2025. Proc Ling Soc Amer 10(1). 5884. https://doi.org/10.3765/plsa.v10i1.5884.

Prosodic boundary processing is language-dependent

Shinobu Mizuguchi & Koichi Tateishi *

Abstract. This paper provides empirical data to support the claim that prosodic boundary processing is language-dependent. Prosody plays an important role in determining the meaning of the utterance. Despite their importance in speech recognition, prosodic boundaries in spontaneous speech are understudied, and the finding that syntactic and prosodic boundaries are not isomorphic complicates automatic speech recognition. This paper considers how prosodic boundaries are perceived in spontaneous speech via perception experiments.

We will consider Japanese first. It is a mora-timed pitch language and typologically different from stress languages like English. Japanese prosody is complex; it is compositionally formed by the lexical pitch accents H*L, phrasal tones, and boundary tones (L%, H%, LH%, HL%). Though Japanese literature has a long history of prosodic studies on the word-level and the phrase-level, prosody above the φ-level is understudied and no standard view for phrasal patterns has ever been established. We conducted perception experiments on spontaneous Japanese via Rapid Prosody Transcription (RPT) and conducted multiple regression analyses between boundary marking and cue candidates. Our findings are that the primary cue for boundary perception in Japanese is post-boundary pause, followed by syntactic higher categories.

On the other hand, prosody of stress languages is well-studied, and the studies claim that the prosodic boundary cues in such languages are either acoustic (e.g. French, etc.) or syntactic (e.g. American English, Estonian, etc.). If boundary cues vary between Japanese and American English, it is expected that Japanese learners of English process prosodic boundaries differently from native American English listeners. We did a comparative study and analyzed how native listeners and learners would perceive the same English natural speech. We reanalyzed the RPT data in Cole et al. 2010 and Mizuguchi et al. 2016 and found that native listeners and learners use different perception cues.

Keywords. prosody; acoustic and syntactic boundary cues; Japanese; English; spontaneous speech; Rapid Prosody Transcription (RPT)

1. Introduction. Prosody is central to language comprehension by helping listeners segment the incoming text. Prosody denotes properties besides word-level accentual features and is characterized by suprasegmental features such as tonal structure, pitch accents, and phonological boundaries. Prosody is considered to be hierarchically organized in prosodic domains of mora, syllable, foot, prosodic phrase and intonation phrase, as in (1).

* We thank our experiment participants and the audience of our presentation at the 99th General Meeting of LSA 2025 in Philadelphia. Our deepest gratitude goes to Jennifer Cole for letting us share her data, and Timothy Mahrt for his help and suggestions in the course of our experiments. This work is partially supported by the collaborative research project fund 2021 of the National Institute of the Japanese Language and Linguistics (NINJAL), given to the first author. All errors are our own. Authors: Shinobu Mizuguchi (mizuguti@kobe-u.ac.jp) & Koichi Tateishi (tateishi@mail.kobe-c.ac.jp).

(1) Prosodic Hierarchy (Féry 2017:36)

υ	utterance	(corresponds roughly to a paragraph or more)
ι-phrase	intonation phrase	(corresponds roughly to a clause)
φ-phrase	prosodic phrase	(corresponds roughly to a syntactic phrase)
ω-word	prosodic word	(corresponds roughly to a grammatical word)
F	Foot	(metrical unit)
σ	syllable	(strings of segments)
μ	Mora	(unit of syllable weight)

Studies of the prosody above the φ-phrase level are in progress and most of them are on stress-languages like English and French. There is, however, no consensus on methods of segmentation into prosodic categories. As for American English, the literature on boundary perception of natural spontaneous speech in stress-languages find that syntactic categories are the primary boundary cues and the acoustic category of vowel duration is the secondary predictor of prosodic boundaries (Cole et al. 2010), while Smith 2009 suggests that pauses favor the perception of a boundary in English. In French, pauses are the strongest prosodic cues, followed by higher syntactic categories (cf. Simon & Christodoulides 2016). In German, boundary cues are acoustic cues of the pre-boundary and phrase-final lengthening (cf. Petrone et al. 2017). In spontaneous Estonian, syntactic cues are the primary and pause is the secondary boundary predictor (Ots & Taremaa 2023). These studies on stress-languages show that (i) acoustic cues and/or syntactic cues predict boundaries in stress-languages, (ii) acoustic cues vary among languages, and (iii) the choice of the primary boundary cue depends on a language.

This paper considers Japanese. Japanese is a mora-timed pitch language and typologically different from well-studied stress-languages. We would like to study how Japanese boundaries are perceived and find the primary boundary cue. Japanese literature boasts prosodic studies on the word level, but prosodies above the φ-level are understudied. Even the standard view for phrasal patterns above the word level has not been established (cf. Pierrehumbert & Beckman (P&B) 1988). What is worse, no empirical perception data on spontaneous Japanese is available, and the *Corpus of Spontaneous Japanese* marks Intonation Phrase in the framework of ToBI mechanically when boundary pause is more than 0.2 second only by stipulation (cf. Maekawa 2011, among others). We will conduct perception experiments on spontaneous Japanese in this paper to find boundary cues in Japanese. We will further compare Japanese and American English and see whether there is a cross-linguistic difference between the two languages in boundary processing.

The structure of this paper is as follows: Section 2 presents the brief introduction of Japanese prosody and prosodic structure, followed by perception experiments on Japanese spontaneous speech. Section 3 will compare how native listeners of American English and Japanese learners of English (JEFL) perceive the same excerpts of English, and show that boundary processing is under the first-language interference. Section 4 concludes the paper and suggests theoretical implications.

2. Japanese Prosody and Boundary Perception Experiment. In this section, we will briefly review Japanese prosody and previous theoretical analyses first, and then introduces boundary perception experiments we conducted on spontaneous Japanese.

2.1. JAPANESE PROSODY.Japanese is a pitch language and its prosody is complex; it is compositionally formed by the lexical pitch accents H*+L (e.g. H*+L on *Na'oya* and *oyo'ida* in Figure

1), phrasal tones (e.g. L%, H%, HL% in Figure 1), and boundary tones (L%, H%, LH%, HL%). The selection and alignment of phrasal and boundary tones depend on a context.

