Do Japanese speakers always prosodically group wh-elements and their licenser? Implications for Richards' (2010) theory of wh-movement*

Shigeto Kawahara, Jason Shaw, Shinichiro Ishihara

May 11, 2020

Abstract

The relationship between syntactic structure and prosodic structure has received increased theoretical attention in recent years. Richards (2010) proposes that Japanese allows wh-elements to stay in situ because of an aspect of its prosodic system. Specifically, in contrast to some other languages (like English), Japanese can prosodically group wh-elements together with their licensers. This prosodic grouping is signaled by eradication or reduction of the lexical pitch accents of intervening words. In this theory, a question still remains as to whether each syntactic derivation is checked against its prosodic realization, or whether what allows Japanese wh-elements to stay in situ is more abstract in character. To address this question, we test whether and how the evidence of the prosodic grouping is observed in naturalistic utterances. Our analysis makes use of a computational toolkit that allows us to assess the presence of tonal targets on a token-by-token basis. Our analysis shows that almost all speakers produce some sentences that show prosodic grouping of wh-elements and their licensers, as well as some that do not. Our results generally support the prediction of Richard's (2010) theory, and they furthermore suggest that the prosodic properties of Japanese that allow their wh-elements to stay in situ must be an abstract grammatical feature of Japanese. We conclude that, while language-specific syntax operates in the presence of language-specific prosody, syntax does not vary to accommodate variation in prosody, as least not in this case.

1 Introduction

The relation between syntactic structure and prosodic structure has received increased theoretical attention in recent years. The standard feedforward model, in which the syntactic derivation is computed first and prosody follows, has been challenged by apparent cases in which prosody and other phonological factors influence the syntactic derivation (e.g. Anttila,

*We received helpful comments from the participants at the following conferences, AMP 2019, HisPhonCog 2019, and ICPP 2019, especially Mary Beckman, Ryan Bennett, Edward Flemming, and Haruo Kubozono. They bear no responsibilities for the remaining errors.

Adams, and Speriosu 2020; Bennett, Elfner, and McCloskey 2016; Breiss and Hayes 2020; Shih and Gribanova 2016; Shih and Zuraw 2017).[1] We take up an example from the widely discussed syntactic phenomenon of wh-movement.

Languages differ with respect to whether wh-phrases move overtly or not: Tagalog wh-phrases move overtly, whereas Japanese wh-phrases can stay in situ (or move covertly). In Minimalist Syntax (Chomsky 1995), this difference was stipulated to derive from a difference in feature strength. Strong (or uninterpretable) features, which Tagalog wh-elements have, need to be checked in the syntax, which requires overt movement, whereas weak features associated with Japanese wh-elements can be checked at LF. This account restates the difference between overt and covert movement in terms of feature strength. Richards (2010) pursues a more explanatory theory of this cross-linguistic variation in wh-movement and attempts to derive differences between overt movement and covert movement (or lack of overt movement) from independent properties of each language (see also Richards 2016 for an extension of this proposal to a wider range of syntactic phenomena).

More specifically, Richards (2010: 145) argues that there is a language universal principle—all languages attempt "to create a prosodic structure for wh-questions in which the wh-phrase and corresponding complementizer are separated by as few prosodic boundaries as possible." This prosodic grouping is accomplished in Tagalog via wh-movement. Inspired by a body of work on Japanese intonational patterns (Deguchi and Kitagawa 2002; Ishihara 2003; Hirotani 2005; Sugahara 2003; Smith 2005), Richards (2010) proposes that Japanese has a prosodic means to group the wh-phrase and its complementizer, and hence does not need to resort to overt wh-movement.[2]

This theory can be interpreted in two ways: one is that whether a wh-element moves or not is determined for each syntactic derivation; the other is that there is an abstract prosodic feature of Japanese that allows wh-elements to stay in situ, and it is not necessarily the case that each syntactic derivation is checked against its prosodic realization. To address this question, one of the aims of this paper is to elicit prosodic patterns from naive participants in a controlled experiment in order to examine if—and how—wh-sentences in Japanese show evidence of prosodic grouping between wh-elements and their licensers. Our results show that there is prosodic variation, which implies that the same syntactic derivation can be produced with different prosodic structures. Some instances show evidence for prosodic grouping, while others do not, indicating that what allows Japanese to have wh-elements in situ is abstract in character.

Some background information regarding the Japanese prosodic system is in order. In Japanese, a Minor Phrase (MiP) contains at most one accented lexical item, and is signaled by a phrase-initial %LH rise and a H*+L accentual fall (e.g. Igarashi 2015; Pierrehumbert and Beckman 1988; Venditti 2005; Venditti, Maekawa, and Beckman 2008).[3] Deguchi and

[1] See footnote 1 in Bennett, Elfner, and McCloskey (2016) for a list of relevant proposals in which phonological factors can influence word order.

[2] Smith (2011) argues based on the data from Fukuoka Japanese that it is the complementizer, not the wh-elements, that derives this phrasing pattern. We are not concerned in this paper what triggers this prosodic grouping.

