BEYOND PITCH: EXPLORING DURATION, INTENSITY, AND SILENCE IN JAPANESE FOCUS MARKING

Jun Nagao
Gifu Shotoku Gakuen University, Japan
junnagao@gifu.shotoku.ac.jp
https://orcid.org/0000-0001-7618-0640

Marta Ortega-Llebaria
University of Pittsburgh, United States
mao61@pitt.edu
https://orcid.org/0000-0003-1637-3541

Abstract

Building on Jun's prosodic typology (2005, 2014), this study examines how head and edge languages differ in their strategies for expressing focus. While head languages like English rely on pitch expansion in focus and post-focal compression, edge languages like Japanese are thought to emphasize boundary cues. Inspired by the framework proposed by Mizuguchi and Tateishi (2023), we hypothesized that in educated standard Japanese, focus marking extends beyond pitch modulation to include the boundary cues such as duration and silence insertions. To test this, we analyzed recordings of native speakers of educated standard Japanese producing noun phrases under broad and narrow focus conditions, examining duration, F0 maxima, intensity, and silence insertion at both the word and morpheme levels. The results demonstrate that while F0 maxima cue focus, duration and silence insertion play dominant roles in marking focus and focus position, especially at the word level. Intensity, in contrast, primarily cues accent, though it also signals focus position at the morpheme level in contexts with unaccented initial words, a particularly challenging environment for focus marking. These findings reveal a hybrid prosodic system in educated standard Japanese, where temporal and boundary cues dominate but pitch modulation remains an auxiliary tool. This system reflects the typological distinctions between head and edge languages, with educated standard Japanese relying on a flexible combination of global and local cues to signal focus. Future research should investigate how these cues are processed in perception and whether similar strategies are employed in other syntactic structures, offering broader insights into the prosodic systems of edge languages.

Keywords: Narrow Focus vs. Broad Focus; Japanese; Pitch; Duration; Silence

Més enllà del to: Exploració de la durada, de la intensitat i del silenci en les marques del focus en japonès

Resum: Partint de la tipologia prosòdica de Jun (2005, 2014), aquest estudi analitza com les llengües *head* (cap) i *edge* (marge) difereixen en les seves estratègies per expressar el focus. Mentre que les llengües *head,* com l'anglès, es basen en l'expansió del to (F0) en el focus i la compressió post-focal, les llengües *edge,* com el japonès, es caracteritzen per l'èmfasi en els senyals de límit. Inspirats en el marc proposat per Mizuguchi i Tateishi (2023), vam hipotetitzar que en el japonès estàndard, la marca de focus va més enllà del to, incloent-hi senyals com la durada i les insercions de silencis. Per comprovar-ho, vam analitzar gravacions de parlants nadius de japonès estàndard produint frases nominals sota condicions de focus ampli i estret, de les quals es va analitzar la durada, els màxims de F0, la intensitat i les insercions de silencis, tant a nivell de paraula com de morfema. Els resultats mostren que mentre que els màxims de F0 indiquen el focus, la durada i les insercions de silenci juguen un paper dominant en marcar el focus i la seva posició, especialment, a nivell de paraula. La intensitat, en canvi, se centra principalment en l'accent, tot i que també assenyala la posició del focus a nivell de morfema en contextos amb paraules inicials sense accent. Aquestes troballes revelen un sistema prosòdic híbrid en el japonès estàndard, en el qual els senyals temporals i de límit dominen, mentre la modulació del to continua sent una eina auxiliar. Aquest sistema reflecteix les distincions tipològiques entre les llengües *head* i *edge* amb el japonès estàndard, fent ús d'una combinació flexible de senyals globals i locals per indicar el focus. Futures investigacions haurien d'examinar com aquests senyals són processats en la percepció i si s'utilitzen estratègies similars en altres estructures sintàctiques, oferint així una visió més àmplia dels sistemes prosòdics en llengües *edge*.

Paraules clau: Focus estret vs. focus ampli; japonès; to; durada; silenci.

Más allá del tono: Exploración de la duración, de la intensidad y del silencio en el marcado del foco en japonés

Resumen: Basándonos en la tipología prosódica de Jun (2005, 2014), este estudio examina cómo las lenguas *head* (cabeza) y *edge* (borde) difieren en sus estrategias para expresar el foco. Mientras que las lenguas *head,* como el inglés, dependen de la expansión tonal (F0) en el foco y la compresión post-focal, las lenguas *edge,* como el japonés, enfatizan las señales relacionadas con los límites. Inspirados en el marco propuesto por Mizuguchi y Tateishi (2023), planteamos la hipótesis de que, en el japonés estándar, la marca del foco va más allá de la modulación tonal, incluyendo señales como la duración y las inserciones de silencio. Para comprobarlo, analizamos grabaciones de hablantes nativos de japonés estándar produciendo frases nominales bajo condiciones de foco amplio y estrecho, examinando la duración, los máximos de F0, la intensidad y las inserciones de silencio, tanto a nivel de palabra como de morfema. Los resultados demuestran que, aunque los máximos de F0 marcan el foco, la duración y las inserciones de silencio desempeñan un papel dominante en el marcado del foco y su posición, especialmente a nivel de palabra. La intensidad, por el contrario, se centra principalmente en el acento, aunque también señala la posición del foco a nivel de morfema en palabras iniciales sin acento. Estos hallazgos revelan un sistema prosódico híbrido en el japonés estándar, donde las señales temporales y de límite dominan, y la modulación tonal sigue siendo una herramienta auxiliar. Este sistema refleja las distinciones tipológicas entre lenguas *head* y *edge* con el japonés estándar, utilizando una combinación flexible de señales globales y locales para marcar el foco. Investigaciones futuras deberían explorar cómo estas señales son procesadas en la percepción y si estrategias similares son empleadas en otras estructuras sintácticas, aportando así una comprensión más amplia de los sistemas prosódicos en las lenguas *edge*.

Palabras clave: Foco estrecho vs. foco amplio; japonés; tono; duración; silencio

1. Introduction

In the prosodic typology proposed by Jun (2024, p. 535), languages are categorized as either head or edge languages, based on how prosodic prominence is realized. English, as a prototypical head language, relies on pitch expansion and compression at prosodic heads, such as syllables bearing lexical stress (e.g., *Ma-* in *Mary*), to signal distinctions like broad versus narrow focus (see Figure 1a and 1b). In contrast, Japanese, classified as an edge language, marks prominence through boundary-related cues within Accentual Phrases (APs, see Figure 3 for examples illustrating APs), such as boundary tones (Venditti, 2005, pp. 181-184, 186-188) and dephrasing (Igarashi, 2014, p. 476). However, the role of additional edge-related prosodic features, beyond these traditional cues, warrants further exploration.