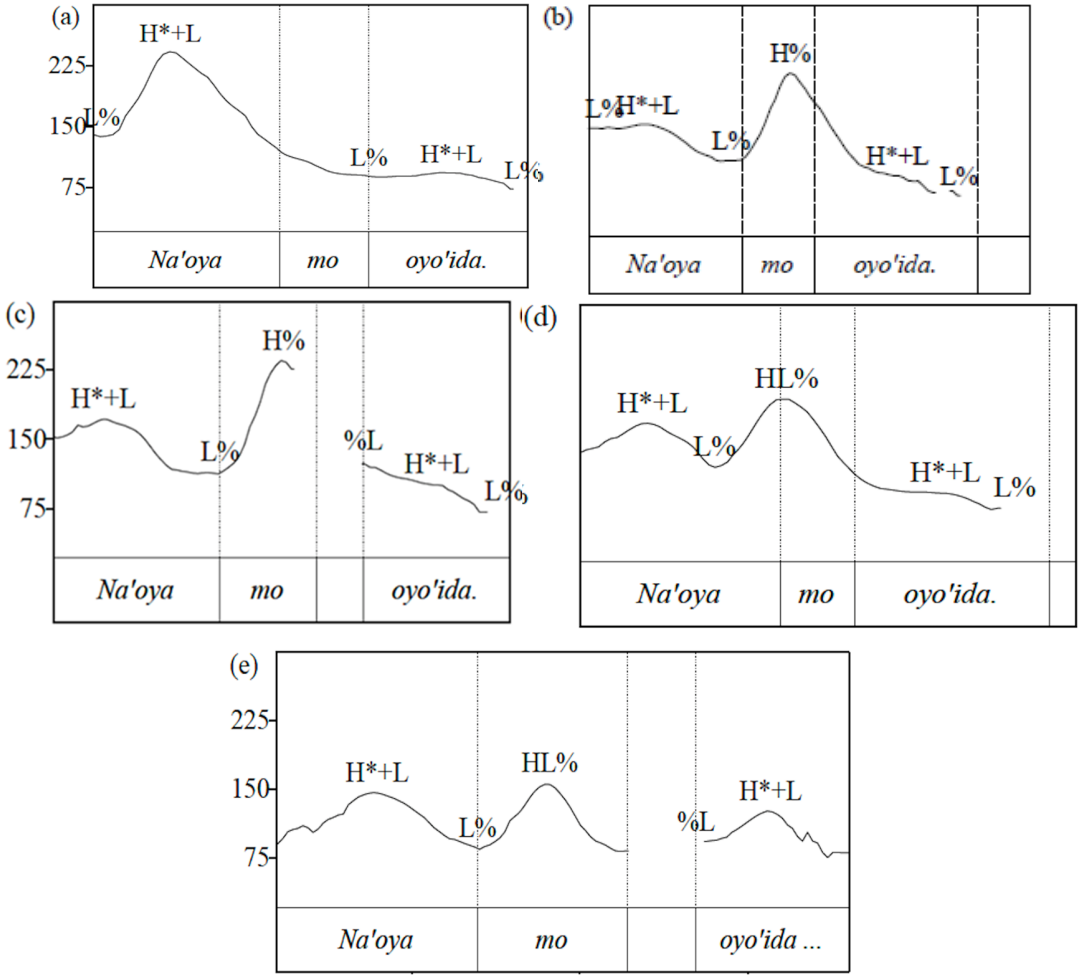

Figure 1. Pitch movements of *Naoya-mo oyoida* 'Naoya also swam.'
(Venditti et al. 2008: 488)

Japanese literature has a long history of prosodic studies on the word-level and the phrase-level (cf. McCawley 1968, Poser 1984, Pierrhumbert & Beckman (P&B) 1988, Kubozono 1993, Venditti et al. 2008, Ito & Mester 2012, among many others). Studies above the φ-level are, however, understudied and no standard view for phrasal patterns above the word level has ever been established, as shown in Table 1.

McCawley 1968 Poser 1984 Kubozono 1993	Pierrehumbert & Beckman 1988	(X)J-ToBI	Ito & Mester 2012 Ishihara 2022	Cross-linguistic phrasing
	ι phrase		ι phrase	utterance υ ι phrase
Major Phrase	Intermediate phrase	I(ntonation) P(hrase)	φ	φ phrase
Minor Phrase	A(ccentual) P(hrase)		φ	φ phrase

Table 1: Prosodic phrasings proposed in the Japanese literature

As for phrasal tones, Venditti et al. 2008 claim that L%, H%, LH%, and LHL% are four major phrase-final tones, and they are aligned at the AP in their framework, although Koiso et al. 2020 consider that they are aligned at the AP and Intonation Phrase (on their own definition, i.e. cross-linguistic φ-phrase). (X)J-ToBI labels an IP (i.e. cross-linguistic φ-phrase) only by stipulation when a boundary pause is more than 0.2 sec. (cf. Maekawa 2011, among others).

Interface studies, though not many, have been proposed in the Japanese literature; P&B 1988 proposed the Reset Theory and Selkirk & Tateishi (S&T) 1988 proposed the Left Edge Theory.

(2) a. The Reset Theory (P&B 1988): Focus appears at the leftmost position of the phrase (their Intermediate Phrase) and hence it resets a prosodic phrase.
b. The Left Edge Theory (S&T 1988): Japanese is a left-branching language and prominence is aligned leftmost within a φ-phrase.

These classical theories defined a prosodic phrase via prominence. They were, however, empirically refuted; Shinya 1999 and Kubozono 2007 report that prominence is observed not only at the phrase-initial position but also at the phrase-mid position. Mizuguchi & Tateishi (M&T) 2023 conducted perception experiments on spontaneous Japanese and show that (2) is not empirically supported (cf. Table 2 below).

