High Vowel Devoicing in Tohoku Japanese is Conditioned by Foot Structure

Hanyoung Byun

University of California, Santa Cruz

1 Introduction

It is controversial whether the position of feet in a word can be influenced by segmental sonority. Kenstowicz (1997) and de Lacy (2002, et seq.) proposed that the sonority of vowels within a word can affect the footing of the word. Their proposal is based on cross-linguistic findings that stress can shift in order to satisfy either of the following universal tendencies about metrical structure: (i) high-sonority vowels are preferred in a foot head position, and (ii) low-sonority vowels are preferred in a non-head position of a foot. In Kenstowicz (1997) and de Lacy (2002, et seq.), these tendencies are captured by a set of markedness constraints that ban vowels of a certain sonority level in a certain position of a prosodic constituent.

However, more recent literature has questioned the adequacy of the empirical data that were used to support the claim that stress assignment can be sensitive to vowel sonority. For example, Shih (2018) and Shih and de Lacy (2019) argue that the empirical evidence for sonority-driven stress discussed in previous literature is either methodologically dubious, or can be accounted for by separate factors such as the moraicity of a vowel. In a similar vein, Rasin (2018, 2020) proposes reanalyses of some purported cases of sonority-driven stress. These previous studies have been mostly concerned with whether stress can shift due to segmental sonority. However, the proposal on sonority-driven stress by Kenstowicz (1997) and de Lacy (2002, et seq.) extends beyond stress assignment, since it is a mechanism for sonority-sensitivity in prosodic structure in general. Then, we predict its effects in other domains of phonology, such as prosodically driven allophony or pitch accent assignment.

In this study, I present evidence for sonority-sensitivity in foot structure that is independent of stress assignment. The empirical evidence comes from high vowel devoicing and loanword accentuation in Tohoku Japanese, an understudied dialect of Japanese. I propose a modified vowel sonority hierarchy, by adding voiceless vowels at the bottom of the hierarchy. Then, based on this hierarchy, I introduce two foot well-formedness constraints that regulate the prosodic structure of Tohoku Japanese: one that disfavors voiceless vowels (low sonority) in foot heads, and another that disfavors non-high vowels (high sonority) in the non-head position of a foot. Crucially, I claim that high vowel devoicing is conditioned by these constraints — high vowel devoicing is only licensed when its application does not create an ill-formed foot. In addition, I propose that loanword accentuation patterns in Tohoku Japanese provide converging evidence for sonority-driven foot structure. Specifically, I show that the default accent in disyllabic and trisyllabic words shifts in order to avoid violations of foot well-formedness constraints.

2 High vowel devoicing in Tohoku Japanese

Tohoku Japanese is a group of Japanese dialects spoken in the northern area of Honshu (the main island of Japan). While high vowel devoicing is a phenomenon attested in many dialects of Japanese, its properties, such as the categoricity of the process and the phonological environment where it applies, differ by dialect (Fujimoto 2015, Byun 2012). Especially, as shown below, the application of high vowel devoicing in Tohoku Japanese is determined by multiple phonological factors, such as (i) the local segmental context, (ii) the height of the vowel in the following syllable, (iii) the position of the target vowel in a word, and (iv) the

*Many thanks to Ryan Bennett, Rachel Walker, Jaye Padgett, Junko Ito, Niko Webster and the members of Phlunch (UCSC Phonetics/Phonology Lunch) for their helpful comments and feedback. I would also like to thank the reviewers and participants of AMP 2024 for helpful discussion. All remaining errors are my own.

Proceedings of *AMP 2023/2024*

weight of the following syllable. The main claim of this study is that these factors can be accounted for by making reference to sonority-sensitive foot structure.

The rest of the section is dedicated to describing the phonological factors that condition high vowel devoicing in Tohoku Japanese. The examples in this section come from Saito (1987, 1993), Uwano (2021, 2022, 2023), and the audio recordings from the Dialect Conversation Data (National Language Research Institute 1981)[1], all of which describe Tohoku Japanese as spoken in Iwate Prefecture. In the examples below, I mark each example with its source(s)[2]. All the words are from the native or Sino-Japanese vocabulary (see Ito and Mester 1995 for the relevance of lexical strata in Japanese phonology). For the sake of simplicity, I omit the information on pitch accent in the examples in this section.

First, high vowel devoicing applies when a phonemically short high vowel /i, ɯ/ is surrounded by voiceless consonants, as illustrated in (1).

(1) Application of high vowel devoicing between voiceless consonants
a. [kɯ̥sa] 'grass' (S93) b. [ɸɯ̥ta] 'lid' (S93)
c. [sɯ̥so] 'shiso (herb)' (U21) d. [ikɯ̥sa] 'battle' (S87)

However, devoicing is blocked *when the syllable following the target vowel contains another high vowel*. This is shown in (2). The first vowel in each word is otherwise eligible for devoicing but fails to actually devoice. Notice that [ɸɯtsɯ̥ka] '2nd day' in (2d) exemplifies a case where the first two vowels are high, and the final vowel is non-high. While each of the first two vowels is surrounded by voiceless consonants, only the second vowel, which is followed by a non-high vowel, undergoes devoicing.