The seminal work by Mizuguchi and Tateishi (2023) underscores this distinction by demonstrating that pitch modulation strategies, effective in head languages, fail to convey focus distinctions in educated standard Japanese, particularly when unaccented words are in focus (see Chapters 3 and 4). This highlights the inadequacy of applying head-language prosodic frameworks to edge languages like Japanese, emphasizing the need to examine alternative focus-marking strategies.

Our study investigates whether prosodic cues such as duration and silence—beyond pitch modulation, boundary tones, and dephrasing—play a role in marking focus in Japanese. Given the limitations of pitch-based approaches in edge languages, we aim to explore how these lesser-studied cues function in production data, paving the way for future perceptual studies.

The remainder of this introduction is structured as follows: Section 1.1 discusses the intonational strategies for expressing focus in English, a head language. Section 1.2 reviews the focus-marking strategies in Japanese, emphasizing the limitations of pitch-based cues. Finally, Section 1.3 presents our hypothesis and research questions.

1.1. Broad and Narrow Sentence Focus in English

Broad and narrow focus are universal linguistic categories, but languages vary in how they express these distinctions. Some rely on syntactic reordering (e.g., Hungarian; É. Kiss, 1998), while others use morphological markers (e.g., Akan; Bodomo & Marfo, 1996). In many languages, prosody plays a central role in signaling focus. According to Jun's prosodic typology framework (2005, 2014), head languages, like English, mark prosodic prominence at the heads of prosodic units, whereas edge languages, like Japanese, emphasize prominence at unit boundaries (e.g., Beckman & Pierrehumbert, 1986; Beckman & Venditti, 2011; Beckman, 2012). This typological distinction shapes how focus is conveyed: head languages rely on pitch modulation and stress expansion at prominent syllables, while edge languages use boundary cues such as phrase-final lengthening or boundary tones.

In head languages, sentence intonation is closely tied to lexical stress, with prosodic prominence realized at the stressed syllables of words (Beckman & Pierrehumbert, 1986; Beckman, 2012; Ortega-Llebaria 2006; Ortega-Llebaria & Prieto, 2011). Each word contains a syllable bearing primary stress, such as *Ma-* in *Mary*, *co-* in *coming*, and *mo-* in *tomorrow*. This stress is phonetically marked by increased duration and intensity, serving as the landing site for pitch accents that shape sentence intonation. Broad focus, which conveys general or new information, distributes prominence evenly across the sentence. As shown in Figure 1a, broad focus intonation involves progressively lower pitch accents on stressed syllables, with no single syllable standing out significantly. For example, in the response "Mary's coming tomorrow" to *What's happening?*, each stressed syllable (*Ma-*, *co-*, *mo-*) contributes equally to the overall intonation.

In contrast, as shown in Figure 1b, narrow focus highlights a specific word, emphasizing its importance or contrasting it with other elements. For instance, in response to *Who's coming tomorrow? Peter?*, narrow focus on *Mary* amplifies the stressed syllable *Ma-*, making it the most salient part of the utterance. This is achieved by increasing the pitch, intensity, and duration of the focused word's stressed syllable, while compressing the prosodic cues of post-focal elements. Intonation and stress work together, with the stressed syllable serving as the primary site for these enhancements. Figure 1b illustrates this: under narrow focus, *Ma-* in *Mary* becomes significantly more prominent, while *coming tomorrow* shows reduced pitch and duration.
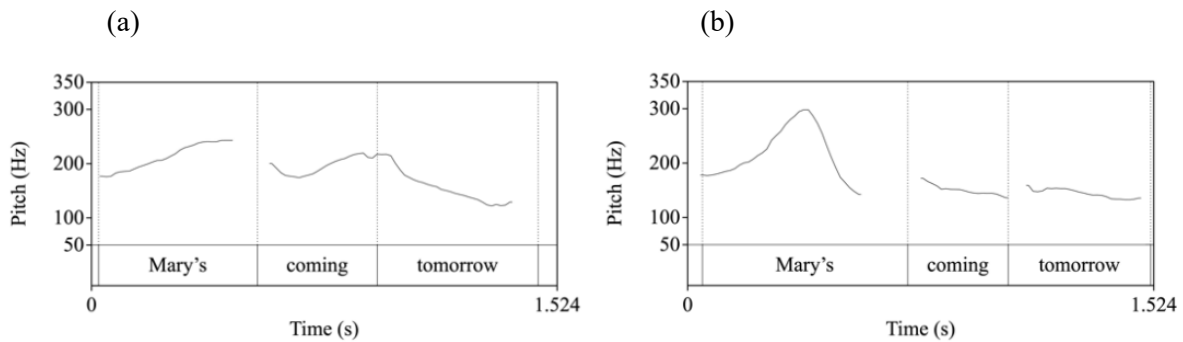
(a)   (b)

Figure 1. Broad Focus (left) and Narrow Focus (right) in English.

1.2. Broad and narrow focus in educated standard Japanese

Japanese is classified as an edge language in Jun's prosodic typology, meaning that its prosodic structure relies heavily on edge marking through pitch reset and boundary tones (Jun, 2005; 2014). Within this framework, dialectal variation introduces additional complexity. Accentless Japanese dialects align solely with the edge typology, while accented dialects, like Tokyo Japanese, exhibit characteristics of both edge and head languages (Venditti, 2005; Igarashi, 2014, 2015). This dual typology is grounded in the presence of pitch accents.

Pitch accents in Japanese, particularly in Tokyo Japanese, differ fundamentally from lexical stress in languages like English. They are cued exclusively by pitch, with a fixed High-Low (HL) tonal pattern, and are not accompanied by changes in duration (Beckman & Pierrehumbert, 1986; Venditti et al., 2008). Furthermore, not all words in Japanese carry a pitch accent, resulting in a mix of accented and unaccented words, which can form minimal pairs (e.g., *hashi* 'chopsticks' [accented] vs. *hashi* 'edge' [unaccented] in Figure 2) (Beckman & Pierrehumbert, 1986; Igarashi, 2015). This contrasts sharply with English, where lexical stress is phonetically complex, involving multiple cues like pitch, duration, and intensity (e.g., Fry, 1955; Lieberman, 1960; Ladd, 2008). English stress patterns are more dynamic, adapting to sentence-level intonation, and every content word contains one syllable with primary stress (e.g., Beckman, 1986; Fry, 1955).
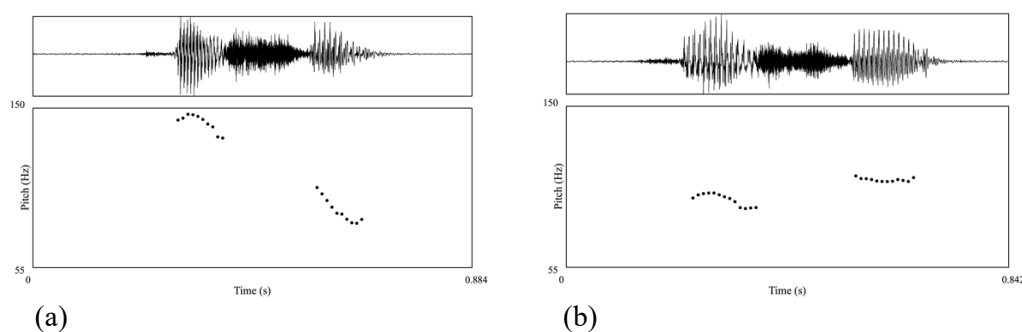

(a)   (b)

Figure 2. hashi "chopsticks" (accented) hashi "edge" (unaccented) produced in isolation.