[3] Minor Phrase is also known as an Accentual Phrase. Major Phrase, the level above Minor Phrase, is also known as an Intermediate Phrase. Terminological differences do not concern us here (see Igarashi 2015 for a recent systematic review). We use the term Minor Phrase, because this is what Richards (2010) uses. See Richards (2016) for a proposal which deploys a recursive prosodic structure without a Minor Phrase/Major

Kitagawa (2002), one source of inspiration for Richards (2010), argue that these tonal events associated with Minor Phrases are *eradicated* after wh-elements up to the complementizer that licenses the wh-elements, effectively grouping the wh-phrase and complementizer within a single Minor Phrase.[4] This eradication is accompanied by a boost of accentual rise on the wh-element itself, which is arguably an instance of a more general focus-induced prominence boost (Igarashi 2015, Ishihara 2015 and Venditti, Maekawa, and Beckman 2008 and many references cited therein). Deguchi and Kitagawa (2002: 74) state:

> Another important prosodic effect of focus pointed out by Ishihara (2000) (extending the original observation by Ladd (1996)) is that an emphatic accent is accompanied by what we label as "**eradication**" of lexical accents. That is, when one or more of [the] lexical accents follow an emphatic accent, their H tones (H*) are all suppressed. As a result, the lowest f0 induced by the emphatic accent is inherited and prolonged with further gradual declination up to the right boundary of some clausal structure (emphasis in the original).

An oft-cited pair of pitch tracks is provided in Figure 1, which is reproduced from Ishihara (2003: 53). The top panel shows a declarative sentence consisting of four words which are all lexically accented. The bottom panel shows a corresponding wh-sentence, in which the second word is a wh-element, *nani-o* "what-ACC," shown by an arrow. The domain of eradication in the wh-sentence (or reduction: see below) is shown in grey.

If Deguchi and Kitagawa's observation is correct, then we have a very simple story in accordance with Richards' theory—Japanese groups wh-elements and the complementizer within a single Minor Phrase without any intervening Minor Phrase boundaries. The prosodic structure that would reflect Deguchi and Kitagawa's observation is schematically depicted in (1). This structure is also entertained by Richards (2010: 145), although he ultimately adopts a different structure, to be discussed below, in (2).

(1) $_{\mathrm{MiP}}$[wh DP DP V Comp]

If (1) is correct, any tonal events in the intervening DPs should be lost, *a la* Deguchi and Kitagawa (2002); i.e., the intonational contour should be "flat" throughout the whole Minor Phrase. We note at this point that the claim by Deguchi and Kitagawa (2002) is based on the intonational contours produced by the authors. They note the preliminary nature of the data and the need to test the claim about tone eradication with more objective methodology. We provide this test in the current study.

Some later studies cast doubt on Deguchi and Kitagawa's (2002) claim that accent following wh-elements is eradicated—instead, the post-wh accent may simply be reduced (see Hirotani 2005; Ishihara 2003; Ishihara 2011; Sugahara 2003—see also Maekawa 1994). Ishihara (2003: 32) for example notes that if all tonal events after wh-elements are eradicated,

Phrase distinction (e.g. Ito and Mester 2012). In this paper we follow Richards' (2010) conventions, as the predictions regarding phonetic realizations are straightforward to illustrate.

[4]See also Igarashi (2015), Pirrehumbert and Beckman (1988), Venditti et al. (2008) and Ishihara (2015) and works cited therein for (de)phrasing that may occur in post-focal positions in general. Most of these studies, however, posit that dephrasing occurs at the level of the Major Phrase rather than the Minor Phrase. In the current paper, we focus on the proposal by Deguchi and Kitagawa (2002) and specifically analyze those contexts that are relevant to wh-constructions in Japanese.

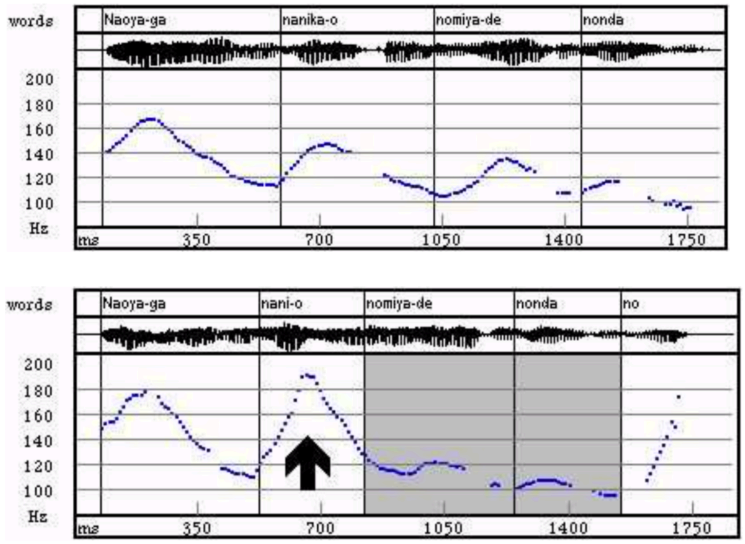

Figure 1: Illustrative pitch tracks of a declarative sentence and a wh-sentence in Japanese. Reproduced from Ishihara (2003: 53). These tokens are based on the production of Ishihara himself. Top: declarative, "Naoya drank something at a bar." Bottom: wh-sentence, "What did Naoya drink at a bar?"