Selkirk 2009 proposes 'The Match Theory' to map syntactic categories N, V, and A to ω, XP to φ, and CP to *ι*, with OT-theoretic constraints such as *MATCH(clause, ι)*. Ishihara 2022 proposes each speech act is mapped to ι : *MATCH(SA, ι)*. These theories are not empirically proven yet.

The brief review above on the prosody studies in the Japanese literature shows that empirical data, especially perceptual data, needs to be improved. We will conduct perception experiments on spontaneous Japanese below. Our research questions (RQ) are the following:

RQ1: What is the primary cue in boundary perception of spontaneous Japanese?

RQ2: Is boundary processing dependent on language, i.e., is it different between Japanese and other languages?

2.2. BOUNDARY PERCEPTION EXPERIMENTS. We will conduct Rapid Prosody Transcription (RPT) perception experiments on Japanese spontaneous speech. RPT was developed by Cole and her colleagues (cf. Cole et al. 2010) as a tool for prosody research to see how acoustic, phonological, syntactic, semantic, and pragmatic properties determine a listener's perception. In RPT, untrained transcribers mark boundaries and prominences on unpunctuated texts while listening to

spontaneous speech, based on an auditory impression, with minimal instructions and without examples of transcriptions or feedback. The p(rominence)-score and b(oundary)-score are calculated; they indicate the proportion of transcribers who underscore the respective word, and higher values indicate strong perceptual salience of the prosodic element.

2.2.1. METHODOLOGY. Our method is RPT, and our materials are 13 excerpts of *Corpus of Spontaneous Japanese* (CSJ) released by the National Institute for Japanese Language and Linguistics (NINJAL). CSJ is a corpus of monologues and dialogues by more than 1,400 Tokyo Japanese native speakers, with the total recorded time being about 660 hours. We used 6 lecture-type monologues and 7 pseudo-lecture-type monologues. Our excerpts are 16 to 41 seconds long. Since Japanese is an agglutinative language, we segmented our data set on the morpheme level (cf. (3)); our materials contain 490 content words and 490 function morphemes. For the syntactic analysis of boundary-marking, syntactic categories of S, S-bar, Conjunction, XP, Particle, and non-syntactic categories of Disfluency and Discourse Marker (DM) are assigned at the left edge ('[X') and the right edge (']Y') of each morpheme, following Cole et al. 2010, as in (3).

(3) [S [Adv mazu]Adv [NP [N watashi] -[Part no]]NP [NP [V sukina [N mono]]]NP [[N inu] -
first I -GEN like thing dog-
[V desu]]VP]]S
COPULA

'First, what I like is a dog.'

(N.B. In the experiment, the text is written in Japanese *Hiragana,* i.e. Japanese cursive syllabary character, without punctuation.)

We recruited three groups of transcribers without hearing difficulties: 35 Tokyo Japanese (TJ) listeners (mean age 24.8, SD=0.5), 27 Osaka Japanese (OJ) listeners (mean age 25.3, SD=3), and 11 Northern Kanto (NK) listeners (mean age 23.2, SD=4)[1]). The prosody varies among dialects in Japanese; on the word (ω) level, TJ and OJ have accented (H*+L) and unaccented (LH) words, while NK has only unaccented words. On the phrase (φ) level, TJ and NK allow dephrasing, i.e., the process of deleting a φ-phrase when syntactic words form a prosodic phrase (e.g. (4) below), while OJ does not (cf. Igarashi 2014). On the intonation phrase (ι) level, TJ is downward (L%), NK is upward (H%), and OJ is both downward and upward. We recruited these three groups of participants to examine whether listeners with different dialectal backgrounds process boundaries differently or not. We predicted that the dialectal differences would affect boundary perception and OJ listeners would mark more boundaries than TJ and NK listeners.

We conducted online perception experiments via Yahoo! Crowd-Sourcing Service[2]. After the exercise session, our participants listened to each material twice via PC, while marking boundaries and prominences on the text by clicking a mouse. Their responses were saved on the

[1] We conducted our experiments during the pandemic of Covid 19 and recruited 50 participants for each group online. Due to the mismatch between a self-claimed native dialect and the target dialect, we regret that we had to dismiss the mismatched data.

[2] Our experiments were approved by the Ethical Committee of NINJAL.

computer via LMEDS, an experimental platform developed by Tim Mahrt (Mahrt 2016). It took about 30 minutes for our participants to complete the task, and they were paid in Yahoo! points.

2.2.2. RESULTS. Our major findings are the following. First, the overall inter-listener agreements were κ0.638 on the boundary (b-) score and κ0.359 on the prominence (p-)-score on Fleiss' Kappa, which are above the chance level and show that our data are reliable.

Second, the correlation between the p-score and b-score is weak (r=0.12 on Pearson's Correlation). Table 2 shows the numbers of prominence and boundary markings, and we see that they do not match.

	Prominence Marking (p>0.2)				Boundary Marking (b>0.2)
	Phrase-initial	Phrase-mid	Phrase-final	Total	Total
TJ	74	145	53	272	167
OJ	63	91	26	180	181
NK	52	54	22	128	163

Table 2: Numbers of prominence and boundary markings

OJ is significantly different from TJ and NK in boundary marking on One-way ANOVA (p=0.004, p=0.017, respectively), while TJ and NK are not.

Third, the b-scores per syntactic category are given in Figure 2.

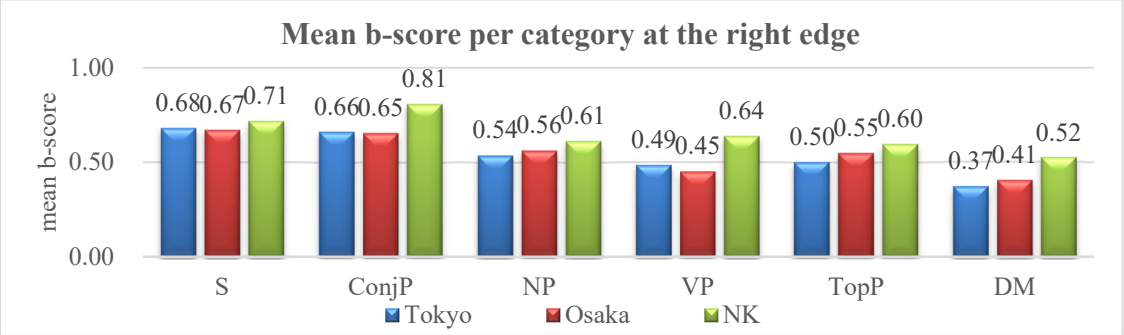

Figure 2: Mean b-scores per category at the right edge
(N.B. S=Sentence, CONJP=Conjunction Phrase, NP=Noun Phrase, VP=Verb Phrase, TOPP=Topic Phrase, DM=Discourse Marker)

The b-scores at the edge of higher syntactic categories (i.e., S, CONJP) are higher than those at the edge of lower categories (i.e., XP), and the difference is significant ($F(1, 1936)=3.864$, $p<0.001$).