These typological distinctions between Japanese pitch accent and English lexical stress significantly influence the prosodic strategies available for marking focus. As explained in 1.1., focus in English is typically expressed through pitch range expansion and post-focus compression (e.g., Cooper et al., 1985; Xu & Xu, 2005). Mizuguchi and Tateishi (2023) demonstrate that this mechanism although possible, as illustrated in Figure 3, is less effective in Japanese, particularly when the focused word lacks a pitch accent. In such cases, pitch range expansion is limited, and the post-focus compression strategy of head languages is ineffective. However, accented words in Japanese can use pitch expansion to mark focus, albeit with certain constraints (Mizuguchi & Tateishi, 2023).
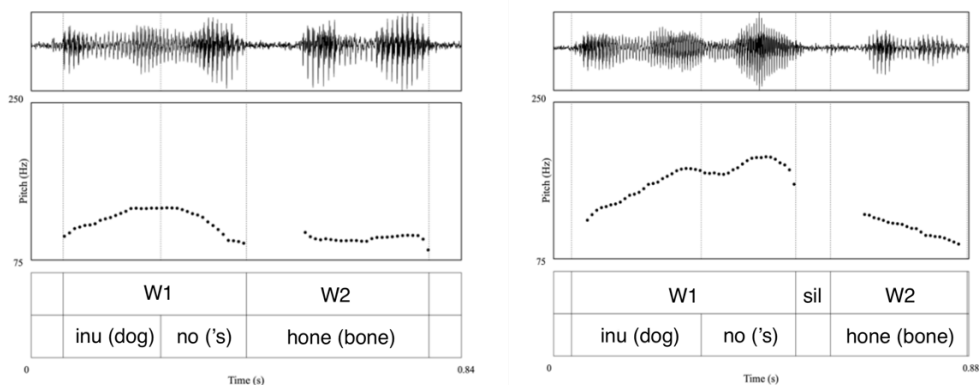
Figure 3. Japanese "inuno hone" in Broad Focus (left) and Narrow Focus on "inuno"(right).

For broad focus, the entire phrase maintains a default pitch contour within its Accentual Phrases (AP) structure, with no specific pitch prominence. These two APs tend to be perceived as a single intonation unit or intonational phrase. When expressing narrow focus on "*inuno*", either the root "*inu*" or the genitive "*no*" expands the AP pitch range.

Instead, Japanese relies on edge-oriented prosodic cues, such as pitch resets at the boundaries of Accentual Phrases (APs) and Intonation Phrases (IPs) (Pierrehumbert & Beckman, 1988; Venditti, 2005). When narrow focus is applied, the focused AP exhibits an expanded pitch range, and its right edge is reinforced through additional prosodic features, including prominence tone boundaries, boundary pauses, and sometimes lengthening of the AP (e.g., Seo et al. 2019; Maeda & Venditti, 1998). These edge strategies are efficient for marking focus, even in phrases with unaccented words.

Mizuguchi and Tateishi's findings (2023) highlight that focusing primarily on pitch range changes without considering edge-related cues like pauses and duration might lead to an incomplete understanding of focus marking in Japanese. This study addresses this gap by examining how these edge cues, alongside pitch, contribute to the prosodic expression of focus, emphasizing their role in both accented and unaccented contexts.

1.3. Goal of the Current Study and Research Questions

Building on the insights of Mizuguchi and Tateishi (2023), this study aims to expand our understanding of focus marking in educated standard Japanese by examining how prosodic cues beyond pitch contribute to the distinction between broad and narrow focus. While pitch resets and range expansions at AP boundaries are well-documented as primary mechanisms for signaling focus (e.g., Venditti et al., 1998, 2008), the potential roles of duration, intensity, and silence remain underexplored. These additional cues may play a critical role in refining focus marking, particularly in contexts involving unaccented words, where pitch cues alone may be insufficient. Our goal is to examine these cues on Genitive Phrases such as *umá no hizume* ("horse's hoof")

and *tora no kawa* ("tiger's skin") produced in broad and narrow focus contexts. This approach allows us to analyze how prosodic cues operate within and across APs in both accented and unaccented conditions.

We hypothesize that:

1. Duration and intensity will interact with pitch to enhance the marking of narrow focus, with higher ratios observed in focused APs (e.g., *umá no* in *umá no hizume*) and morphemes of the in-focus word (e.g. *umá* and *no*) compared to those in broad focus contexts.
2. Silence, particularly in the form of boundary pauses between APs, will occur more frequently in narrow focus contexts, providing an additional prosodic marker.
3. The interplay of these cues will yield distinct prosodic patterns, influenced by the presence or absence of pitch accent in the focused words.

The following research questions will guide our investigation:

1. Do duration ratios, pitch ratios, and intensity ratios distinguish broad focus intonation from narrow focus intonation when:
   a. Ratios are measured at the AP level (e.g., between *umá no* and *hizume* in the sentence *umá no hizume* ("horse's hoof")?
   b. Ratios are measured at the morpheme level (e.g. between *umá* and *no*)?
2. Are silences between APs inserted more frequently in narrow focus than in broad focus?
3. How do the cues relevant to Questions 1 and 2 interact in the expression of focus?

By addressing these questions, this study aims to provide a comprehensive understanding of how Japanese leverages a combination of prosodic strategies to express focus, offering new insights into the interaction of pitch, duration, intensity, and silence in both accented and unaccented contexts.

2. Methodology

2.1. Participants

Six native speakers of Japanese (three females, three males; average age 41) participated in this study. Although all participants are fluent in standard Japanese, they come from various regions of Japan, including, Miyagi (Participant 1), Hyogo (Participant 2), Nagano (Participant 3), Tochigi (Participant 4), Saitama (Participant 5), and Tokyo (Participant 6). Given that some participants are from regions outside of Tokyo, there may be slight dialectal variations that influence their prosodic features. However, the formal register was used in the collection of the data to ensure uniformity in language use. In this context, all participants spoke in a standard Japanese style, minimizing any influence of dialectal variation. Additionally, each participant exhibited expected accentuation patterns (distinctions between accented and unaccented words and consistent accent locations within words), allowing us to consider them as speakers of a standard or near-standard Japanese variety.