(2) Blocking of high vowel devoicing by the following high vowel
a. [kɯsɯ] 'comb' (S93) b. [ɸɯsɯ] 'joint' (S93, DCD)
c. [sisi] 'deer' (U21) d. [ɸɯtsɯ̥ka] '2nd day' (S93, U21)

While the observation that high vowel devoicing in Tohoku Japanese is conditioned by the height of the following vowel has been pointed out by several previous studies (Inoue 1968, Saito 1993, Ohashi 1995, Matsui et al. 2023), there are exceptional phonological environments whereby devoicing applies even though the following syllable contains a high vowel. One such environment is where the target high vowel is in the second syllable of a tetrasyllabic word with light syllables. In the examples in (3), the second vowel of each word is followed by a high vowel. Nevertheless, the second vowel has undergone devoicing. This suggests that the position of the target vowel in a word and the number of syllables are relevant factors for the application of high vowel devoicing.

(3) Application of high vowel devoicing in even-parity words
a. [hotsi̥kɯɾɯ] 'to dig up' (U21) b. [kaɕi̥-kiɾi] 'rent' (U22)

Another exceptional phonological environment where devoicing applies is when the following high vowel is phonemically long, as shown in (4). This suggests that the weight of the following syllable is also a relevant factor for high vowel devoicing.

(4) Application of high vowel devoicing before a long high vowel
a. [ɸɯ̥tsɯː] 'usually' (DCD) b. [ɸɯ̥kjɯː] 'dissemination' (DCD)

While these conditioning factors are rather complex, I show that the blocking environments can be unified by making reference to the foot structure of Tohoku Japanese. To preview the analysis, I show that the blocking effect of the height of the following vowel is not due to vowel height *per se*, but rather due to the sonority of vowels and the sonority requirements for vowels in certain positions in a foot. I argue that high vowel devoicing is prohibited when its application creates an ill-formed foot.

[1] For the Dialect Conversation Data, I transcribed the relevant examples from the audio recordings.
[2] S87 = Saito (1987), S93 = Saito (1993), U21 = Uwano (2021), U22 = Uwano (2022), U23 = Uwano (2023), DCD = National Language Research Institute (1981)

3 Analysis

In this section, I provide an analysis of high vowel devoicing in Tohoku Japanese. Adopting the proposal by Kenstowicz (1997) and de Lacy (2002, et seq.) that footing can be affected by vowel sonority, I employ a foot well-formedness constraint that bans low-sonority vowels from foot heads, and another constraint that bans high-sonority vowels in the weak branch of a foot. I propose that Tohoku Japanese employs two strategies to avoid violations of these constraints in the devoicing context: (i) blocking high vowel devoicing, and (ii) leaving the target vowel unfooted.

3.1 *Vowel sonority hierarchy and foot well-formedness constraints* Kenstowicz (1997) and de Lacy (2002, et seq.) propose a vowel sonority hierarchy that conditions the well-formedness of feet. The hierarchy is based on the cross-linguistic tendency that the location of a foot is influenced by the type of vowels in a word (but see Shih 2018 and Shih and de Lacy 2019 for arguments about the empirical reliability of such a tendency). Based on the vowel sonority hierarchy in (5), de Lacy (2002) provides the following two generalizations. One is that the more sonorous the vowel is, the more preferred it is as a head of a prosodic constituent. The other is that the less sonorous the vowel is, the more preferred it is as a non-head of a prosodic constituent. So for example, for trochees, *(ba.bi)* will be preferred over *(bi.ba)*[3], all else being equal. Notice that the vowel sonority hierarchy in (5) roughly coincides with vowel height, except for central vowels being less sonorous than peripheral vowels.

(5) Vowel sonority hierarchy (Kenstowicz 1997; de Lacy 2002, et seq.)
Low peripheral vowels (e.g. [a]) > Mid peripheral vowels (e.g. [e, o]) >
High peripheral vowels (e.g. [i, u]) > Mid central vowels (e.g. [ə]) > High central vowels (e.g. [ɨ])
("A > B" means "A is more sonorous than B")

In de Lacy (2002, et seq.), generalizations about sonority requirements for foot heads and non-heads are formalized as markedness constraints for foot head well-formedness and non-head well-formedness, as shown in (6) and (7), respectively. For example, *HEAD$_{FT}$/*ɨ•ə•i•u* bans a foot head that contains a vowel whose sonority is lower than mid peripheral vowels, and *NONHEAD$_{FT}$/*a•e•o•i•u* bans a non-head of a foot that contains any vowel whose sonority is higher than central vowels. These foot well-formedness constraints form a stringency relation; according to de Lacy (2002, et seq.), this allows vowel sonority categories to be collapsed, or 'conflated'.

(6) Foot head well-formedness constraint (de Lacy 2002, et seq.)
*HEAD$_{FT}$/X =$_{def}$ Assign a violation for every foot head that contains X,
where X = {{ɨ}, {ɨ, ə}, {ɨ, ə, i, u}, {ɨ, ə, i, u, e, o}, {ɨ, ə, i, u, e, o, a}}.

(7) Non-head well-formedness constraint (de Lacy 2002, et seq.)
*NONHEAD$_{FT}$/X =$_{def}$ Assign a violation for every non-head of a foot that contains X,
where X = {{a}, {a, e, o}, {a, e, o, i, u}, {a, e, o, i, u, ə}, {a, e, o, i, u, ə, ɨ}}.

