Jun, S.-A.: The accentual phrase in the Korean prosodic hierarchy. Phonology *15*: 189–226 (1998).
Jun, S.-A.: Prosodic typology: the phonology of intonation and phrasing (Oxford University Press, Oxford 2005).
Kager, R.: Optimality theory (Cambridge University Press, Cambridge 1999).
Kanerva, J.: Focusing on phonological phrases in Chichewa; in Inkelas, Zec, The phonology-syntax connection, pp. 145–162 (University of Chicago Press, Chicago 1990).
Kasai, H., Takahashi, S.: An ATB movement approach to gapping; in Jang, Kim, Proc. the 2000 Int. Workshop on Generative Grammar, pp. 75–89 (Hanshin Publishing, Seoul 2000).
Kawahara, S.: On the proper treatment of non-crisp edges; in Endo, Jun, Sells, Japanese/Korean Linguistics *13* (CSLI, Stanford 2007).
Kawahara, S.; Shinya, T.: The intonation of gapping and coordination in Japanese: evidence for Intonational Phrase; in Kawahara, University of Massachusetts Occasional Papers in Linguistics 30: Papers in Prosody, pp. 91–127 (GLSA, Amherst 2005).
Keating, P.: Phonetic encoding of prosodic structure. Proc. 6th Int. Semin. on Speech Production, 2003, pp. 119–124.
Keating, P.; Cho, T.; Fougeron, C.; Hsu, C.-S.: Domain-initial strengthening in four languages; in Papers in Laboratory Phonology 6: Phonetic interpretation, pp. 145–163 (Cambridge University Press, Cambridge 2003).
Kim, H.; Yoon, T.-J.; Cole, J.; Hasegawa-Johnson, M.: Acoustic differentiation of L- and L-L% in switchboard and radio news speech. Proc. Speech Prosody 2006.
Kim, J.-S.: Syntactic focus movement and ellipsis: a minimalist approach; PhD diss. University of Connecticut (1997).
Kristoffersen, G.: The phonology of Norwegian (Oxford University Press, Oxford 2000).
Kubozono, H.: The organization of Japanese prosody. Stud. Jap. Linguist., vol 2 (Kurosio Publishers, Tokyo 1993).
Kuno, S.: Japanese: a characteristic OV language; in Lehmann, Syntactic typology (University of Texas Press, Austin 1978).
Ladd, D. R.: Intonational phonology (Cambridge University Press, Cambridge 1996).
Ladd, R.: Declination 'reset' and the hierarchical organization of utterances. J. acoust. Soc. Am. *84*: 530–544 (1988).
Ladd, R.: Metrical representation of pitch register; in Beckman, Kingston, Papers in Laboratory Phonology: Between the grammar and physics of speech, pp. 35–57 (Cambridge University Press, Cambridge 1990).
Liberman, M.; Pierrehumbert, J.: Intonational invariance under changes in pitch range and length; in Aronoff, Oehrle, Language sound structure, pp. 157–233 (MIT Press, Cambridge 1984).
Liberman, M.; Prince, A.: On stress and linguistic rhythm. Linguist. Inq. *8*: 249–336 (1977).

Maeda, S.: A characterization of American English intonation; PhD diss. MIT (1976).
Maeda, S.; Venditti, J.: Phonetic investigation of boundary pitch movements in Japanese. Proc. Int. Conf. on Spoken Lang. Processing, 1998, pp. 631–634.
Maeda, S.; Venditti, J.; van Santen, J.P.H.: Modeling Japanese boundary pitch movements for speech synthesis. Third ESCA Workshop on Speech Synthesis, 1998.
Maekawa, K.; Kikuchi, H.; Igarashi, Y.; Venditti, J.: An extended JToBI for spontaneous speech. Proc. 7th Int. Congr. on Spoken Lang. Processing, 2002, pp. 1545–1548.
McCarthy, J.: A thematic guide to Optimality Theory (Cambridge University Press, Cambridge 2002).
McCarthy, J.; Prince, A.: Generalized alignment; in Booij, van Marle, Yearbook of morphology, pp. 79–153 (Kluwer, Dordrecht 1993).
McCawley, J.: The phonological component of a grammar of Japanese (Mouton, The Hague 1968).
Nespor, M.; Vogel, I.: Prosodic phonology (Foris, Dordrecht 1986).
Onaka, A.: Domain-initial strengthening in Japanese: an acoustic and articulatory study. Proc. 15th Int. Congr. Phonet. Sci., 2003, pp. 2091–2094.
Pierrehumbert, J.: The phonetics and phonology of English intonation; PhD diss. MIT (1980).
Pierrehumbert, J.; Beckman, M.: Japanese tone structure. LI Monogr. Series No. 15 (MIT Press, Cambridge 1988).
Poser, W.: The phonetics and phonology of tone and intonation in Japanese; PhD diss. MIT (1984).
Potts, C: The logic of conventional implicatures; PhD diss. University of California at Santa Cruz (2003).
Prieto, P.; D'Imperio, M.; Elordieta, G..; Frota, S.; Vigario, M.: Evidence for soft preplanning in tonal production: initial scaling in Romance. Proc. Speech Prosody, 2006, pp. 803–806.
Prince, A.; Smolensky, P.: Optimality Theory: constraint interaction in generative grammar (Blackwell, Malden 2004).