2.2. Materials

Participants were provided with a list of 48 two-word noun phrases, such as *umáno hizume* (horse's hoof) and *hanáno kubiwa* (flower's necklace), to be spoken with both broad and narrow focus intonation. In narrow focus, emphasis was placed either on the first or second word. All sentences were balanced for accent combinations, including:

- Accented-unaccented words (e.g., *tsunó no kubiwa*),
- Unaccented-accented words (e.g., *tora no kawá*),
- Accented-accented words (e.g., *umá no mimí*),
- Unaccented-unaccented words (e.g., *same no kubiwa*).

Since the 48 sentences were produced with both intonations (broad and narrow focus) by six participants, we collected a total of 576 noun phrases as we can see in Table 1.

Accent Pattern	Position of Narrow Focus	Number of sentences	Examples
AU	Word 1 (the first noun + *-no*)	6	tsunóno kubiwa
UA		6	torano kawá
AA		6	umáno mimí
UU		6	sameno kubiwa
aU	Word 2 (the second noun)	6	umáno hizume
uA		6	ushino tsunó
aA		6	ahíno honé
uU		6	ushino kazari
Total Number of Sentences		48	

Table 1. Set of 48 items. (A=Accented word, U= unaccented word).

2.3. Procedure

Participants first read the list of 48 two-word noun phrases aloud to capture broad focus. The experimenter, who conducted the interview and presented the sentences, is a native Japanese speaker from Sendai with fluency in standard Japanese. The experimenter's proficiency in standard Japanese ensures consistency in pronunciation and intonation throughout the procedure. To capture narrow focus, the participants corrected information provided by the experimenter. For instance, the experimenter would say, "*umáno hizume desu ka?*" (Is it a horse's hoof?), while pointing to *ushíno hizume* (a cow's hoof). The participants would then clarify, "*USHI no hizume desu*" (It's a COW's hoof), emphasizing "cow" instead of "horse". Recordings were conducted to ensure clear capture of all acoustic nuances.

The speech data were analyzed to extract measurements of pitch, duration, and intensity. Specifically, the F0 maxima, duration, and intensity were measured for:

- The root (e.g., *umá* in *umáno hizume*),
- The genitive particle (e.g., *no* in *umáno hizume*),
- Word 1 (e.g., *umáno* in *umáno hizume*),
- Word 2 (e.g., *hizume* in *umáno hizume*).

For each measurement, namely, F0 maxima, duration, and intensity, ratios were computed between the genitive particle *no* and the root, and between Word 1 and Word 2. Additionally, the presence of silences between words as detected by the WEBMAUS forced aligner was recorded.

2.4. Statistics

The first research question, namely whether duration ratios, pitch ratios, and intensity ratios distinguish broad focus intonation from narrow focus intonation, was examined using linear mixed models computed with R and the lme4 package (Bates et al., 2015). For question 1a, the dependent variables were the duration ratios, F0 maxima ratios, and intensity ratios between APs

(e.g., *umáno* and *hizume* in *umá no hizume*). For question 1b, the same ratios were measured between morphemes (e.g., the root and genitive *no* in *umá no*).

For each model, the fixed factors were Focus, Accent, and their interaction. Focus was coded as an orthogonal contrast to account for the differences between broad and narrow focus as well as narrow focus on Word 1 versus narrow focus on Word 2. The fixed factor Accent was sum coded. The random factors were Participant and Sentence. Due to crossed random effects, relevant random slopes were included in the models. Effect sizes were calculated for each statistically significant factor. When relevant, post-hoc comparisons were performed with Welchs t-test and Bonferroni corrections to adjust for multiple comparisons.

Examples of the models include:

- Question 1a: Duration ratio between APs ~ 1 + Focus * Accent + (1+Accent+Focus|Participant) + (1|Sentence)

- Question 1b: Duration ratio between morphemes ~ 1 + Focus * Accent + (1+Accent+Focus|Participant) + (1|Sentence)

For question 2, namely, are silences between APs inserted more often in narrow focus than in broad focus, mean differences in the number of inserted silences between broad focus and narrow focus were compared using a t-test.

For the third research question, which examines how relevant cues and silence insertion interact in the expression of focus, logit mixed-effect models with categorical dependent variables were used. Dependent variables compared broad vs. narrow focus and narrow focus on Word 1 vs. narrow focus on Word 2. Fixed factors included variables with significant effects in question 1 and large effect sizes. Random effects included Participant and Sentence. Examples of models include:

- Broad vs Narrow Focus ~ 1 + duration ratio between APs + F0 ratio between APs + duration ratio between morphemes + Silence + (1|Participant) + (1|Sentence)

- Narrow Focus on Word 1 vs Narrow Focus on Word 2 ~ 1 + duration ratio between APs + F0 ratio between APs + duration ratio between morphemes + Silence + (1|Participant) + (1|Sentence)

3. Results

3.1. Duration, F0, intensity and the insertion of silences

Table 2 summarizes the results of the mixed models addressing research question 1a, which investigates how duration, F0 maxima, and intensity ratios between Word 1 and Word 2 cue focus. The analysis compares Broad Focus to Narrow Focus, and within Narrow Focus, it distinguishes between focus on Word 1 and Word 2. By coding Focus orthogonally, we isolate these comparisons.

The results reveal that duration and F0 models explain more variation in focus marking (29% and 31%, respectively) compared to intensity models (23%). Importantly, Focus consistently shows significant main effects in both duration and F0 models, indicating these cues reliably distinguish broad focus from narrow focus and identify the position of narrow focus.

In contrast, intensity primarily cues Accent, as evidenced by a significant main effect. On average, accented words exhibit higher intensity values than unaccented words. However, the interaction between Narrow Focus position and Accent suggests that the role of intensity in marking accent is modulated by focus position, as shown by significant interaction effects in the models.

| Fixed Factors	Dependent Variables: Ratios Between APs		
	Duration	F0	Intensity
NF position	**	**	**
BF vs NF	***	*	
Accent			***
NF position * Accent	.		*
BF vs NF * Accent			
Model's size effect	29%	31%	23%

Table 2. Summary of linear mixed models examining the effects of focus and accent on the duration, F0, and intensity ratios between APs. All two-word phrases contain two APs. P–values are indicated as follows: '***' ≤ *0.0001*, '**' ≤ *0.001*, '*' ≤ 0.01, '.' ≤ 0.05.

The main patterns from Table 2 are illustrated in Figures 4a and 4b, which provide detailed insights into the significant effects of duration and F0 maxima in distinguishing focus types.