Fourth, among the four major boundary tones in Venditti et al. 2008, L% and HL% cover nearly 90% of the boundary tones aligned, but H% and LH% are relatively few in use. Observe Table 3.

	TJ	OJ	NK
L%	61	67	57
HL%	62	65	59
H%	9	9	9
LH%	10	12	10
total	142	153	135

Table 3: The number of four phrasal tones assigned at the right boundary edge (b>0.2)

Fifth, Table 4 shows the results of regression analyses between b-score and acoustic cues (z-standardized). Among the acoustic cues (MaxF0, range F0, mora-duration, intensity, pre-boundary pause, and post-boundary pause), post-boundary pauses are the primary acoustic boundary cue in spontaneous Japanese.

	TJ		OJ		NK	
	t	Adj. R^2	t	Adj. R^2	t	Adj. R^2
b × MaxF0	-0.973	< 0.001	-0.111	-0.001	-0.534	-0.0007
b × range F0	3.138**	0.009	3.96***	0.015	3.492***	0.011
b × duration	6.117***	0.036	7.724***	0.056	10.321***	0.097
b × intensity	7.313***	0.05	-0.091	-0.001	6.491***	-0.0007
b × post-boundary pause	41.09***	0.635	41.078***	0.635	40.744***	0.631
b × pre-boundary pause	3.03**	0.008	4.87***	0.023	3.1**	0.009

Table 4: Results of regression analyses between b-score and acoustic cue (z-standardized) (where ***: p<0.001, **: p<0.01)

2.3. DISCUSSION. P&M 1988 claim that 'focus' appears at the leftmost position of a φ-phrase (i.e., the Intermediate Phrase in their term) and resets a prosodic phrase. In their framework, a boundary is inserted before a focus. Their theory predicts that every φ-phrase has a focus at its leftmost position. However, Shinya 1999 and Kubozono 2007 empirically refuted P&B's reset theory on the sentence level. Our RPT experiments found that the correlation between boundary and prominence perception is very weak (r=0.12 on Pearson's Correlation) and also that focal prominence appears in the phrase-initial, phrase-mid, and phrase-final positions in spontaneous speech (cf. Table 2). We found that some phrases do not mark focal prominence at all (cf. Figure 3 Left) and that a single phrase contains more than one prominence (cf. Figure 3 (Right)).

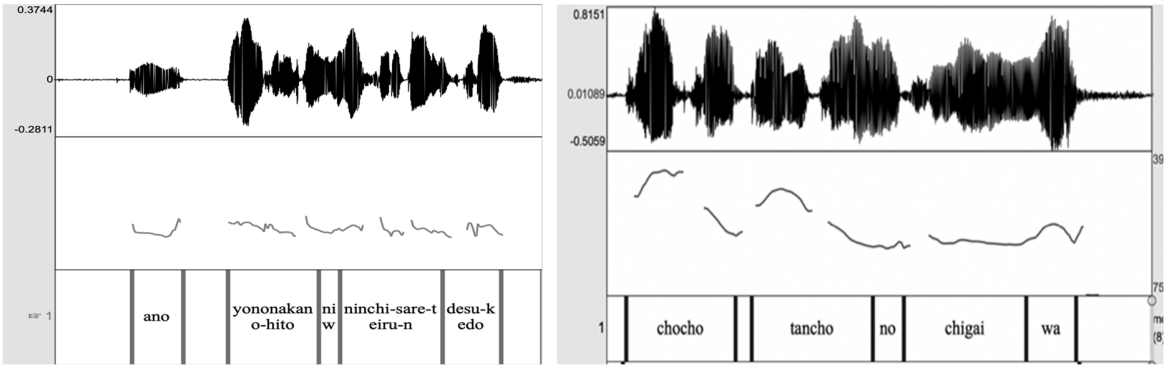

Figure 3: Pitch movement in a prosodic boundary (Left: without a prominence marking (extracted from CSJ file #A01f0055), Right: with three prominence markings (extracted from CSJ file #S00f0082))

We consider these as serious problems to the Reset theory and the Left-edge theory (cf. (2))[3]. In other words, prosodic boundaries are not aligned based on prominence in Japanese.

What are boundary cues in spontaneous Japanese, then? Ots & Taremaa 2023 claim that there are 'bottom-up' processing and 'top-down' processing of boundaries in natural speech; the former employs acoustic cues like vowel lengthening, duration, pause, intensity, and F0, and the

[3] We will not go into the details of prominence perception here. For further details, see Mizuguchi & Tateishi 2023.

latter is based on syntactic cues. Our results above show that Japanese uses post-boundary pause, phrasal tones of L% and HL%, and syntactic categories for boundary-processing. Kawahara 2012 conducted a production experiment on parentheticals in Japanese and claims that there is a substantial pause before, but not after, the parenthetical clause. Our perception data, however, gives a different picture: Table 4 shows that post-boundary pause, but not pre-boundary pause, is a strong boundary predictor. We focus on the post-boundary pause below and conduct regression analyses between b-scores and cue candidates to see whether Japanese is of bottom-up or top-down type in the sense of Ots & Taremaa. Table 5 shows the results.