the lexical distinction between accented and unaccented words should be neutralized after wh-elements. Though he did not report any experimental results, Ishihara suggests that this neutralization does not occur. Hirotani's (2005) production study showed substantial variation across speakers in the degree of compression after wh-elements. Richards (2010: 149-151) posits the prosodic structure schematized in (2) for Japanese, in which wh-elements and the complementizers are grouped in a higher recursive Minor Phrase. The recursive Minor Phrase was first proposed for Japanese by Kubozono (1988); more recent work has identified potential problems with this analysis, suggesting instead that the higher prosodic level is a Major Phrase. The issues with the recursive Minor Phrase, pointed out by Ito and Mester (2012), is that since Minor Phrases are usually defined in terms of accent culminativity (i.e., at most one accent), a recursive structure should not be possible (except for very special cases in which all the terminal Minor Phrases contain an unaccented item). For our purposes, the difference between a Major Phrase and recursive Minor Phrases is minimal. However, if the higher level prosodic structure is indeed a Major Phrase, then the tonal events of the intervening DPs should be reduced due to independently observed downstep, as the domain of downstep is a Major Phrase (McCawley 1968; Pierrehumbert and Beckman 1988; Poser 1984) (cf. Ishihara 2016). Unlike (1), then, the prosodic structure in (2) predicts that accentual rises in the intervening DPs should be reduced, compared to when they occur in non-wh-contexts.

(2) ${}_{\text{MiP(MaP)}}[{}_{\text{MiP}}[\text{wh}]\ {}_{\text{MiP}}[\text{DP}]\ {}_{\text{MiP}}[\text{DP V Comp}]]$

Ishihara (2011) reports an instrumental study that suggests that reduction (instead of eradication) takes place after wh-elements, as predicted by the structure in (2). However, as we will discuss more fully below, Ishihara (2011) only examines averaged contours, and a token-by-token analysis is not reported. As we show below, however, looking at averaged contours can be misleading. From averages, it is not possible to differentiate between reduction and variability, phonetic outcomes which often map onto distinct theoretical proposals both in this case and more generally (Cohn 2006; Shaw and Kawahara 2018). In other words, if the phonetic realizations are variably either "fully realized" or "completely eradiated," averaging over these two categorical generalizations would not be distinguishable from "reduction."

To summarize, Richards' (2010) theory of wh-movement assumes that Japanese resorts to some sort of prosodic means in one way or another to group wh-elements and their licensers. Since past work has shown variation in f0 patterns following wh-elements (e.g., Hirotani 2005), we feel that it is important to test for tone eradication/reduction on a token-by-token basis. The time is particularly appropriate given the recent development of new computational methods for assessing on a token-by-token basis whether a phonetic target is reduced, deleted or fully realized (Shaw and Kawahara 2018). This new method allows us to tease apart two possible interpretations of Richards' (2010) theory: whether each syntactic derivation is checked against its prosodic realization, or whether what allows Japanese wh-elements to stay in situ is more abstract in character.

2 The current study

Against this theoretical background, this paper reexamines wh-phrase conditioned tone eradication/reduction in Japanese. We adopt a token-by-token analysis of intonational contours, using the computational toolkit developed by Shaw and Kawahara (2018). The basic approach is to classify intonational contours based upon competing phonological hypotheses, in this case the presence/absence of an H tone. Each intonational contour is assigned a probability of being generated from a LHL pitch accent or a LΦL pitch accent, where Φ represents syllables unspecified for f0. A key step in mapping the continuous trajectories to discrete phonological hypotheses is a low-dimensional representation of the f0 signal, which is accomplished using Discrete Cosine Transform. More broadly, the approach fits within the broader analytical strategy of functional data analysis (e.g., Beddor et al. 2018; Krivokapić, Styler, and Parrell 2020; Lee, Byrd, and Krivokapić 2006), as the low dimensional representation is achieved by fitting non-linear functions to the data. A key innovation of the approach is that phonetic interpolation, characteristic of LΦL, is formalized stochastically. First developed to assess the presence/absence of a lingual vowel target of devoiced vowels in articulatory trajectories, the approach is general and has been extended to other types of continuous phonetic data, including nasal reduction in Ende (Brickhouse and Lindsey 2020) and tone reduction in Mandarin Chinese (Zhang, Geissler, and Shaw 2019). In addition to being able to analyze each tonal contour on a token-by-token basis, this method has an advantage of being able to analyze the whole intonational contour without relying on "magic moments" (Mücke, Grice, and Cho 2014), i.e., particular aspects of the phonetic signal, such as f0 minima or maxima, are not given privileged status in the analyses.

To preview the results, our analysis shows that there is prosodic variation. Almost all speakers produce some sentences that behave as predicted by Richards' (2010) theory, as well as some that do not. We conclude that the prosodic properties of Japanese that allow their wh-elements to stay in situ is an abstract feature of Japanese, and it is not the case that each syntactic derivation has to be checked against its phonetic/prosodic realizations. The overall conclusion with respect to the case under study is that, while language-specific syntax may operate in the presence of language-specific prosody, syntax does not vary to accommodate variation in prosody.

2.1 Methods

2.1.1 Materials

The current study reanalyzes a subset of the data recorded by Ishihara (2011). The corpus features carefully controlled pairs consisting of a declarative sentence and wh-question counterpart. All sentences consisted of five words. Schemata of the item pairs are given in (3) and (4), together with one example sentence for each condition. Accent is shown by an apostrophe (') following accented vowels. Word2, Word3, and Word4 were all lexically accented. For wh-questions, the second word was the wh-phrase. There were six types of sentences with different lexical items for (3) and (4).