In order to account for the Tohoku data, I make one modification to the vowel sonority hierarchy in (5): adding *voiceless vowels* at the bottom of the vowel sonority hierarchy, which makes them the least sonorous vowels.[4] This is illustrated in (8), where I collapse some vowel categories irrelevant for Tohoku Japanese. While this modification of the hierarchy is motivated by the patterns of high vowel devoicing in Tohoku Japanese, it can be justified from a phonetic perspective. Given that sonority (of vowels and consonants) is correlated with acoustic intensity (Parker 2002), voiceless vowels are predicted to have lower sonority than voiced vowels since voiceless (or whispered) vowels are generally less loud than voiced vowels (Jesus et al. 2023, Heeren 2015).

[3] From now on, parantheses are used to group feet and an underline is used to express the head of a foot.
[4] It is possible that there is a sub-hierarchy for the sonority of voiceless vowels (e.g. voiceless high vowels vs. voiceless non-high vowels). But at least in Tohoku Japanese, there is no evidence for such a distinction since only high vowels undergo devoicing.

(8) Modified vowel sonority hierarchy (categories irrelevant for Tohoku Japanese are collapsed.)
Voiced non-high vowels (e.g., [a, e, o]) > Voiced high vowels (e.g., [i, ɯ]) >
Voiceless vowels (e.g., [i̥, ɯ̥])

The sonority hierarchy in (8) predicts that voiceless vowels are least preferred in the head position of a foot, and most preferred in the non-head position of a foot. Conversely, voiced non-high vowels are predicted to be least preferred in a non-head of a foot, and most preferred in a head of a foot. Based on this, I introduce two foot well-formedness constraints in (9), which play an important role in the analysis below. The first is *HEAD$_{FT}$/V̥, which bans a voiceless vowel (i.e. low-sonority vowel) in a foot head, and the second is *NONHEAD$_{FT}$/V$_{[-high]}$, which bans a non-high vowel (i.e. high-sonority vowel) in a non-head of a foot.

(9) Foot well-formedness constraints
a. *HEAD$_{FT}$/V̥: Assign a violation for every foot head that contains a voiceless vowel.
b. *NONHEAD$_{FT}$/V$_{[-high]}$: Assign a violation for every non-head of a foot that contains a voiced non-high vowel.

3.2 *Analysis of high vowel devoicing* Based on the discussion so far, I present an OT analysis of high vowel devoicing in Tohoku Japanese. I mostly focus on the (non-)application of devoicing in disyllabic words (e.g. *kɯ̥sa* 'grass' vs. *kusɯ* 'comb'), and the exceptional application of devoicing in tetrasyllabic words (e.g. *hotsɨ̥kuɾɯ* 'dig up') and words with a heavy syllable (e.g. *ɸɯ̥tsɯː* 'usually').

I first discuss the set of constraints used to derive the basic high vowel devoicing pattern between voiceless consonants. The constraints and their definition are shown in (10). In the current analysis, I employ HVD (10a) as a context-sensitive markedness constraint that bans a voiced short high vowel between voiceless consonants. This constraint will act as the main driving force of high vowel devoicing. The context-free markedness constraint *V̥ (10b) bans voiceless vowels, which are reported to be rare cross-linguistically (Blevins 2018). Finally, the faithfulness constraint ID(voi) (10c) is in conflict with these markedness constraints.

(10) Constraints for high vowel devoicing
a. HVD: Assign a violation for every voiced short high vowel that is surrounded by voiceless consonants.
b. *V̥: Assign a violation for every voiceless vowel.
c. ID(voi): Assign a violation if correspondent input and output segments do not have identical values for the feature [voice].

As shown in Tableau (11), the ranking of HVD ≫ *V̥ ≫ ID(voi) derives devoicing only when a high vowel is flanked by voiceless consonants. If a hypothetical input contains a voiceless vowel that is not flanked by voiceless consonants, as in /kɯ̥ma/ (11ii), an output with a voiced vowel (11iia) is chosen as the optimal candidate since it satisfies the context-free markedness constraint *V̥.

(11) Basic high vowel devoicing

(i) /kɯsa/ 'grass'	HVD	*V̥	ID(voi)	(ii) /kɯ̥ma/	HVD	*V̥	ID(voi)
☞a. kɯ̥sa		*	*	☞a. kɯma			*
b. kɯsa	*! W	L	L	b. kɯ̥ma		*! W	L

Having set the basic analysis of high vowel devoicing, I now turn to a case where devoicing is blocked, e.g. *kusɯ* 'comb'. The main proposal is that all Tohoku words involve foot structure, and the blocking of devoicing is attributed to the effect of a high-ranked foot well-formedness constraint, *HEAD$_{FT}$/V̥. One important assumption that I make here is that the default foot form in Tohoku Japanese is trochaic (which is independently supported by the default penultimate accent in loanwords; Section 4). This assumption is implemented by the constraint ranking of TROCHEE (= Every foot is left-headed) over IAMB (= Every foot is right-headed). For example, a disyllabic word such as *kusɯ* will be parsed as *(kusɯ)*, where the target vowel

for devoicing is in a foot head position. Then, if *HEAD$_{FT}$/V̥ outranks HVD, devoicing the target vowel (i.e. *(kɯ̥sɯ)*) will result in a fatal violation of *HEAD$_{FT}$/V̥. The tableau for this analysis is given in (12).[5]

(12) Blocking of devoicing

/kɯsɯ/ 'comb'	*HD$_{FT}$/ V̥	TROCHEE	FTBIN	IAMB	HVD	ID(voi)
☞a. (kɯsɯ)				*	*	
b. (kɯ̥sɯ)	*! W			*	L	* W
c. (kɯ̥sɯ)		*! W		L	L	* W
d. kɯ̥(sɯ)			*! W	L	L	* W
e. (kɯ)sɯ			*! W	L	*	