Figure 4a highlights how duration ratios differ between focus conditions. In Broad Focus, the durations of Word 1 and Word 2 are nearly identical, resulting in balanced timing across the phrase. However, in Narrow Focus, durations shift significantly to enhance the prominence of the focused word. When focus is on Word 1, its duration increases, while Word 2 shortens. Conversely, when focus is on Word 2, its duration extends in comparison to that of Narrow Focus-W1, while Word 1 compresses. These findings underscore the importance of temporal organization in marking focus, an area that has been relatively underexplored compared to pitch.

Similarly, Figure 4b demonstrates how F0 maxima ratios differentiate focus conditions. Under Broad Focus, the F0 values of Word 1 and Word 2 are relatively stable, producing a flat F0 slope across the noun phrase. This even distribution reflects the neutral prosody of broad focus. In Narrow Focus, however, the F0 contour dynamically adjusts to highlight the focused word. When Word 1 is in focus, its F0 maximum increases, while Word 2's F0 decreases, resulting in a descending slope. Conversely, when Word 2 is in focus, its F0 maximum rises while Word 1 decreases, creating an ascending slope. These shifts in pitch contour emphasize the focused element, reinforcing its perceptual salience.
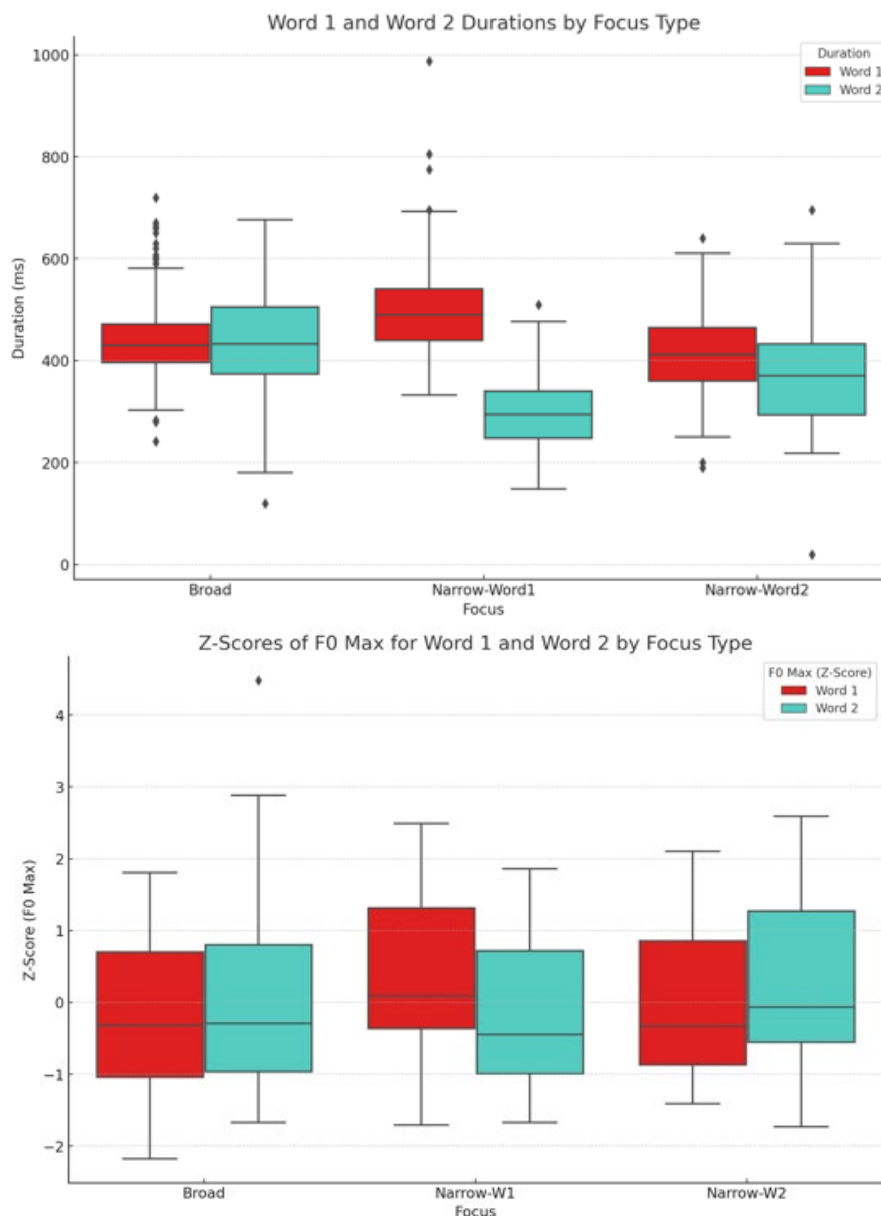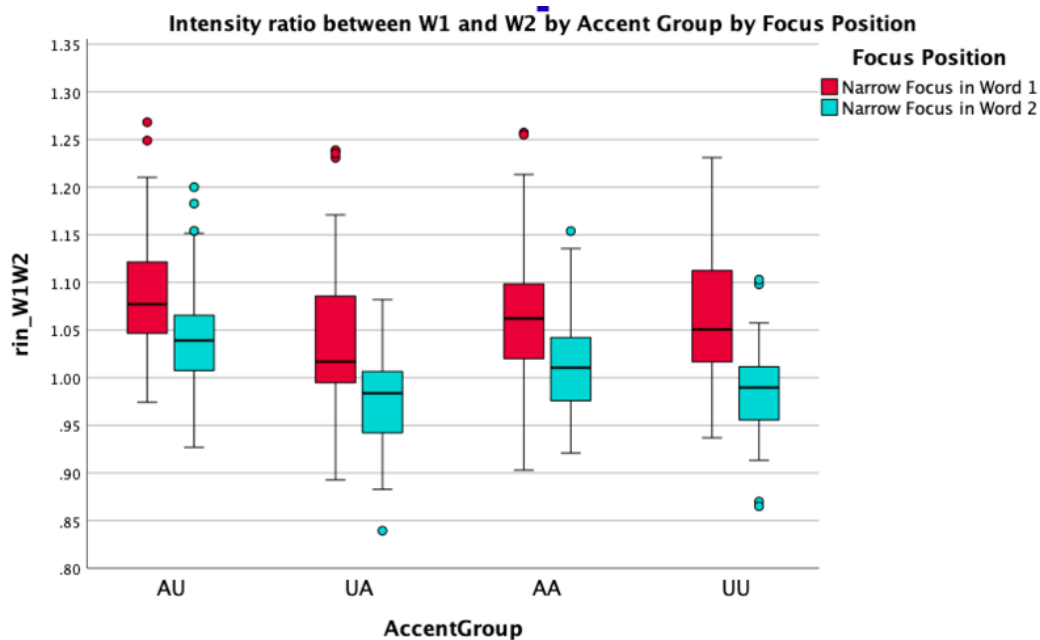
Figure 4. Word 1 and Word 2 Durations (a) and F0 maxima (b) by Focus Type.