Rialland, A.: Anticipatory raising in downstep realization: evidence for preplanning in tone production; in Kaji, Proc. Symp. Cross-Linguistic Studies of Tonal Phenomena: Tonogenesis, Accentology, and Other Topics, pp. 301–321 (Tokyo University of Foreign Studies, Tokyo 2001).
Rooth, M.: On the interface principles for intonational focus. Proc. SALT *6*: 202–226 (1996).
Saito, M.: Three notes on syntactic movement in Japanese; in Imai, Saito, Issues in Japanese linguistics (Foris, Dordrecht 1987).
Schafer, A.: Prosodic parsing: the role of prosody in sentence comprehension; PhD diss. University of Massachusetts, Amherst (1997).
Selkirk, E.: The role of prosodic categories in English word stress. Linguist. Inq. *11*: 563–605 (1980).
Selkirk, E.: On derived domains in sentence phonology. Phonol. Yb. *3*: 371–405 (1986).
Selkirk, E.: Sentence prosody: intonation, stress, and phrasing; in Goldsmith, The handbook of phonological theory, pp. 550–569 (Blackwell, Cambridge 1995).
Selkirk, E.: The interaction of constraints on prosodic phrasing; in Horn, Prosody: theory and experiment, pp. 231–271 (Kluwer Academic Publishers, Dordrecht 2000).
Selkirk, E.: Testing the prosodic structure hypothesis. J. Phonet. Soc. Japan *5*: 53–56 (2001).
Selkirk, E.: Contrastive FOCUS vs. presentational focus: prosodic evidence from right node raising in English. Speech Prosody, 2002, pp. 643–646.
Selkirk, E.: Comments on intonational phrasing in English; in Frota, Vigario, Freitas, Prosodies: selected papers from the Phonetics and Phonology in Iberia Conference, 2003. Phonetics and Phonology Series, pp. 11–58 (Mouton de Gruyter, The Hague 2005).
Selkirk, E.; Shinya, T.; Kawahara, S.: Phonological and phonetic effects of minor phrases length on f0 in Japanese. Proc. 2nd Int. Conf. on Speech Prosody, 2004, pp. 183–186.
Selkirk, E.; Shinya, T.; Sugahara, M.: Degree of initial lowering in Japanese as a reflex of prosodic structure organization. Proc. 15th Int. Congr. Phonet. Sci., 2003, pp. 491–494.
Selkirk, E.; Tateishi, K.: Syntax and downstep in Japanese; in Georgopoulos, Ishihara, Interdisciplinary approaches to language: essays in honor of S.-Y. Kuroda, pp. 519–544 (Kluwer, Dordrecht 1991).
Shinya, T.: The intonational asymmetry between argument and adjunct in Japanese; in Kawahara, University of Massachusetts Occasional Papers in Linguistics 30: Papers on Prosody, pp. 185–218 (GLSA, Amherst 2005).
Silverman, D.: Phasing and recoverability; PhD diss. University of California, Los Angeles (1995).
Sohn, K.-W.: On gapping and right node raising; in Kim, Lee, Lee, Yang, Yoon, Explorations in generative grammar: a festschrift for Dong-Whee Yang, pp. 589–611 (Hankuk Publishing, Seoul 1994).
Steriade, D.: Phonetics in phonology: the case of laryngeal neutralization; University of California, Los Angeles (unpublished, 1997).
Sugahara, M.: Downtrends and post-focus intonation in Tokyo Japanese; PhD diss. University of Massachusetts, Amherst (2003).
Sugahara, M.; Turk, A.: Durational correlates of sub-lexical constituent structure in English; Doshisha University and the University of Edinburgh (unpublished, 2006).
'tHart, J.; Cohen, A.: Intonation by rule: a perceptual quest. J. Phonet. *1*: 309–327 (1973).
Thorsen, N.: A study of the perception of sentence intonation: evidence from Danish. J. acoust. Soc. Am. *64*: 1014–1030 (1980).
Truckenbrodt, H.: Phonological phrases: their relation to syntax, focus, and prominence; PhD diss. Massachusetts Institute of Technology (1995).
Truckenbrodt, H.: Upstep and embedded register levels. Phonology *19*: 77–120 (2002).

Truckenbrodt, H.: Final lowering in non-final position. J. Phonet. *32*: 313–348 (2004).
Venditti, J.: Japanese ToBI labeling guidelines; in Ainsworth-Darnell, D'Imperio, Ohio State University Working Papers in Linguistics *50*: 127–162 (1997).
Venditti, J.: The ToBI model of Japanese intonation; in Jun, Prosodic typology: the phonology of intonation and phrasing, pp. 172–200 (Oxford University Press, Oxford 2005).
Vogel, I.; Kenesei, I.: The interface between phonology and other components of grammar: the case of Hungarian. Phonol. Yb. *4*: 243–263 (1987).
Warren, P.: Prosody and language processing; in Garrod, Pickering, Language processing, pp. 155–188 (Psychology Press, Hove 1999).
Wightman, C.; Shattuck-Hufnagel, S.; Ostendorf, M.; Price, P.: Segmental durations in the vicinity of prosodic phrase boundaries. J. acoust. Soc. Am. *91*: 1707–1717 (1992).
Xu, Y.: Production and perception of coarticulated tones. J. acoust. Soc. Am. *95*: 2240–2253 (1994).
Xu, Y.: Contextual tonal variations in Mandarin. J. Phonet. *25*: 62–83 (1997).
Yip, M.: Tone (Cambridge University Press, Cambridge 2002).