(3) Declarative sentence (control): Word1 Word 2[-wh] Word3 Word4 Verb

Maruyama-wa	e'rumesu-no	eri'maki-ni	nomi'mono-o	kobo'shita
NAME-TOP	Hermes-GEN	scarf-DAT	drink-ACC	spilt

"Maruyama spilled drink over Hermes scarf."

(4) Wh-question sentence (test): Word 1 Word 2[+wh] Word3 Word4 Verb

Maruyama-wa	do'nohito-no	eri'maki-ni	nomi'mono-o	koboshi	mashi'ta	ka?
NAME-TOP	whose	scarf-DAT	drink-ACC	spilt	POL-PAST	COMP

"What type of scarf did Maruyama spilled drink over?"

The sentences in (3) serve as the control sentences. Both Word3 and Word4 are Minor Phrases; as such, the phrasal tones (%LH) and lexical accent (H*+L) of both Word3 and Word4 are realized (i.e. full target), as shown in a sample pitch track given in Figure 2 (top).

We are interested in whether Word3 and Word4 in (4) (=Figure 2, bottom) also fully retain these %LH*+L tones, or whether these tones are completely eradicated or reduced.

2.1.2 Procedure

The stimuli contain six items per each of the two conditions shown in (3) and (4) (i.e., $2 \times 6 = 12$ target sentences). They were recorded together with another 142 filler sentences (which were recorded for other studies).

Nine native speakers of Tokyo Japanese (4 females and 5 males) read all the stimulus sentences, twice each. One stimulus sentence was presented to a speaker per trial on a computer screen. Speakers were allowed to repeat the sentence when they made a mistake or they felt that their utterance was unnatural, in which case only the last rendition was used for the following analysis. The order of the stimuli was pseudo-randomized. Two recordings

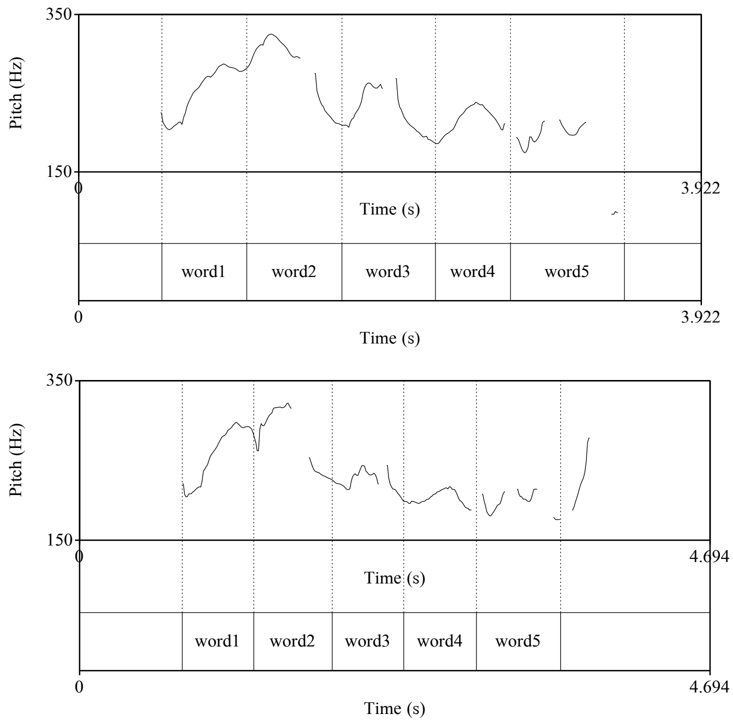

Figure 2: Sample pitch tracks from the current recordings. Top: declarative sentence (= sentence (3)). Bottom: wh-sentence (= sentence (4)).

per speaker were made in two different randomized orders. A total of 432 tokens (12 target sentences read by 9 speakers twice each for Word3 and Word4) entered into the subsequent analysis.

2.1.3 Analysis

The intonational contours of Word3 and Word4, delimited by %L and +L, were extracted using YAAPT, a robust pitch tracking algorithm (Yet Another Algorithm for Pitch Tracking: Zahorian and Hu 2008). We then applied the computational toolkit developed by Shaw and Kawahara (2018).

The starting point of the approach is to recognize that, given a phonetic trajectory, it is often hard to decide, especially if we rely on visual inspection of pitch tracks, whether that trajectory should be characterized as linear interpolation between two targets (with declination), or whether it has a distinct phonetic target (i.e. in the case at hand, a H tone), as schematically illustrated in Figure 3. This is particularly so because actual intonational contours always involve natural variability, due to various factors such as consonantal perturbations (Hombert, Ohala, and Ewan 1979), the influences of vowel height (Whalen and Levitt 1995), and others; in other words, intonational contours are usually “bumpy”, and it is often hard to tell whether a “bump” comes from an actual phonological specification or merely due to random variation (noise).

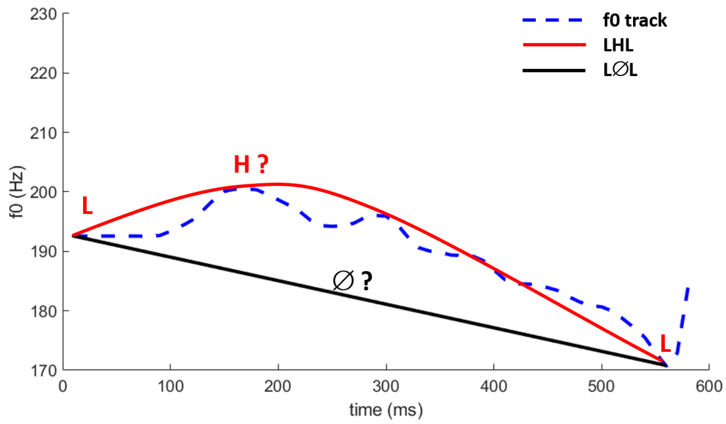

Figure 3: Deciphering whether a contour has a H(igh) tonal target. The dotted blue line, representing an actual phonetic contour, is shown together with phonetic schemata for competing phonological hypotheses, LΦL (solid black line) and LHL (solid red line).