In (12), the non-devoiced output *(kɯsɯ)* (12a) only violates IAMB and HVD. On the other hand, the losing candidate *(kɯ̥sɯ)* (12b) satisfies HVD but crucially violates higher-ranked *HEAD$_{FT}$/V̥, due to the voiceless vowel in the foot head. Changing the foot form to an iamb (12c) is also ruled out, due to high-ranked TROCHEE. Finally, output candidates where the target vowel is unfooted (= (12d, e)) need to be considered as well. If the target vowel is unfooted, it is no longer subject to foot well-formedness constraints (assuming that degenerate feet satisfy TROCHEE and IAMB) and it will be forced to undergo devoicing (i.e. (12d)) by the ranking of HVD over ID(voi). To rule this out, I introduce FTBIN (= "Feet are binary under moraic or syllabic analysis"; Prince and Smolensky 1993/2004), which is ranked higher than IAMB and HVD. Similarly, an output candidate where the initial syllable forms a degenerate foot (12e) is also ruled out. The optimal output (12a) satisfies the default trochaic form and the foot well-formedness constraint. Put differently, we can say that blocking of devoicing is used as a strategy to avoid a violation of *HEAD$_{FT}$/V̥.

Now, let us revisit the application of devoicing in examples like *kɯ̥sa* 'grass', where the height of the following vowel is non-high. Now that we have high-ranked *HEAD$_{FT}$/V̥, devoicing in *kɯ̥sa* cannot be derived by a trochaic structure like *(kɯsa)*. Furthermore, a trochaic parse such as *(kɯ̥sa)* is ill-formed not only due to the voiceless vowel in the foot head, but also due to the non-high vowel (high-sonority) in non-head position. In other words, the trochaic parse incurs a violation of another foot well-formedness constraint, *NONHEAD$_{FT}$/V$_{[–high]}$, which bans a non-high vowel in the non-head position of a foot.

Given this, there are two options to avoid a violation of *HEAD$_{FT}$/V̥ and *NONHEAD$_{FT}$/V$_{[–high]}$. The first is to change the foot form to an iamb (i.e. *(kɯ̥sa)*), which will incur a violation of TROCHEE. The target vowel is expected to undergo devoicing since it is no longer in a foot head position and is not subject to *HEAD$_{FT}$/V̥. If this is the case, then the ranking should be *HEAD$_{FT}$/V̥, *NONHEAD$_{FT}$/V$_{[–high]}$ ≫ TROCHEE, FTBIN.

The second is to leave the target vowel unfooted and create a degenerate foot (i.e. *kɯ̥(sa)*), which will incur a violation of FTBIN. In this scenario, too, the target vowel is not in a foot head position and hence expected to undergo devoicing without violating *HEAD$_{FT}$/V̥. If this is the case, then the ranking should be *HEAD$_{FT}$/V̥, *NONHEAD$_{FT}$/V$_{[–high]}$, TROCHEE ≫ FTBIN.

At this point, both options can derive the pattern. However, evidence from loanword accentuation patterns suggests that what happens in Tohoku Japanese is the second option. As will be shown in more detail in Section 4, the default accent position in disyllabic loans in Tohoku Japanese is the penultimate mora (e.g. *gásɯ* 'gas'). This pattern is maintained even if the penultimate vowel is high and the final vowel is non-high (e.g. *kíro* 'kilo'). If we assume that the default penultimate accent is an effect of TROCHEE, the ranking of *NONHEAD$_{FT}$/V$_{[–high]}$ over TROCHEE will incorrectly predict iambic **kiró*. For this reason, I propose that the right analysis for Tohoku Japanese is *HEAD$_{FT}$/V̥, *NONHEAD$_{FT}$/V$_{[–high]}$, TROCHEE ≫ FTBIN, under which devoicing can apply via creating a degenerate foot (e.g. *kɯ̥(sa)*).

Tableau (13) illustrates this point with the example *kɯ̥sa* 'grass'. The optimal output (13a) undergoes devoicing by making the target vowel unfooted. The unfooted vowel can undergo devoicing because it is exempt from violations of the foot well-formedness constraints. The outputs with a trochee (13b, c) are ruled out by high-ranked *NONHEAD$_{FT}$/V$_{[–high]}$, and the iambic output (13d) is ruled out by TROCHEE. The non-devoiced output with a left-aligned degenerate foot (13e) violates FTBIN just as the optimal output (13a), but it loses due to the violation of HVD.

[5] Constraints that are not crucial for the analysis will be omitted in the tableaux hereafter.

(13) Application of devoicing via creating a degenerate foot

/kɯsa/ ‘grass’	*HDFT/ V̥	*NONHDFT/V[–hi]	TROCHEE	FTBIN	HVD	ID(voi)
☞a. kɯ̥(sa)				*		*
b. (kɯsa)		*! W		L	* W	L
c. (kɯ̥sa)	*! W	*! W		L		*
d. (kɯ̥sa)			*! W	L		*
e. (kɯ)sa				*	*! W	L

To summarize, the current analysis suggests that the vowel height difference conditioning high vowel devoicing is only apparent, and it is actually vowel *sonority* and the sonority-sensitive foot structure that condition devoicing. Under this analysis, high vowel devoicing is blocked when the target vowel is forced to be in a foot head. Conversely, it is allowed when the target vowel can escape from a foot head position.