	TJ		OJ		NK	
	t	Adj.R^2	t	Adj.R^2	t	Adj.R^2
higher categories	24.74***	0.386	23.44***	0.361	25.22***	0.396
lower categories	20.35***	0.299	22.8***	0.348	18.92***	0.27
L%	8.154***	0.063	8.402***	0.067	8.698***	0.07
post-boundary pause	41.08***	0.635	41.08***	0.635	40.74***	0.631

Table 5: Results of regression analyses between b-scores and cue candidates (where ***: $p<0.001$)

Table 5 shows that post-boundary pause is the primary boundary cue in all three dialects, followed by higher categories (i.e., S and ConjP) and lower categories (XP). Our answer to RQ1, 'What is the primary cue in boundary perception of spontaneous Japanese?' is that post-boundary pause is the primary boundary cue, followed by higher and lower syntactic categories. We can claim that Japanese is 'bottom-up' type in boundary processing.

We need to refer to the IP markings in CSJ. They stipulated to mark an IP (i.e., cross-linguistic phrase) when a boundary pause is more than 0.2 sec. (cf. Maekawa 2011, among others), but their markings do not match our findings; the mean averages of post-boundary pauses are 0.129 sec. in lecture-type speeches and 0.237 sec. in pseudo-lecture-type speeches. The IP labels in CSJ need to be reconsidered.

Before closing this section, let us take a quick look at our finding that there is a cross-dialectal difference in the perception of boundaries. Let us recall that OJ is significantly different in the number of boundary markings from TJ and NK Japanese, while the difference between TJ and NK is not (cf. Table 2). Japanese literature has found that prosody varies among dialects in Japanese. Igarashi 2014 shows that TJ and NK allow 'dephrasing', i.e., the process of deleting a φ-phrase (Intermediate Phrase in Igarashi's term) when conjoining syntactic words to form a prosodic phrase, while OJ does not. Before the RPT experiments, we never expected that native Japanese perceive the same Tokyo Japanese differently depending on the dialect they speak, but Table 2 shows that OJ listeners mark boundaries more than TJ and NK listeners. Figure 4 is an example where only OJ listeners mark a boundary (b-score 0.296) between object *takuan-o* (pickled radish-ACC) and subject *haha-ga* (mother-SUBJ) in (4a).

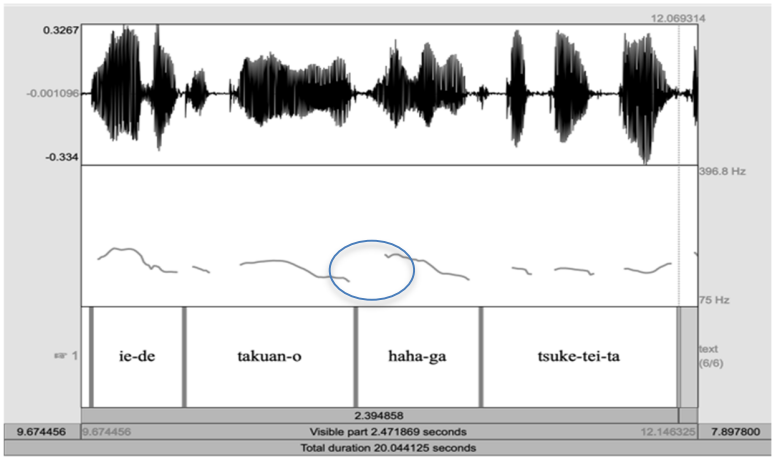

Figure 4: Pitch movement of (4a) (extracted from CSJ file #S00f0095)

(4) a. [S[AdvP ie -de] [NP takuan -o] [NP haha -ga] [VP tsuke
House -at picked radish-ACC mother -NOM pickle
-tei -ta]]
-PROG -PAST
'My mother pickled radish at home.'

b. (ι (φ ie-de) (φ takuan-o))ι (ι (φ haha-ga) (φ tsuke-tei-ta))ι (OJ)

c. (ι (φ ie-de) (φ takuan-o haha-ga) (φ tsuke-tei-ta))ι (others)

Dephrasing is a process of deleting a φ boundary when we spell out the phonological output from a morphosyntactic input. Krazter & Selkirk 2020 propose a constraint *DephraseGiven* in English, and recent studies on tones in Lekeitio Basque (cf. Elordieta & Selkirk 2022) and Xitsonnga (cf. Lee & Selkirk 2022) show that dephrasing applies in these tone languages and affects prosody. (4a) is a morphosyntactic input, and (4b) and (4c) are phonological outputs without and with dephasing, respectively.

We know that (4b) and (4c) are production models, and we still do not know if production models affect perception. But Table 5 shows that lower syntactic phrases of XP are more effective for boundary marking in OJ than TJ and NK, and dephrasing accounts for this fact. If we were on the right track to assume some dialect in Japanese allows dephrasing and production strategy induces perception bias, we can account for why OJ differs from TJ and NK in perceiving boundaries, since OJ does not dephrase prosodic boundaries both in production and perception.

3. Cross-linguistic difference in boundary perception. In Introduction, we have addressed that languages vary in boundary processing between 'top-down' and 'bottom-up' tactics in the sense of Ots & Taremaa 2023; the former is syntax-oriented, and the latter is acoustic-oriented. In the previous section, we have considered how Japanese native listeners process prosodic boundaries in spontaneous Japanese and found that post-boundary pause is the primary and higher syntactic categories are the secondary boundary cues. We claim that Japanese belongs to the 'bottom-up' group, like French (cf. Simon & Christodoulides 2016). On the other hand, American-English (cf. Cole et al. 2010) and Estonian (cf. Ots & Taremaa 2023) employ the 'top-down' processing tactics. When we compare the b-scores per syntactic category of native American English listeners and those of native Japanese listeners, we see a big difference. Observe Figure 5.