To address this question in an objective fashion, we first transform the phonetic trajectories (i.e. intonational contours) into a low-dimensional space, so that we have a mathematical handle on them. For this purpose, we use Discrete Cosine Transform (DCT) to transform phonetic signals from the time domain (changes in f0 over time) to the frequency domain (sums of cosines of different frequencies and amplitudes). The numerical expression of DCT is given in formula (1)-(2). Like Fourier Transform, this analysis decomposes trajectories into a set of frequency components.

$$y(k) = w(k) \sum_{n-1}^{L} \cos \frac{\pi(2n-1)(k-1)}{2L} \quad k = 1, 2, ...L \tag{1}$$

where

$$w(k) = \begin{cases} \dfrac{1}{\sqrt{L}} & k = 1 \\ \sqrt{\dfrac{2}{L}} & 2 \leq k \leq L \end{cases} \tag{2}$$

The f0 contour across Word3 and Word4 can be represented faithfully in the frequency domain using four DCT components. We determined this using the inverse function for DCT, iDCT, which transforms cosines (frequency domain) back to f0 trajectories in the time domain. The numerical expression of iDCT is given in formula (3)-(5).

$$y(k) \sim \; N(\mu(k), \sigma(k)) \tag{3}$$

$$x(n) = \sum_{n-1}^{L} w(k) y(k) \cos \frac{\pi(2n-1)(k-1)}{2L} \qquad n = 1, 2, ...L \tag{4}$$

where

$$\begin{cases} \frac{1}{\sqrt{L}} & k = 1 \\ \sqrt{\frac{2}{L}} & 2 \leq k \leq L \end{cases} \tag{5}$$

Using iDCT, we simulated pitch trajectories from different numbers of DCT coefficients and compared the simulated trajectories to the actual trajectories. For the case at hand, using four DCT coefficients achieves higher than 95% fit between actual and simulated trajectories.

Having verified that we can faithfully represent f0 trajectories (time domain) with four DCT coefficients (frequency domain), we proceeded to set up stochastic generators of our two competing phonological hypotheses, LHL and LΦL in Figure 3, in the frequency domain. Gaussian distributions over DCT components for the control sentences (3) describe the LHL hypothesis. Both the mean and the standard deviation of the distributions were determined by the data. Gaussian distributions for LΦL were defined with reference to linear interpolation between the two L tones in the test sentences (4). The mean of the distributions was determined by DCTs fit to the linear interpolation. The standard deviation was set to the same standard deviation for the LHL hypothesis. The LΦL is thus a realization of linear interpolation with the same level of variability as the control sentences (3). Contours created from the LΦL generator, using iDCT, simulate how the LΦL line in Figure 3 would be phonetically realized given naturalistic variability.

Finally, we used the stochastic generators of our phonological hypotheses as a Bayesian classifier (formula (6)). The classifier assigned posterior probabilities to each test token, i.e., Word3 and Word4 following wh-elements in the test sentences (4). The posterior probability represents the likelihood that the token was generated by one phonological hypothesis or the other. Since the probabilities are complementary, we report, for each token, the posterior probability that it was generated by the LΦL hypothesis, indicating tone eradication.

$$\prod_{i=1}^{n} p(T|Co_1, ..., Co_n) = \frac{p(T) \times \prod_{i=1}^{n} p(Co_i|T)}{\prod_{i=1}^{n} p(Co_i)} \tag{6}$$

2.2 Results

Figure 4 shows the posterior probability of eradication for Word3. Tokens on the right have a high probability of being generated by the LΦL model (tone eradication). Tokens on the left have a high probability of being generated by the LHL model, as in our control sentences (3) (=full target). Many speakers (Speakers 1, 2, 4, 5, 7, 8, 9) show at least some tokens that have a high posterior probability of eradication (right). These tokens thus support the view expressed by Deguchi and Kitagawa (2002); i.e. they instantiate tonal patterns deriving from the prosodic structure posited in (1) and directly support Richards' (2010) argument that wh-elements and their licensers are grouped within a single Minor Phrase in Japanese.

However, Speakers 6, 7, 8, and 9 show a large number of tokens that have a high probability of being generated by the LHL model, i.e., no different from tonal realizations in declarative sentences. These tokens show no or very little trace of reduction (near 0 probability of LΦL). We finally observe those tokens whose posterior probabilities are in the middle range (Speakers 2, 4, 5, 7). These tokens show properties that are intermediate between the two phonological hypotheses, and thus are best viewed as phonetically reduced, instantiating tonal patterns predicted by the structure in (2). For most speakers (all except Speaker 3), within-speaker variability is also evident.