The current foot-based account further predicts the exceptional application of devoicing in tetrasyllabic words with light syllables and words with a heavy syllable. Tetrasyllabic words with light syllables such as *hotsi̥kɯɾɯ* ‘dig up’ (3a) will be optimally parsed as (σσ)(σσ) under the current ranking, if a constraint like PARSE-σ, which demands all syllables to be parsed, is ranked high enough. If the target vowel is in the second syllable, it will be in a non-head position of the foot and it can undergo devoicing without violating foot well-formedness constraints. This is illustrated in Tableau (14), where I only included relevant constraints. In the winning candidate (14a), all syllables are parsed as trochees and the second vowel is in a non-head position. Creating degenerate feet (14b) and leaving a syllable unfooted (14c) will incur fatal violations of FTBIN.

(14) Application of devoicing in even-parity words with light syllables

/hotsikɯɾɯ/ ‘dig up’	FTBIN	PARSE-σ	HVD	ID(voi)
☞a. (hotsi̥)(kɯɾɯ)				*
b. (ho)(tsi)(kɯɾɯ)	*!* W		* W	L
c. (ho)tsi̥(kɯɾɯ)	*! W	* W		*

In words where the target high vowel is followed by a heavy syllable, such as *ɸɯ̥tsɯː* ‘usually’ (4a), the heavy syllable can form a bimoraic trochee, without violating TROCHEE or FTBIN. In Tableau (15), the winning candidate has a bimoraic trochee with its initial syllable unfooted. The initial vowel can undergo devoicing since it is unfooted. On the other hand, creating an iamb (15b) incurs a fatal violation of TROCHEE, and creating a degenerate foot (15c) fatally violates FTBIN. PARSE-σ should be ranked lower than FTBIN so that unfooted syllables are allowed (also in (13)).

(15) Application of devoicing in a word with a heavy syllable

/ɸɯtsɯː/ ‘usually’	TROCHEE	FTBIN	PARSE-σ	HVD	ID(voi)
☞a. ɸɯ̥(tsɯː)			*		*
b. (ɸɯ̥tsɯː)	*! W		L		*
c. (ɸɯ)(tsɯː)		*! W	L	* W	L

In summary, the foot-based account of high vowel devoicing successfully unifies its conditioning factors, such as the height of the following vowel, the position of the target vowel in a word, and the weight of the following syllable. In addition, the current account makes further predictions that go beyond the scope of the current dataset. For example, if there is a tetrasyllabic word whose first and fourth syllables are heavy (i.e. HLLH, where H = heavy syllable and L = light syllable) and the second vowel is subject to devoicing, the current analysis predicts that this vowel will not undergo devoicing because the word will be parsed as (H)(LL)(H). I have not yet found an example to test this prediction. Also, a fuller analysis should account for devoicing in words with an odd number of syllables, which is not addressed in this study. I leave a fuller analysis of Tohoku high vowel devoicing for future study.

4 Accent shift in loanwords

The existence of *HEAD$_{FT}$/V̥ and *NONHEAD$_{FT}$/V$_{[-high]}$ in the grammar of Tohoku Japanese is independently evidenced by accent shift in loanwords. In Japanese dialects, the distribution of pitch accent is more or less predictable in loanwords, whereas the location of pitch accent in native vocabulary is mostly lexically specified and hence largely unpredictable (Kawahara 2015, Kubozono 2008 among many others). Given this, loanword accentuation patterns can provide evidence for the basic foot structure (i.e. its form and location) in Tohoku Japanese, if we assume a tight correspondence between accent assignment and foot structure (e.g. by employing WORD PROMINENCE TO WORD HEAD by Ito and Mester 2016). Specifically, I show that accent shift in disyllabic and trisyllabic loans provides further evidence for the dispreference for voiceless vowels in a foot head position, and for non-high vowels in a non-head position.

All the examples of loanwords in this section come from Uwano (2020). While there seems to be some variation in accentuation patterns among Tohoku Japanese dialects and even within a specific regional variety of Tohoku Japanese, the description by Uwano (2020) lays out some general tendencies in loanword accentuation. In disyllabic loans, the most common accentuation pattern is penultimate accent (I only focus on words with light syllables, so the syllable count and the mora count are identical.). This is shown in (16). Notice that the penultimate accent is maintained regardless of the height of the vowels within a word.

(16) Penultimate accent in disyllabic loans
a. [gásɯ] ‘gas’ b. [zéɾo] ‘zero’ c. [bíɾɯ] ‘building’ d. [kíɾo] ‘kilo’

However, as shown in (17), the default penultimate accent is shifted due to an interaction with high vowel devoicing. Specifically, the penultimate accent shifts to the final syllable when the penultimate vowel undergoes devoicing. Note that devoicing is licensed in (17) since the final vowels are non-high.

(17) Accent shift to the final syllable when the penultimate vowel is devoiced
a. [sɯ̥tó] ‘strike’ b. [pi̥ké] ‘picket’

It is worth noting that this type of accent shift due to vowel devoicing is also reported for Tokyo Japanese (Haraguchi 1977, Yokotani 1997, but also see Sugito 1982, Kitahara 1998). The current analysis predicts this accent shift: it shows that pitch accent, which falls on a foot head, avoids voiceless vowels, which have the lowest sonority.