Figure 5 illustrates how intensity ratios are influenced by both accent and focus position. The data reveal that intensity plays a crucial role in marking accent, with accented words consistently exhibiting higher intensity values than unaccented ones. This pattern holds across all focus conditions, but its effect varies depending on the accent group.

For phrases that begin with an accented word (e.g., AU, AA combinations), intensity ratios display notably higher values compared to those starting with an unaccented word. This leads to steeper intensity slopes, emphasizing the prominence of accented words over their unaccented counterparts.

Moreover, focus position modulates these intensity slopes. When focus is on the first word, the intensity slope is steeper, meaning the intensity difference between the first and second word is more pronounced. In contrast, when focus shifts to the second word, the intensity slope becomes less steep, as the intensity of the second word increases relative to the first. This modulation underscores how intensity not only cues accent but also interacts with focus to adjust prominence within the phrase.

In summary, Figure 5 highlights the dual role of intensity: it robustly signals accent and adapts dynamically based on focus position, further fine-tuning the prosodic marking of focus in educated standard Japanese.

Figure 5. Intensity ratios of Word 1 and Word 2 by Accent Group and by Focus position.

In summary, the results addressing research question 1a show that at the word level, duration and F0 maxima are the primary prosodic cues distinguishing Broad Focus from Narrow Focus and signaling the position of Narrow Focus. In Broad Focus, durations and F0 maxima are similar across words, while in Narrow Focus, the focused word is enhanced through increased duration and F0 maxima, with corresponding reductions in the non-focused word. These adjustments result in distinct F0 slopes and temporal patterns that highlight the focused element. In contrast, intensity primarily cues accent, with accented words displaying higher intensity than unaccented ones, but its role in focus marking is minimal. These findings confirm that in educated standard Japanese, focus at the word level is marked through temporal and pitch adjustments, with intensity playing a limited, accent-related role.

Table 3 summarizes the results of the linear mixed models addressing research question 1b, which examines whether duration, F0, and intensity ratios between the morphemes of the first word (e.g., *uma/no* from *umano hizume*) cue focus. Unlike the results at the word level, where duration and F0 played a significant role (Table 2), the intensity models at the morpheme level explain the most variation (32%), followed by F0 (25%), while duration models account for only 6% of the variance. This indicates that duration ratios between morphemes are largely irrelevant for explaining focus at this finer granularity.

F0 ratios remain effective in cueing focus, but not accent, consistent with the word-level findings. The higher values of F0 maxima in the root and -no in Narrow Focus-W1 differentiates this focus type from the rest. However, intensity emerges as the dominant cue at the morpheme level, primarily marking accent by producing accented words with higher intensity than unaccented words. Additionally, the main effect of Narrow Focus position and its interaction with Accent show that at this level of granularity intensity marks focus position. As shown in Figure 6, unaccented first words resort to an increase in intensity to signal focus. Specifically, when focus is on Word 1, the root morpheme is slightly louder than -*no*, maintaining a subtle prominence pattern. Conversely, when focus shifts to Word 2, -*no* becomes clearly louder than the root, creating a sharp prosodic distinction within Word 1. These results highlight that while

F0 continues to play a general role in distinguishing broad and narrow focus, intensity becomes a more critical factor in marking focus position in unaccented words.

Fixed Factors	Dependent Variables: Ratios Between Root and -no in Word 1		
	Duration	F0	Intensity
NF position	*	***	***
BF vs NF		**	
Accent			***
NF position * Accent			*
BF vs NF * Accent			
Model's size effect	6%	25%	32%

Table 3. Summary of linear mixed models examining the effects of focus and accent on the duration, F0, and intensity ratios between the root and the -no ending within Word 1. P-values are indicated as follows: '***' ≤ *0.0001*, '**' ≤ *0.001*, '*' ≤ 0.01, '.' ≤ 0.05.
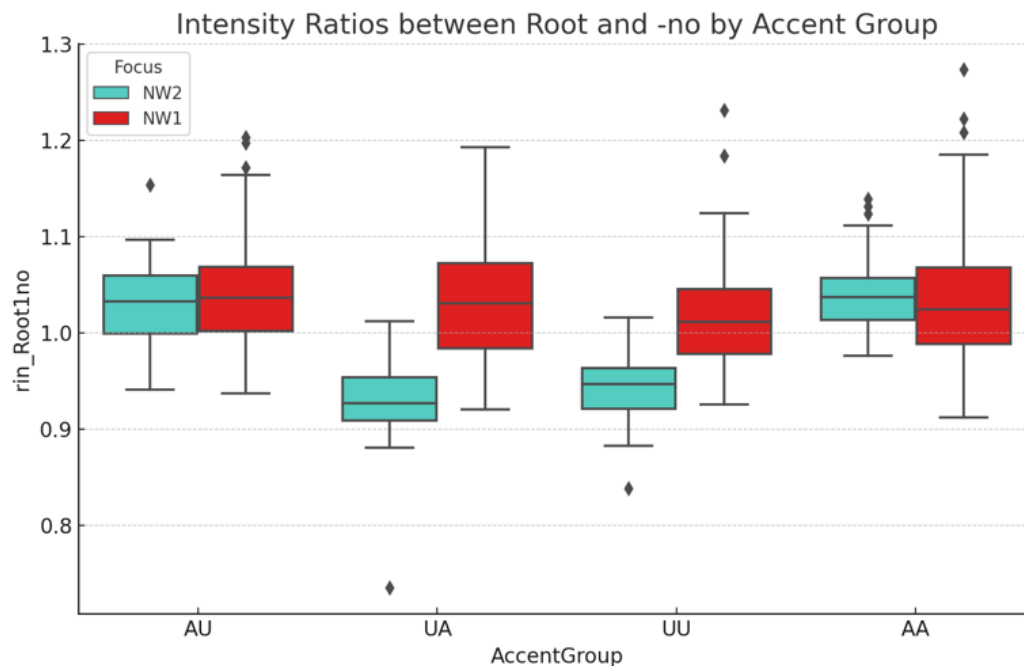
Figure 6. Intensity ratios of Root and -no by Accent Group and by Focus position.