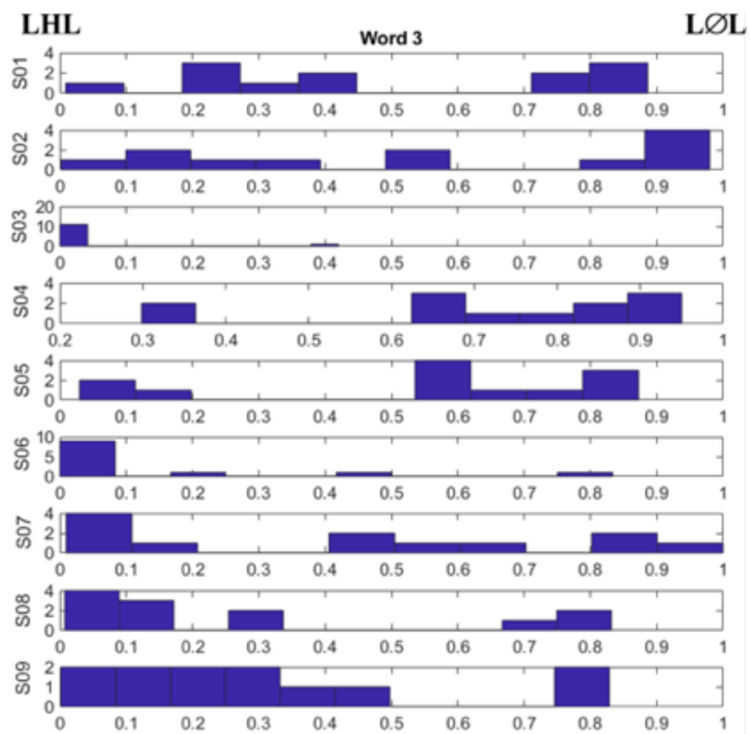

Figure 4: Posterior probabilities of tone eradication for Word3. For visibility, the scale of the vertical axis is optimized for each speaker.

Figure 5 shows the posterior probability of eradication for Word4. The structure of the figure is the same as Figure 4. Most speakers (all but Speakers 1 and 6) show tokens of high eradication probability (right). Speakers 1, 3, 5, 6, and 8 also produced tokens with full tonal targets (left) and there are many tokens as well that are phonetically reduced (middle). Again, just like Word3, we observe both inter- and intra- speaker variability.

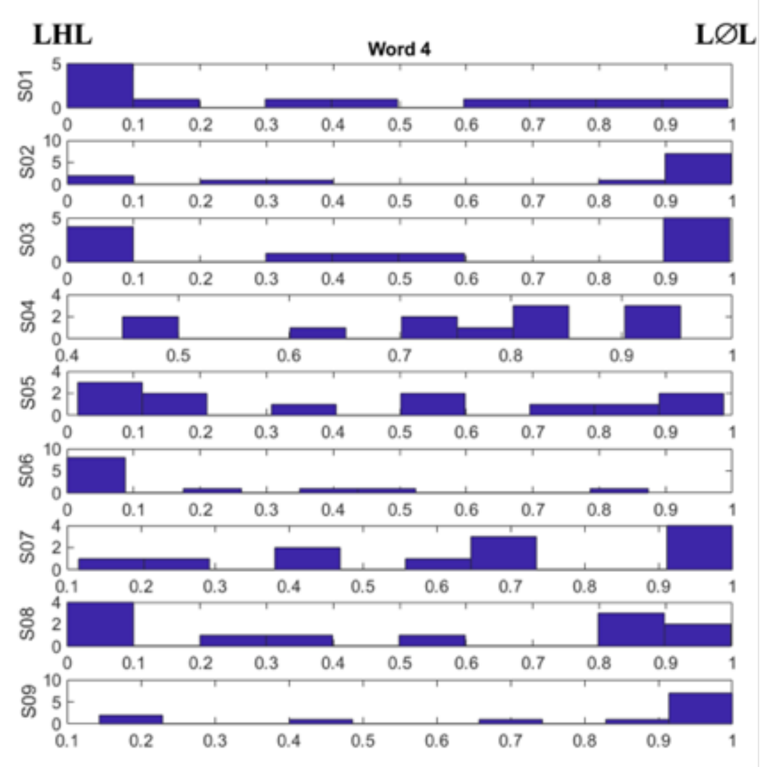

Figure 5: Posterior probabilities of tone eradication for Word4. For visibility, the scale of the vertical axis is optimized for each speaker..

At this point we would like to highlight the importance of analyzing each token separately, instead of just looking at averaged contours, as Ishihara (2011) did. Take the case of Word4 for Speakers 3 and 4, for example. Speaker 4 shows reduction for all tokens; Speaker 3 on the other hand shows a bimodal distribution of full targets and eradication. If we were to be only looking at averages, we would have erroneously concluded that both speakers show reduction (Cohn 2006; Shaw and Kawahara 2018). This comparison highlights the importance of analyzing each token separately. We would also like to note that our analyses do not grant any privileges to "magic momoents" in intonational contours, such as f0 minima or maxima, but analyze the whole contours in their entity, thus eschewing the danger of missing important aspects of dynamic speech (Cho 2016; Mücke, Grice, and Cho 2014; Vatikiotis-Bateson, Barbosa, and Best 2014).