Let us now move on to the accentuation pattern in trisyllabic loans. Uwano (2020) states that some old (or frequently used) loanwords are unaccented, as illustrated in (18).[6]

(18) Unaccented trisyllabic loans
a. [kabotʃa] ‘pumpkin’ b. [boɾɯto] ‘bolt’ c. [ɾamɯne] ‘lemonade’

Just as disyllabic loans, the default accentuation pattern for trisyllabic loans is penultimate accent. It is worth noting that this pattern is different from the default accentuation pattern in Tokyo Japanese, where accent falls on the antepenultimate mora (e.g. *bánana* in Tokyo Japanese; Ito and Mester 2016, Kubozono 2008, among others).

(19) Penultimate accent in trisyllabic loans
a. [banána] ‘banana’ b. [gaɾásɯ] ‘glass’ c. [teɾébi] ‘TV’ d. [aɾúmi] ‘aluminum’

However, the accent shifts to the antepenultimate syllable *when the penultimate syllable contains a high vowel and the final syllable contains a non-high vowel*. This is illustrated in (20). Notice that this environment bears a similarity to one of the contexts where high vowel devoicing is licensed, namely, a syllable with a high vowel followed by a syllable with a non-high vowel (i.e. CV$_{[+high]}$CV$_{[-high]}$).

[6] Deaccentuation of frequently used loanwords is also attested in Tokyo Japanese; see Kawahara (2015: 459) and references therein.

(20) Accent shift to the antepenultimate syllable
a. [káɾɯta] ‘card’ b. [góɾiɾa] ‘gorilla’ c. [míkɯɾo] ‘micro’ d. [bíɾɯma] ‘Burma’

As also pointed out by Uwano (2020), it is important to note that the accent shift in (20) is not solely due to the high vowel in the penultimate syllable. This is because *arúmi* 'aluminum' (19d), which contains high vowels in both penultimate and final syllables, does not exhibit the accent shift. Also, it is not the case that the accent shift occurs in order to avoid an accent on an epenthetic vowel (not present in the source language), as suggested by Ito and Mester (2016), Kubozono (2011, 2015), among others, for loanwords in Tokyo Japanese. This is because accent shift occurs in *góɾiɾa* ‘gorilla’ in (20b), even though [i] in *góɾiɾa* is not epenthetic.

Let us now discuss what these data tell us about foot structure in Tohoku Japanese. An important assumption here is that the pitch accent of a word coincides with its prosodic head (Ito and Mester 2016). Then, the fact that accent is assigned on the penultimate mora by default (excluding some old loans), as in (17) and (20), suggests that the default foot form is trochaic, and the directionality of footing is right to left. For example, [gásɯ] ‘gas’ (16a) should be parsed into trochaic *(gaṣu)*, and not iambic *(gaṣu)*. For trisyllabic loans, such as [banána] ‘banana’ (19a), the last two syllables should be footed together and form a trochee, i.e. *ba(nána)*. This can be accounted for by the ranking of TROCHEE ≫ IAMB and ALLFTR ≫ ALLFTL. This is illustrated in (21).

(21) Analysis for the default penultimate accent in trisyllabic loans

/banana/ ‘banana’	TROCHEE	ALLFTR	IAMB	ALLFTL
☞a. ba(nána)			*	*
b. (bána)na		*! W	*	L
c. ba(naná)	*! W		L	*
d. (baná)na	*! W	*! W	L	L

However, as briefly discussed in Section 3, the penultimate accent in *kíɾo* ‘kilo’ (16d) cannot be attributed to a trochaic foot (e.g. *(kiɾo)*). This trochaic output is ruled out by high-ranked *NONHEAD$_{FT}$/V$_{[-high]}$, since the final vowel is non-high. Furthermore, the ranking in (21) incorrectly predicts **ki(ɾó)*, with a right-aligned degenerate foot. This result is due to the ranking of ALLFTR, as shown in Tableau (22).

(22) The current ranking incorrectly predicts **ki(ɾó)*

/kiɾo/ ‘kilo’	*NONHD$_{FT}$/V$_{[-hi]}$	TROCHEE	FTBIN	ALLFTR	IAMB	ALLFTL
☞a. ki(ɾó)			*			**
☹b. (kí)ɾo			*	*!* W		L
c. (kíɾo)	*! W		L		* W	L
d. (kiɾó)		*! W	L			L

One solution to this is to rank NONFIN (= “No prosodic head is final in PrWd”; Prince and Smolensky 1993/2004) over ALLFTR. Words that end in a foot will violate NONFIN. This ranking is motivated by the fact that a final accent in Tohoku Japanese is generally avoided, except for cases of accent shift. This ranking will prefer *(kí)ɾo* over **ki(ɾó)*, as shown in Tableau (23).