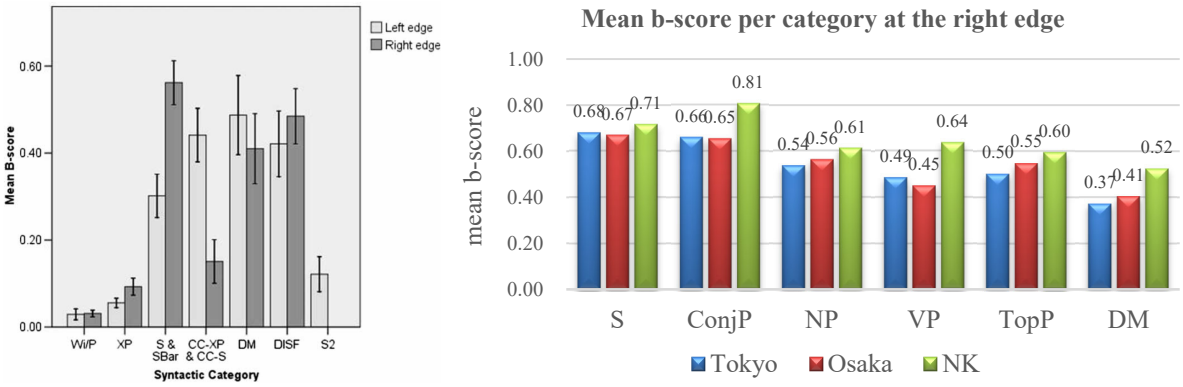

Figure 5: Mean b-score per category (Left: American English (Cole et al. 2010: 1161), Right: Japanese (replication of Figure 2))

We suspect this difference comes from boundary processing tactics: syntax-oriented English versus acoustic-oriented Japanese. Our RQ2 was, 'Is boundary processing dependent on a language, i.e., is it different between Japanese with an acoustic boundary cue and American English with a syntactic boundary cue?' To answer this RQ, we will compare how native listeners and learners process boundaries in spontaneous American English. Generally, Japanese learners of English as a foreign language (JEFL) have much difficulty in processing English prosody. We predict JEFLs, especially those whose proficiency of English is not high, do not use syntactic boundary cues when they perceive English. We will compare the results of RPT experiments on spontaneous speech in Cole et al. 2010 and Mizuguchi et al. 2016 below and see whether our prediction is borne out.

3.1. METHODOLOGY. Cole et al. 2010 conducted RPT experiments on 72 excerpts (11-22 seconds long) of Buckeye Corpus of American English by the total of 97 annotators. Mizuguchi et al. 2016 replicated Cole et al.'s RPT experiments using a portion of the same materials[4]; the materials were 11 excerpts of Buckeye Corpus provided by Prof. Cole. They were 10 to 24 seconds long, as in (5).

(5) i really don't know i think in today's world what they call the nineties that uh it's like everything is changed like when i grew up ...

We recruited 108 Japanese students who were intermediate learners of English (Int, TOEFL PBT mean=493.7) and 15 advanced learners (Adv, TOEFL PBT mean=595). The participants listened to the excerpts twice through a room speaker and marked boundaries and prominences on the transcriptions without punctuation and capitalization with a pencil. It took them about 30 minutes to complete the task, including the exercise session. They were given a part of a course credit for the task. Our experiments were approved by the Ethical Committee of Kobe University.

3.2. RESULTS. We compared the results of RPT experiments on the same materials of 11 excerpts of spontaneous American English. The number of transcribers were 16 native American listeners (L1), 108 Intermediate-level JEFLs (Int), and 15 Advanced-level JEFLs (Adv)[5].

[4] We thank Prof. Jennifer Cole for providing the experiment materials and sharing her data with us.

[5] The numbers of transcribers vary among the groups. Based on an analysis of inter-annotator agreement proposed by Roy et al. 2017, we consider RPT annotations from a group of 13 annotators to be reliable in the sense that they are reproducible with an expected difference of less than 5% of the estimated s.d. of the true population. Our thanks go to Tim Mahrt, who did the calculation for us.

We will list three important results by comparing RPT experiments by native listeners and JEFLs. First, the inter-speaker agreements were 0.63 for L1, 0.521 for Adv, and 0.458 for Int on Fleiss' Kappa. All were above the chance level, and we consider that our data are reliable.

Second, the difference in b-scores is significant between native listeners and JEFLs ($F(1, 1286)=3.85$, $p<0.001$ on One-way ANOVA. Figure 6 shows the mean b-scores per categories at the left edge, where a prosodic phrase starts, and at the right edge, where a prosodic boundary ends. At the right edge, syntactic categories S and SBar function as the primary boundary cues for L1 and Adv, but not for Int; the difference in b-score of S and SBar between L1 and Int is significant on One-way ANOVA ($F(1,22)=4.664$, $p=0.042$), but not between L1 and Adv ($F(1,22)=0.049$, $p=0.44$).

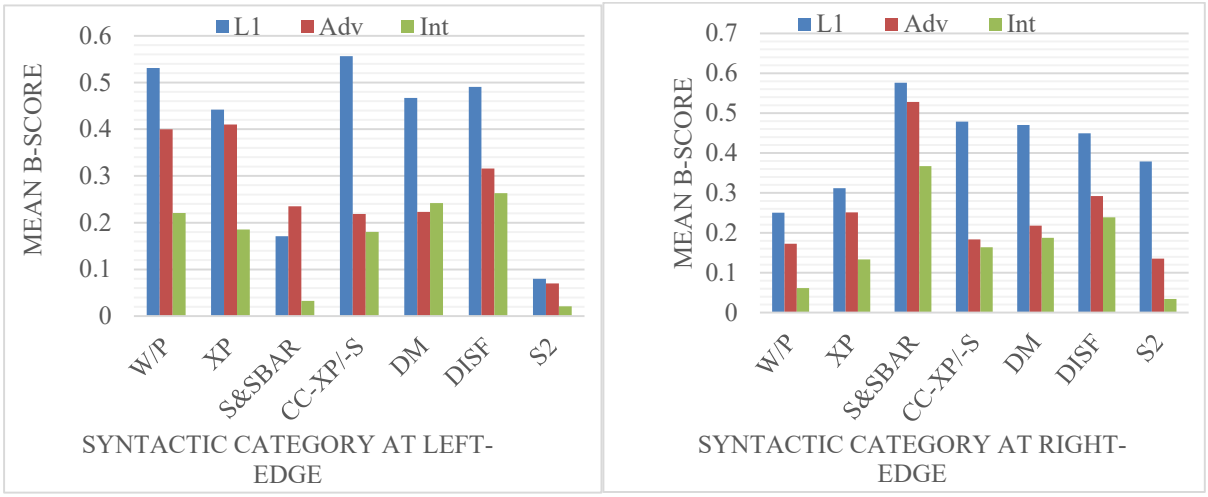

Figure 6: Mean b-score at the left and the right edge per category (N.B. W/P=phrase-medial, XP =NP, VP, AdjP, CC-XP/S=coordinate conjunction, DM=discourse marker, DISF=disfluency, S2=subordinate clause)