For research question 2, the analysis revealed that the number of detected silences significantly differentiates Broad Focus from Narrow Focus and further distinguishes between Narrow Focus on Word 1 and Narrow Focus on Word 2(p < 0.001). As shown in Figure 6, silences were inserted in only 24% of cases under broad focus. In contrast, silence insertion increased substantially in narrow focus contexts, occurring in 69% of cases when focus was on Word 1 and 49% when focus was on Word 2. This pattern highlights the role of silences as a key prosodic boundary marker in distinguishing focus types in educated standard Japanese.
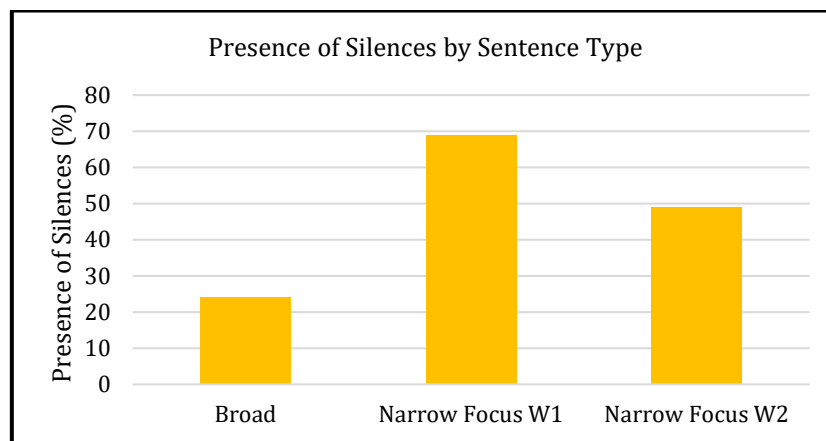
Figure 7. Percentages of inserted silences between Word 1 and Word 2 in Broad Focus sentences, Narrow Focus in Word 1, and Narrow Focus in Word 2

In conclusion, the results demonstrate that duration and F0 maxima are the primary cues for marking Broad Focus vs. Narrow Focus at the word level, with both cues enhancing the prominence of the focused word while compressing the non-focused word, thereby signaling both the presence and position of focus (research question 1a). At the morpheme level, intensity becomes the dominant cue, primarily marking accent, with its role modulated by focus position, particularly in unaccented contexts, while F0 continues to play a secondary role in marking focus (research question 1b). Additionally, silence insertion emerges as a significant prosodic boundary marker, with silences more frequently inserted in narrow focus contexts, particularly when focus is on Word 1, further differentiating focus types and positions (research question 2). Together, these findings highlight a nuanced prosodic system in educated standard Japanese, where duration and F0 are central to marking focus, while intensity cues accent and silences reinforce prosodic boundaries.

3.2. Interactions of Duration, F0, and Silences

A logit linear mixed model revealed that the two duration ratios, namely between words and morphemes, the two F0 maxima ratios, and the insertion of a silence between Word 1 and Word 2 all significantly contributed to differentiating broad and narrow focus. The model explained 42% of the variation, with the duration ratio between Word 1 and Word 2 and the insertion of a silence accounting for most of the variation. The contributions were as follows:

Effect	Rsq
Model	42%
Ratio between APs, duration	28%
Silence insertion	11%
Ratio between APs, F0 max	7%
Ratio between morphemes, F0 max	6%
Ratio between morphemes, Duration	4%

Table 4. Size effects of all the significant factors that contribute to differentiating broad focus from narrow focus.

A similar logit model to differentiate Narrow Focus on Word 1 from Narrow Focus on Word 2 failed to converge, indicating potential complexities or insufficient data in distinguishing between these specific focus conditions.

4. Discussion

This study addresses a key question in prosodic typology: do edge languages like educated standard Japanese employ different strategies than head languages like English to mark focus? This inquiry stems from the findings of Mizuguchi & Tateishi (2023), who showed that the focus expansion and compression strategy, central to English, works effectively in Japanese only when the focused word is accented and not placed at the end of a sentence. However, this strategy fails other contexts, e.g., unaccented words in-focus, raising the broader question of how focus is conveyed in such cases. To explore this, we investigated whether duration, F0, intensity, and silence insertion—examined at two levels of granularity, word and morpheme—offer alternative mechanisms for marking focus in educated standard Japanese. Our hypothesis was that these cues could shed light on the prosodic strategies employed by an edge language like educated Japanese, helping to clarify how focus is signaled across diverse contexts. In particular, we investigated first the contribution to focus marking of duration, F0 maxima, and intensity cues (question 1) as well as the addition of silences between words (question 2). Once relevant cues emerged, we explored the most effective cue combinations to mark focus (question 3).

In summary, those were our findings for each of the research questions. The results for question 1 demonstrate that at the word level, both duration and F0 maxima independently cue focus, irrespective of accent. In Broad Focus, durations and F0 maxima are similar across words, while in Narrow Focus, the focused word is enhanced through increased duration and F0 maxima, with corresponding reductions in the non-focused word. These adjustments result in distinct F0 slopes and temporal patterns that highlight the focused element. In contrast, intensity primarily cues accent, with accented words displaying higher intensity than unaccented ones, but its role in focus marking is minimal.

In contrast, at the morpheme level, intensity emerges as the most influential cue marking both accent and focus position. F0 maxima remains an effective tool marking focus independently of accent, while duration becomes largely irrelevant. Intensity not only marks accent—with accented roots and genitive morphemes (*-no*) being louder than their unaccented counterparts—but also plays a crucial role in signaling focus position in Noun Phrases containing an initial unaccented word. Specifically, when focus is on Word 1, the root morpheme is slightly louder than *-no*, maintaining a subtle prominence pattern. Conversely, when focus shifts to Word 2, *-no* becomes clearly louder than the root, creating a sharp prosodic distinction within Word 1.

The analysis for question 2 confirms that silence insertion between words 1 and word 2 is a crucial boundary marker, particularly in narrow focus contexts. Silences were significantly more frequent in narrow focus than in broad focus noun phrases, especially when focus was on Word 1. The increased use of silence reflects the edge-oriented strategy of Japanese, where prosodic boundaries—not heads—are emphasized to distinguish focus types and positions.

Putting together our findings for research questions 1 and 2, the emerging picture is as follows: there appears to be a clear division of labor among prosodic cues. Duration, F0, and silence insertion primarily cue focus, while intensity cues accent. In broad focus, words have similar durations, a flatter F0 slope, and no silences inserted, resulting in a uniform prosodic pattern. In narrow focus, this pattern shifts: silences are more frequently inserted, and focus position is signaled by enhancing the duration and F0 maxima of the focused word. However, this division of labor—where intensity marks accent while duration, F0, and silence cue focus—is not equally effective in all contexts. In cases where the first word is unaccented, a context identified by Mizuguchi & Tateishi (2023) as particularly challenging for focus marking, intensity takes on an additional role: it cues focus position at the morpheme level. Specifically, the intensity ratio between the root and -no in Word 1 signals focus: when focus is on Word 1, the root is louder than -no; when focus shifts to Word 2, *-no* becomes louder, marking the shift in focus.