3 General Discussion

Our analysis shows that most speakers do show some tokens in which %LH*+L tones are eradicated (i.e. tokens which can be characterized as linear interpolation between %L and L), supporting the prosodic structure in (1). Some tokens were best viewed as showing reduction, suggesting the prosodic structure in (2). At the same time, however, no speakers consistently show eradication, and more importantly, there are tokens that show a high probability of LHL structure (i.e. no eradication or reduction). These results allow us to tease apart two interpretations of Richards' (2010) theory, laid out in the introduction: whether (1) each syntactic derivation is checked against its prosodic manifestations, or whether (2) the relationship between wh-movement and the prosodic structure is more abstract. Our results are consistent only with the second interpretation. The first interpretation would predict that all sentences with wh elements in situ prosodically group the wh-element and its complementizer. Our data show that this is not always the case, but there is instead prosodic variation. Nevertheless, with the possible exception of Speaker 6, who produced just two out of 24 tokens (one for Word3 and one for Word4) with a high (greater than .6) probability of LΦL, all of our speakers produced at least some tokens with a high probability of tone eradication. That is, prosodically grouping the wh-word and the complementizer is a strategy that is available to all speakers, even though it is not obligatory. We surmise that the availability of prosodic grouping is sufficient to license wh in-situ.

Throughout this paper, we have been following a basic assumption of research on intonation, that prosodic structure can be inferred from surface tonal events. This assumption led us to the conclusion that the source of variability comes from variable prosodic structures. It is possible, however, that the prosodic structure is always the same, but the relationship between the prosodic structure and their phonetic cues can be variable. Under this interpretation, whose general implications for theories of intonation are yet to be explored, speakers do not necessarily signal the prosodic grouping between wh-elements and their licensers phonetically. This interpretation likewise requires us to posit some abstraction: speakers have abstract prosodic structures whose phonetic cues are not necessarily signaled in actual utterances.[5]

All in all, our results provide robust quantitative support from a reasonable sample of naturally produced utterances that tone eradication/reduction, as assumed by Richards (2010), does indeed happen in Japanese. However, the process is variable, which implies that what is crucial is not tone eradication/reduction *per se* but rather the availability of the prosodic grouping of wh-phrases and their licensers. It is not the case that a prosodic grouping pattern is checked for each syntactic derivation. That would prevent wh in situ just when the wh-phrase and its licenser are not prosodically grouped. Instead we conclude that it is a more abstract prosodic feature of Japanese, related to the availability of the grouping, that allows wh-phrases to stay in situ.

Our results show that the same syntactic derivation can map to different prosodic struc-

[5]This possibility may have been anticipated by Richards, when he states (2010: 148) that "[w]hat kind of effect these wh-domains have on f0 is not part of the theory: wh-domains might involve f0 compression, a high tone, or (in principle) no prosodic effects at all." Richards (2010) thus does allow for the presence of a language that groups wh-elements and their licensers together, but does not overtly signal that grouping in any prosodic means. As we have seen, however, Japanese is a language that does signal wh-domains either by reduction or eradication; what we are finding is that not every token shows that evidence.

tures. Specifically, the same syntactic derivation for Japanese questions is sometimes produced with tone eradication or reduction, indicating shared prosodic constituency, can also be produced with prosodic boundaries intervening between the wh-element and its complementizer. In the growing literature on syntax-prosody interface, it is still unclear how such prosodic variation may influence syntactic derivations. Our results may suggest that the syntactic derivation can proceed without necessarily referencing the phonological representation of the utterance under construction. The mere availability of prosodic grouping may be sufficient for the syntax to license wh in situ. In this sense, the syntax may "trust" in appropriate prosodic grouping, allowing the derivation to proceed, even as the prosody at times betrays this trust. To be clear, we are not presenting the current results as evidence against Richards' (2010) proposal. Language-specific syntax operates in the presence of language-specific prosody. Our conclusion regards the nature of this relation. The strongest hypothesis that follows from our results is that syntax does not vary to accommodate variation in prosody.

References

Anttila, Arto, Matthew Adams, and Michael Speriosu (2020). "The role of prosody in the English dative alternation". In: *Language and Cognitive Processes* 25, pp. 946–981.

Beddor, Patrice S. et al. (2018). "The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change". In: *Language* 94.4, pp. 931–968.

Bennett, Ryan, Emily Elfner, and James McCloskey (2016). "Lightest to the right: an anomalous displacement in Irish". In: *Linguistic Inquiry* 47.2, pp. 169–234.

Breiss, Canaan and Bruce Hayes (2020). "Phonological markedness effects in sentential formation". In: *Language* to appear.

Brickhouse, C.J. and Kate Lindsey (2020). *Investigating the phonetics-phonology interface with field data: Assessing phonological specification through acoustic trajectories.* Poster presented at the 96th meeting of the Linguistics Society of America.

Cho, Taehong (2016). "Prosodic boundary strengthening in the phonetics-prosody interface". In: *Language and Linguistic Compass* 10.3, pp. 120–141.

Chomsky, Noam (1995). *The Minimalist Program.* Cambridge, MA: MIT Press.

Cohn, Abigail (2006). "Is there gradient phonology?" In: *Gradience in Grammar: Generative Perspectives.* Ed. by Gisbert Fanselow et al. Oxford: Oxford University Press, pp. 25–44.

Deguchi, Masanori and Yoshihisa Kitagawa (2002). "Prosody and wh-questions". In: *Proceedings of NELS* 32, pp. 73–92.

Hirotani, Masako (2005). "Prosody and LF Interpretation: Processing Japanese Wh-questions". Doctoral Dissertation. University of Massachusetts, Amherst.