At the left edge, Coordinate Conjunction (CC-XP/-S) function as the primary boundary perception cue for L1. JEFLs, on the other hand, do not use CC-XP/-S as a boundary cue, and differences in b-score of CC-XP/-S between L1 and JEFLS are significant ($F(1,22)=4.806$, $p=0.038$ for Int, $F(1,22)=4.419$, $p=0.047$ for Adv).

Third, Cole et al. 2010 did not consider post-boundary pause and claimed that syntactic cues were the primary and vowel duration was the secondary boundary cue. We recalculated their data of the 11 excerpts of their RPT experiment. Table 6 shows that post-boundary pause is a stronger boundary cue than vowel duration for L1 English listeners as well as for JEFLs.

	Kendall's tau		
	L1	JEFL:Adv	JEFL:Int
b-score X vowel duration	0.365	0.291	0.295
b-score X post-boundary pause	0.615	0.550	0.522

Table 6: Correlations between b-scores and acoustic cues

3.3.DISCUSSION. Figure 6 shows that higher syntactic categories of S and SBar, and Coordinate Conjunction function as stronger boundary cues than lower categories of XP for L1 English listeners at the right and the left edge, respectively. Table 6 shows that post-boundary pause is a

strong acoustic boundary cue in American English. We conducted a multiple regression analysis between b-score and cue candidates to see which cue is stronger. Table 7 shows the result.

	t-value		
variables	L1	JEFL: Adv	JEFL: Int
post-boundary pause	17.61	13.38	3.71
higher category at the right edge	27.18	16.35	4.42
lower categories at the right edge	26.74	5.90	4.75
Adv.R^2	0.859	0.657	0.165

Table 7: Results of regression analysis between b-scores and cue candidates

We can see that syntactic higher and lower categories are stronger cues than the acoustic cues of post-boundary pause for L1 listeners. For Adv, higher categories are a bit stronger than post-boundary pause, but for Int, all the cues are weak, and their predictability of boundaries is very low.

By the reanalysis of our RPT experiments, we come to the conclusion that American English employs syntactic categories as the primary cue for boundary processing. Japanese, on the other hand, the acoustic cue of post-boundary pause is the primary boundary perception cue (cf. Table 5). Since American English and Japanese are different in boundary cues, it is a natural consequence that JEFLs, especially those with low English proficiency, have difficulties in boundary perception in English. Our prediction is borne out, and our answer for RQ #2, 'Is boundary processing different between Japanese, a pitch language, and American English, a stressed language?' is 'yes.'

Before closing this section, we would like to compare the correlation between variable candidates of English and Japanese. Observe Table 8.

	post-boundary pause		minor syntactic category	
	English	Japanese	English	Japanese
post-boundary pause	1	1		
major syntactic category	0.48	0.47	-0.11	-0.1
minor syntactic category	0.35	0.46	1	1

Table 8: Correlations between explanatory variables in Pearson's r

We see that post-boundary pause and syntactic categories are moderately correlated in both languages. What is interesting is that major syntactic category and minor syntactic category are correlated with post-boundary pause more in Japanese than in American English. This is probably because Japanese allows scrambling and minor categories of XP often form an independent prosodic phrase. Recall (4) above. Japanese basic structure is SOV, but the object is scrambled before S and aligns an ι-phrase in (4b). Scrambling allows word orders of OSV and OVS in Japanese, and often case markers such as *o* 'Accusative' and *ga* 'Subjective' are missing in spontaneous Japanese, forcing another cue to mark phrases. Also, Japanese has the topic markers *wa* and *mo*, and when a pause follows after the topic marker, as in Figure 1, they form a prosodic phrase. We speculate that scrambling and topic markers align independent prosodic phrases and make minor syntactic categories easier to mark prosodic boundaries in Japanese than in English.

These may make Japanese use acoustic boundary cues more than English. Further studies are of course in demand.

4. Conclusion. Ots & Taremaa 2023 argue that languages vary in boundary processing between 'top-down' and 'bottom-up'; the former is syntax-oriented, and the latter is acoustic-oriented. We conducted boundary perception experiments and found that post-boundary pause is the primary boundary cue, followed by higher syntactic categories in Japanese. In Ots & Taremaa's framework, Japanese is grouped as a bottom-up type in boundary processing. If we are correct to assume that languages vary in boundary processing, learners of some languages will naturally have difficulties in mastering the target language where the boundary processing type of their mother tongue is different from that of their target language. As a first step to proving our claim, we compared boundary processing between Japanese and American English, which is grouped as 'top-down'. We predicted that Japanese learners would have difficulty processing boundaries in American English due to their mother tongue interference. Our perception experiments proved our prediction. Table 7, in fact, shows that JEFLs are improving their boundary perception ability as their proficiency in English goes up. We would suggest language learners to learn processing tactics when their mother tongue uses different processing tactics from the target language.

Prosodic boundaries in spontaneous speech are important in speech recognition, and our study shows that prosodic boundary processing is language dependent. The literature in this field is still limited, and cross-linguistic analyses are still missing. Future research in this field is much needed.

References

Borowsky, Toni, Shigeto Kawahara, Takahiro Shinya & Mariko Sugahara (eds.). 2012. *Prosody Matters*. London: Equinox.

Cole, Jennifer , Yoonsook Mo & Soondo Baek. 2010. The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech, *Language and Cognitive Processes*, 25: 7, 1141‑1177. http://dx.doi.org/10.1080/01690960903525507.