These patterns confirm Mizuguchi & Tateishi's (2023) observation that focus marking in an edge language like educated Japanese involves a more complex system than in head languages like English, where F0 expansion in focus and post-focal compression are effective across a wide range of contexts. In Japanese, this pitch-based strategy works in tandem with duration and silence insertion at the word level. While silence insertion signals narrow focus by reinforcing word boundaries, focus position is conveyed by increasing the duration and F0 maxima of the focused word, enhancing its prominence. Thus, both boundary cues and prominence cues are employed to convey focus. However, these strategies do not work equally well in all contexts. In more challenging cases, such as unaccented initial words, Japanese compensates by using intensity ratios between morphemes to signal focus, demonstrating a flexible and context-sensitive prosodic system.

This leads us to question 3: do the cues that signal focus independently of accent—duration and F0 at the word level, and silences—contribute equally? The results highlight the complementary roles of these cues in focus marking. The combined model explained 42% of the variation in distinguishing broad focus from narrow focus, with duration and silence accounting for the majority of the variance. While F0 contributed less, its role in shaping prosodic prominence remains significant. These findings reinforce the idea that Japanese employs a nuanced system, where temporal and boundary cues dominate, but pitch modulation provides an auxiliary layer of focus marking.

In light of our findings, we return to the initial question: do edge languages like educated standard Japanese employ different focus-marking strategies than head languages like English? The emerging picture clearly suggests that they do. While head languages rely heavily on F0 expansion in focus and post-focal compression, these strategies are only partially effective in Japanese and are primarily limited to accented contexts. Instead, Japanese employs a more nuanced, context-sensitive system where temporal cues (e.g., duration) and boundary cues (e.g., silence insertion) dominate, while pitch modulation provides an auxiliary layer of focus marking. Furthermore, in challenging contexts like unaccented initial words, intensity becomes a critical cue for signaling focus, particularly at the morpheme level. This hybrid system reflects the typological differences between edge and head languages, with edge languages emphasizing prosodic boundaries and cue integration over reliance on pitch alone.

5. Conclusions and Implications for Further Research

This study contributes to the growing body of research exploring how edge languages like educated standard Japanese mark focus differently from head languages such as English. Our findings confirm that while Japanese employs pitch modulation in focus marking, it does so as part of a more complex system that relies heavily on temporal cues (e.g., duration) and boundary cues (e.g., silence insertion). These strategies effectively signal focus, particularly in contexts where F0 expansion and post-focal compression—the hallmark strategies of head languages—prove insufficient, such as with unaccented initial words. In these challenging contexts, intensity emerges as a critical cue, marking both accent and focus position at the morpheme level.

These results support the hypothesis that edge languages prioritize prosodic boundaries over prominence at prosodic heads and suggest that Japanese leverages a hybrid strategy. This system allows it to flexibly combine cues like duration, F0 maxima, intensity, and silence to adapt to different linguistic contexts.

Future research should explore how these production-based findings translate to perception, investigating whether listeners rely on the same prosodic cues to interpret focus distinctions. Additionally, studies should extend beyond noun phrases to other syntactic structures to assess

the generalizability of these focus-marking strategies. These avenues of research will provide further insights into the typological framework of prosodic systems and deepen our understanding of how edge languages manage focus marking across diverse communicative contexts.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models Usinglme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Beckman, M. E. (2012). *Stress and non-stress accent*. De Gruyter Mouton. https://doi.org/10.1515/9783110874020

Beckman, M. E., & Venditti, J. J. (2011). Intonation. In J. Goldsmith, J. Riggle, & A. Yu (Eds.), *The handbook of phonological theory* (pp. 485-532). Wiley Blackwell.

https://doi.org/10.1002/9781444343069.ch15

Beckman, M. E. & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*. 3, 55-309. https://doi.org/10.1017/S095267570000066X

Bodomo, A. B., & Marfo, C. O. (1996). The morphophonology of noun classes in Dagaare and Akan. *Nordic Journal of African Studies*, 5(2), 31–51.

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question–answer contexts. *The Journal of the Acoustical Society of America*, *77*(6), 2142-2156.

https://doi.org/10.1121/1.392372

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America*, *27*(4), 765-768. https://doi.org/10.1121/1.1908022

Igarashi, Y. (2014). Typology of intonational phrasing in Japanese dialects. *Prosodic typology II: The phonology of intonation and phrasing* (pp. 464–492). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199567300.003.0015

Igarashi, Y (2015). Intonation. In H. Kubozono (Ed.), *The handbook of Japanese phonetics and phonology* (pp. 539–567). De Gruyter Mouton. https://doi.org/10.1515/9781614511984.525

Jun, S. A. (2005). Prosodic typology. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 430–458). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199249633.003.0015

Jun, S. A. (2014). Prosodic typology: By prominence type, word prosody, and macro-rhythm. In S.-A. Jun (Ed.), *Prosodic typology II: The phonology of intonation and phrasing* (pp. 520–539). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199567300.003.0017

Kiss, K. É. (1998). Identificational focus versus information focus. *Language*, *74*(2), 245-273. https://doi.org/10.1353/lan.1998.0211

Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press. https://doi.org/10.1017/CBO9780511808814

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America*, *32*(4), 451-454. https://doi.org/10.1121/1.1908095

Maeda, K., & Venditti, J. J. (1998). Phonetic investigation of boundary pitch movements in Japanese. In *Proceedings of the 5th International Conference on Spoken Language Processing* (paper 0800). https://doi.org/10.21437/ICSLP.1998-124

Mizuguchi, S., & Tateishi, K. (2023). *Prominence in a Pitch Language: The Production and Perception of Japanese*. Rowman & Littlefield.

Ortega-Llebaria, M. (2006). Phonetic cues to stress and accent in Spanish. In *Selected proceedings of the 2nd conference on laboratory approaches to Spanish phonetics and phonology* (pp. 104-118). Cascadilla Proceedings Project.

Ortega-Llebaria, M., & Prieto, P. (2011). Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Language and speech*, *54*(1), 73-97. https://doi.org/10.1177/0023830910388014

Seo, J., Kim, S., Kubozono, H., & Cho, T. (2019). Preboundary lengthening in Japanese: To what extent do lexical pitch accent and moraic structure matter? *The Journal of the Acoustical Society of America, 146*(3), 1817-1823. https://doi.org/10.1121/1.5125186

Venditti, J. J. (2005). The J_ToBI model of Japanese intonation. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 172-200). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199249633.003.0007

Venditti, J. J., Maekawa, K., & Beckman, M. E. (2008). Prominence marking in the Japanese intonation system. In S. Miyagawa & M. Saito (Eds.), *The Oxford handbook of Japanese linguistics* (pp. 456–512). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195307344.013.0017

WEBMAUS forced aligner. (n.d.). https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics, 33*(2), 159–197. https://doi.org/10.1016/j.wocn.2004.11.001