(23) NONFIN outranks ALLFTR

/kiɾo/ ‘kilo’	*NONHD$_{FT}$/V$_{[-hi]}$	TROCHEE	FTBIN	NONFIN	ALLFTR	IAMB	ALLFTL
☞a. (kí)ɾo			*		**		
b. ki(ɾó)			*	*! W	L		** W
c. (kíɾo)	*! W		L	* W	L	* W	
d. (kiɾó)		*! W	L	* W	L		

Let us now turn to the analysis of the accent shift. The accent shift in disyllabic words in (17) can be accounted for by the same analysis for *kɯ̥sa* 'grass' in (13), whose first vowel is subject to devoicing and whose final vowel is non-high. In the current analysis, the driving force behind the accent shift is *HEAD$_{FT}$/V̥ and *NONHEAD$_{FT}$/V$_{[-high]}$. For the input /sɯto/ 'strike' (17a), the output candidate *(sɯ̥́to)*, whose foot head contains a voiceless vowel and non-head contains a non-high vowel, violates high-ranked *HEAD$_{FT}$/V̥ and *NONHEAD$_{FT}$/V$_{[-high]}$. Even if the target vowel does not undergo devoicing, as in *(sɯ̥́to)*, the output still loses since it violates *NONHEAD$_{FT}$/V$_{[-high]}$. As with the analysis for *kɯ̥sa*, I claim that the penultimate vowel can undergo devoicing because it is unfooted, i.e. *sɯ̥(tó)*. We can see that this output satisfies both of the foot well-formedness constraints.

Compare this result with the one for /kiɾo/ 'kilo' in (23), where the optimal candidate is the one with a left-aligned degenerate foot, i.e. *(kí)ɾo*. To derive these results, a right-aligned degenerate foot should be *only allowed in the devoicing environment*. This can be implemented by ranking HVD over NONFIN, as illustrated in Tableau (24). The ranking HVD ≫ NONFIN ≫ ALLFTR will produce final degenerate feet only in the devoicing environment.

(24) HVD outranks NONFIN

/sɯto/ 'strike'	*NONHD$_{FT}$/V$_{[-hi]}$	TROCHEE	FTBIN	HVD	NONFIN	ALLFTR	IAMB	ALLFTL
☞a. sɯ̥(tó)			*		*			*
b. (sɯ́to)	*! W		L	*	L			L
c. (sɯ̥tó)		*! W	L		L		* W	L
d. (sɯ́)to			*	*! W	L	* W		L

The accent shift in trisyllabic loans (20) is also straightforwardly explained by the effect of *NONHEAD$_{FT}$/V$_{[-high]}$. As shown in (25), when the penultimate vowel is high and the final vowel is non-high, as in /goriɾa/, the default footing will yield **go(ɾíɾa)* (25b), where the non-head position of the foot contains a non-high vowel [a]. This output violates the high-ranked *NONHEAD$_{FT}$/V$_{[-high]}$. On the other hand, when the foot shifts leftward, as in *(góɾi)ɾa* (25a), the output satisfies *NONHEAD$_{FT}$/V$_{[-high]}$ at the expense of violating ALLFTR. The ranking *NONHEAD$_{FT}$/V$_{[-high]}$ ≫ ALLFTR chooses *(góɾi)ɾa* as the winning candidate.

(25) Analysis for the accent shift in trisyllabic loans

/goriɾa/ 'gorilla'	*NONHD$_{FT}$/V$_{[-hi]}$	TROCHEE	ALLFTR	IAMB	ALLFTL
☞a. (góɾi)ɾa			*!	*	
b. go(ɾíɾa)	*! W		L	*	* W

To summarize, I showed that the accent shift in Tohoku loanwords occurs in order to avoid violations of the high-ranked foot well-formedness constraints, *HEAD$_{FT}$/V̥ and *NONHEAD$_{FT}$/V$_{[-high]}$. The constraint ranking for the accent shift is consistent with the analysis of high vowel devoicing in Section 3. Then, we can conclude that the loanword accentuation patterns lend independent support for sonority-driven footing in Tohoku Japanese.

5 Conclusion

In this study, I argued that high vowel devoicing and loanword accentuation in Tohoku Japanese constitute evidence for sonority-driven foot structure in natural language. Based on a modified vowel sonority hierarchy where voiceless vowels are the least sonorous, I introduced two foot well-formedness constraints, HEAD$_{FT}$/V̥ and *NONHEAD$_{FT}$/V$_{[-high]}$, which ban voiceless vowels (low sonority) in a foot head, and non-high vowels (high sonority) in a non-head position, respectively. By employing these constraints, blocking of high vowel devoicing is understood as a strategy to avoid violations of the foot well-formedness constraints when a target vowel for devoicing is in a foot head. Conversely, the application of high vowel devoicing is only allowed when the target vowel is not in a foot head (e.g. in a non-head position of a foot, or in an unfooted

position). In addition, I showed that the current analysis gains further support from loanword accentuation patterns. Specifically, I showed that accent shift from the default penultimate position is attributed to the effect of *HEAD$_{FT}$/V̥ and *NONHEAD$_{FT}$/V$_{[-high]}$. While there has been a controversy in the literature over the existence of sonority-driven stress in natural language (Shih 2018, Shih and de Lacy 2019, Rasin 2018, 2020), this study provides novel evidence for sonority-sensitivity in foot structure, independent of stress assignment.

References

Blevins, Juliette (2018). Evolutionary Phonology and the life cycle of voiceless sonorants. In Sonia Cristofaro and Zúñiga (eds.) *Typological hierarchies in synchrony and diachrony*. Amsterdam/Philadelphia: John Benjamins. 29–60.

Byun, Hi-Gyung (2012). High vowel devoicing in Japanese: As an indicator of standardization of dialect [written in Japanese]. PhD dissertation. University of Tokyo.

de Lacy, Paul (2002). *The formal expression of markedness*. PhD dissertation. UMass Amherst.