Elordieta, Gorka & Elizabeth Selkirk. 2022. Unaccentedness and the formation of prosodic structure in Lekeitio Basque. In Haruo Kubozono et al. (eds.), *Prosody and Prosodic Interfaces*, 374‑419. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9380198869740.003.0013.

Féry, Caroline. 2017. *Intonation and Prosodic Structure*. Cambridge: Cambridge University Press.

Igarashi, Yosuke. 2014. Typology of intonational phrasing in Japanese dialects. In Jun, S-A. (ed.). *Prosodic Typology*, II, 464‑491. Oxford: Oxford University Press

Ishihara, Shinichiro. 2022. Match Theory: An overview. *Language and Linguistics Compass* 16. https://doi.org/10.1111/lnc3.12446.

Ito, Junko & Armin Mester. 2012. Recursive prosodic phrasing in Japanese. In Toni Borowsky, et al. (eds.), *Prosody Matters*, 280‑303. London: Equinox.

Kawahara, Shigeto. 2012. The intonation of nominal parentheticals in Japanese. In Toni Borowsky, et al. (eds.), *Prosody Matters*, 304‑340. London: Equinox

Koiso, Hanae, Hideaki Kikuchi & Taka'aki Yamada. 2020. Intonation labeling for the corpus of everyday Japanese conversation: The labeling scheme and prosodic features of everyday conversation [in Japanese]. *JSAI Technical Report*, SIG-SLUD-B903, pp. 34‑39.

Kubozono, Haruo. 1993. *The Organization of Japanese Prosody*. Tokyo: Kurosio Syuppan.

Kubozono, Haruo. 2007. Focus and intonation in Japanese: does focus trigger pitch reset? *Workshop on Prosody Syntax, and Information Structure* (WPSI) 2. 1‑27.
Kubozono, Haruo, Junko Ito, & Armin Mester. (eds.). 2022. *Prosody and Prosodic Interfaces*. Oxford: Oxford University Press.
Lee, Seunghum J. & Elizabeth Selkirk. 2022. Xitsonga tone: The syntax-phonology interface. In Haruo Kubozono et al. (eds.), *Prosody and Prosodic Interfaces*. 337‑73. Oxford: Oxford University Press.
Maekawa, Kikuo. 2011. *Kopasu-o riyo-shita shizen-onsei-no kenky*u [Study of natural language speech based on corpus]. Tokyo: Tokyo Institute of Technology dissertation.
Mahrt, T. 2016. *LMEDS: Language makeup and experimental design software.* https://githum.com/timmahrt/LMEDS.
McCawley, James. 1968. T*he Phonological Component of a Grammar of Japanese*. The Hague: Mouton.
Mizuguchi, Shinobu, Gábor Pintér, & Koichi Tateishi. 2016. Natural speech perception cues by Japanese learners of English. *Proceedings of Pacific Second Language Research Forum (PacSLRF) 2016*. 151‑156. Tokyo: Association for Pacific Second Language Research Forum.
Mizuguchi, Shinobu & Koichi Tateishi. 2023. *Prominence in a Pitch Language: The Production and Perception of Japanese.* Lanham, MD: Lexington Books.
National Institute for Japanese Language and Linguistics, Communication Research Laboratory & Tokyo Institute of Technology (eds.). 2001. *Corpus of Spontaneous Japanese*. Tokyo: The National Institute for Japanese Language.
Ots, Nele & Piia Taremaa. 2023. Chunking an unfamiliar language: Results from a perception study of German listeners. In Fabian Schubö, Sabine Zerbian, Sandra Hanne & Osabe;; Wartenbirger (eds.). *Prosodic Boundary Phenomena*. Language Science Press, 87‑117. Berlin: Language Science Press. https://doi.org/10.5281/zenodo.777753.
Petrone, Caterina, Hubert Truckenbrodt, Carolinne Wellmann, Julia Holzgrefe-Lang, Isabell Wartenburger & Barbara Höhle. 2017. Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists. *Journal of Phonetics* 61, 71‑92. https//doi.org/10.1016/j.wocn.2017.01.002.
Pierrehumbert, Janet & Mary Beckman. 1988. *Japanese Tone Structure*. Cambridge, MA: MIT Press
Poser, Bill. 1984. *The phonetics and phonology of tone and intonation in Japanese*. Cambridge, MA: MIT dissertation.
Roi, Joseph, Jennifer Cole, & Tim Mahrt. 2017. Individual differences and patterns of convergence in prosody perception. *Laboratory Phonology* 8(1):22. https//doi.org/ 10.5334/labphon.108.
Selkirk, Elizabeth & Koichi Tateishi. 1988. Constraints on minor phrase formation in Japanese. *Proceedings of the 24th Annual Meeting of the Chicago Linguistics Society*. 316‑336.
Selkirk, Elizabeth. 2009. On clause and intonational phrase in Japanese: The syntactic grounding of prosodic constituent structure. *Gengo Kenkyu* 136. 35‑73.
Shinya, Takahiro. 1999. Eigo to Nihongo ni okeru fokasu ni yoru daunsuteppu no soshi to cho'on-undo no chogo [The blocking of downstep by focus and articulatory overlap in English and Japanese]. *Proceedings of the 13th Annual Meeting of the Sophia University Linguistic Society* 14. 35‑51.

Simon, Anne C. and George Christodoulides. 2016. Perception of prosodic boundaries by naïve listeners in French. *Speech Prosody 2016*, 1158–1162. https://doi.org/10.21437/SpeechProsody.2016-238.
Smith, Caroline. 2009. Naïve listeners' perception of French prosody compared to the predictions of theoretical models. *Actes d'IDP 09*. 335‑349.
Venditti, Jennifer, Kikuo Maekawa & Mary Beckman. 2008. Prominence marking in the Japanese intonation system. In Miyagawa, Shigeru & Mamoru Saito (eds.). *The Oxford Handbook of Japanese Linguistics*. 456‑512. Oxford: Oxford University Press.