Hombert, Jean-Marie, John Ohala, and William G. Ewan (1979). "Phonetic explanations for the development of tones". In: *Language* 55, pp. 37–58.

Igarashi, Yosuke (2015). "Intonation". In: *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology.* Ed. by Haruo Kubozono. Mouton: Mouton de Gruyter, pp. 525–568.

Ishihara, Shinichiro (2000). *Scrambling and its interaction with stress and focus.* Ms. MIT.

Ishihara, Shinichiro (2003). “Intonation and Interface Conditions”. Doctoral Dissertation. MIT.
— (2011). “Focus prosody in Tokyo Japanese wh-questions with lexical unaccented wh-phrases”. In: *Proceedings of ICPhS* XVII, pp. 946–949.
— (2015). “Syntax-phonology interface”. In: *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology.* Ed. by Haruo Kubozono. Mouton: Mouton de Gruyter, pp. 569–618.
— (2016). “Japanese downstep revisited”. In: *Natural Language and Linguistic Theory* 34, pp. 1389–1443.
Ito, Junko and Armin Mester (2012). “Recursive prosodic phrasing in Japanese”. In: *Prosody Matters.* Ed. by Toni Borowsky et al. London: Equinox Publishing, pp. 280–303.
Krivokapić, J., W. Styler, and Benjamin Parrell (2020). “Pause postures: The relationship between articulation and cognitive processes during pauses.” In: *Journal of Phonetics* 79.
Kubozono, Haruo (1988). “The Organization of Japanese Prosody”. Published by Kurosio Publishers, Tokyo, 1993. Doctoral dissertation. University of Edinburgh.
Ladd, D. Robert (1996). *Intonational Phonology.* Cambridge, UK: Cambridge University Press.
Lee, S., Dani Byrd, and J. Krivokapić (2006). “Functional data analysis of prosodic effects on articulatory timing”. In: *journal of the Acoustical Society of America* 119.3, pp. 1666–1671.
Maekawa, Kikuo (1994). “Is there ‘dephrasing’ of the accentual phrase in Japanese”. In: *Ohio State University Working Papers in Linguistics*, pp. 146–165.
McCawley, James D. (1968). *The Phonological Component of a Grammar of Japanese.* The Hague: Mouton.
Mücke, D., Martine Grice, and Taehong Cho (2014). “More than a magic moment—Paving the way for dynamics of articulation and prosodic structure.” In: *Journal of Phonetics* 44.1-7.
Pierrehumbert, Janet B. and Mary Beckman (1988). *Japanese Tone Structure.* Cambridge: MIT Press.
Poser, William (1984). “The Phonetics and Phonology of Tone and Intonation in Japanese”. Doctoral dissertation. MIT.
Richards, Norvin (2010). *Uttering Trees.* MIT Press.
— (2016). *Contiguity Theory.* MIT Press.
Shaw, Jason and Shigeto Kawahara (2018). “Assessing surface phonological specification through simulation and classification of phonetic trajectories”. In: *Phonology* 35.
Shih, Stephanie S and Vera Gribanova (2016). “Phonological influences in syntactic alternations”. In: *The morphosyntax-phonology connection: Locality and directionality at the interfaces.* Oxford: Oxford University Press, pp. 223–254.
Shih, Stephanie S and Kie Zuraw (2017). “Phonological conditions on variable adjective-noun word order in Tagalog”. In: *Phonological Analysis* 93.4, e317–e352.
Smith, Jennifer (2005). “on the *wh*-question intonational domain in Fukuoka Japanese: Some implications for the syntax-prosody interface”. In: *UMOP 30: Papers on Prosody.* Ed. by Shigeto Kawahara. Amherst: GSLA.
— (2011). “[+wh] complementizers drive phonological phrasing in Fukuoka Japanese”. In: *Natural Language and Linguistic Theory* 29, pp. 545–559.

Sugahara, Mariko (2003). "Downtrends and Post-focus Intonation in Tokyo Japanese". Doctoral Dissertation. University of Massachusetts, Amherst.

Vatikiotis-Bateson, Eric, Adriano Vilela Barbosa, and Catherine T. Best (2014). "Articulatory coordination of two vocal tracts". In: *Journal of Phonetics* 44, pp. 167–181.

Venditti, Jennifer (2005). "The ToBI model of Japanese intonation". In: *Prosodic Typology: The Phonology of Intonation and Phrasing*. Ed. by Sun-Ah Jun. Oxford: Oxford University Press, pp. 172–200.

Venditti, Jennifer, Kikuo Maekawa, and Mary Beckman (2008). "Prominence marking in the Japanese intonational system". In: *The Handbook of Japanese Linguistics*. Ed. by Shigeru Miyagawa and Mamoru Saito. Oxford: Oxford University Press.

Whalen, Douglas H and Andrea G. Levitt (1995). "The universality of intrinsic F0 of vowels". In: *Journal of Phonetics* 23, pp. 349–366.

Zahorian, Stephan A. and Hongbing Hu (2008). "A spectral/temporal method for robust fundamental frequency tracking". In: *journal of the Acoustical Society of America* 123.6, pp. 4559–4571.

Zhang, M., C. Geissler, and Jason Shaw (2019). "Gestural representations of tone in Mandarin: Evidence from timing alternations". In: *Proceedings of the 19th International Congress of Phonetic Sciences*, pp. 1803–1807.