Fujimoto, Masako (2015). Vowel devoicing. In Haruo Kubozono (eds.) *Handbook of Japanese Phonetics and Phonology*. 167–214. Berlin, München, Boston: De Gruyter Mouton.

Heeren, Willemijn F. L. (2015). Vocalic correlates of pitch in whispered versus normal speech. *Journal of the Acoustical Society of America 138*. 3800–3810.

Inoue, Humio (1968). Consonant system of the Tohoku dialect [written in Japanese]. *Journal of the Linguistic Society of Japan 52.* 80–98.

Ito, Junko, and Armin Mester (2016). Unaccentedness in Japanese. *Linguistic Inquiry 47(3).* 471–526.

Ito, Junko, and Armin Mester (1995). Japanese phonology. In John Goldsmith (ed.) *The Handbook of Phonological Theory*. Blackwell. 817–838.

Jesus, Luis M.T., Sara Castilho, Aníbal Ferreira, Maria Conceição Costa (2023). Discriminative segmental cues to vowel height and consonantal place and voicing in whispered speech. *Journal of Phonetics 97*: 101223.

Kager, René (1999). *Optimality Theory*. Cambridge University Press.

Kawahara, Shigeto (2015). The phonology of Japanese accent. In Haruo Kubozono (ed.) *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology*. Mouton. 445–492.

Kenstowicz, Michael (1997). Quality-sensitive stress. *Rivista di Linguistica 9*. 157–188.

Kitahara, Mafuyu (1999). The interaction of pitch accent and vowel devoicing in Tokyo Japanese. In David James Silva (ed.) *Japanese/Korean Linguistics 8*. Stanford, CA: CSLI Publications. 303–315.

Kubozono, Haruo (2001). Epenthetic Vowels and Accent in Japanese: Facts and Paradoxes. In Jeroen Weijer and Tetsuo Nishihara (eds.) *Issues in Japanese Phonology and Morphology*. Berlin, Boston: De Gruyter Mouton. 111–140.

Kubozono, Haruo (2008). Japanese accent. In Shigeru Miyagawa (ed.) *The Oxford Handbook of Japanese Linguistics*. Oxford University Press. 165–191.

Kubozono, Haruo (2015). Loanword phonology. In Haruo Kubozono (ed.) *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology*. Mouton. 313–361.

Matsui, Sanae, Ai Mizoguchi, Ayako Hashimoto, Chuyu Huang, Naoya Watabe, Hiroto Noguchi, and Mafuyu Kitahara (2023). An interaction of mutually bleeding processes: voicing of intervocalic plosives and vowel devoicing in Tohoku Japanese. *Proceedings of Japanese/Korean Linguistics 30*. 387–395.

National Language Research Institute (1981). *Texts of tape-recorded conversations in Japanese dialects (vol. 5)–Iwate, Miyagi, Chiba, Shizuoka* [written in Japanese]. The National Language Research Institute.

Ohashi, Junichi (1995). *A study of the sounds in Tohoku dialects* [written in Japanese]. Tokyo: Oufu.

Parker, Stephen G. (2002). *Quantifying the sonority hierarchy*. PhD dissertation. UMass Amherst.

Prince, Alan and Paul Smolensky (1993/2004). *Optimality Theory: constraint interaction in generative grammar*. Malden, MA & Oxford, UK: Blackwell. [Revision of 1993 technical report, Rutger University Center of Cognitive Science. Available on Rutgers Optimality Archive, ROA-537].

Rasin, Ezer (2018). *Modular interaction in phonology*. PhD dissertation. MIT.

Rasin, Ezer (2020). Evidence against sonority-driven stress in Nanti. In Maggie Baird and Jonathan Pesetsky (eds.), *Proceedings of NELS 49*, vol. 3. 59–72

Saito, Koji (1987). Phonological interpretation of word-medial consonant voicing [written in Japanese]. *Gobun Ronsō 15*. 64–86.

Saito, Koji (1993). Phonemes of the dialect of Ryōri, Sanriku-chō, Iwate Prefecture [written in Japanese]. *Tohoku Studies in Linguistics 3*. 37–48.

Shih, Shu-hao (2018). On the existence of sonority-driven stress in Gujarati. *Phonology 35*. 327–364.

Shih, Shu-hao, and Paul de Lacy (2019). Evidence for sonority-driven stress. *Catalan Journal of Linguistics 18*. 9–40.

Sugito, Miyoko (1982). *Studies on Japanese accent* [written in Japanese]. Tokyo: Sanseido.

Uwano, Zendo (2020). Accent Data of Foreign Words in the Northern Tohoku Dialects [written in Japanese]. *NINJAL Research Papers 19.* 47–88.

Uwano, Zendo (2021). Accent Data from the Tanohata Dialect, Iwate Prefecture: The Proto-Accent System of Northern Tohoku Dialects [written in Japanese]. *NINJAL Research Papers 20*. 115–147.
Uwano, Zendo (2022). Accent Data of Adjectives in the Tanohata Dialect, Iwate Prefecture: Part 2 [written in Japanese]. *NINJAL Research Papers 23*. 1–28.
Uwano, Zendo (2023). Accent Data of Verbs from the Tanohata Dialect, Iwate Prefecture: Part 2 [written in Japanese]. *Kokugo Kenkyū 84*. 1–29.
Yokotani, Teruo (1997). Accent shift beyond the foot boundary: evidence from Tokyo Japanese compound nouns [written in Japanese]. *Journal of the Phonetic Society of Japan 1*. 54–62.