RESEARCH

The syllable as a prosodic unit in Japanese lexical strata: Evidence from text-setting

Rebecca Lurie Starr[1] and Stephanie S Shih[2]

[1] AS5, Faculty of Arts and Social Sciences, National University of Singapore, SG

[2] University of California, Merced, 5200 N Lake Road, Merced, CA, US

Corresponding author: Rebecca Lurie Starr (rstarr@nus.edu.sg)

Text-setting, the arrangement of language to music, is a common source of evidence in the debate over the relevance of the syllable in Japanese prosody (e.g., Labrune 2012). Although Japanese text-setting is typically treated as mora-based, the present corpus analysis reveals that syllable-based text-setting is pervasive in Japanese. Two studies presented here compare native Japanese songs with those translated into Japanese. The results demonstrate use of syllabic settings throughout the corpora and across the lexical strata of Japanese. Syllabic settings are shown to arise with greater likelihood in response to pressures imposed by restrictive translation contexts, information density mismatch, and knowledge of correspondence to English loans. We argue that, given the viability of syllabic text-setting in Japanese, moraic text-setting is a stylistic norm of Japanese music that is shifting over time, rather than evidence of a lack of syllable structure in the language's prosodic system.

Keywords: prosodic structure; syllable; text-setting; lexical strata; Japanese

1 Introduction

The syllable is hypothesized by many phonologists to be a universal unit in the prosodic hierarchy: "there exists a higher order hierarchical prosodic structure of which syllables must be seen as building blocks" (Selkirk 1982: 338; see also Kubozono 2003; Hyman 2008). Evidence from Japanese phonology, however, has provoked considerable debate over the necessity, reality, and thus universality of the syllable (e.g., Hyman 2003; Labrune 2012; for parallel debates in other languages, see e.g., Gokana; Hyman 1983, 2011). Japanese is commonly described as a mora-based language, meaning that the mora is considered the basic prosodic unit involved in phonological processes (e.g., Vance 1987; Otake et al. 1993; Inaba 1998; Labrune 2012). Several examples of Japanese words broken down by mora and syllable are given in (1):[1]

(1)   a. 限界   b. ラーメン   c. もっと
           genkai   raamen   motto
           'limit'   'ramen noodles'   'more'

---

[1] Japanese Romanization is given in Hepburn format with the exception of long vowels, which are always written with double letters ('oo') rather than macrons. Personal names are written as preferred by the individual.

The mora is generally credited with doing most of the phonological heavy-lifting in Japanese prosody, playing noted roles in weight, speech timing, production, and perception. Because of the predominance of the mora, the place of the syllable as a necessary prosodic unit in Japanese comes under question.

Two basic camps on the syllable in Japanese may be distinguished: (1) the syllable is not necessary in Japanese prosody and thus has no psychological reality (e.g., Labrune 2012), versus (2) the syllable is necessary in Japanese prosody, but exists in a potentially lesser role alongside the mora (e.g., Kubozono 2003). Under the former, mora-only view, Japanese is argued to distinguish between "regular" moras versus "special" moras (e.g., Vance 1987).[2] Special moras, such as final nasals and long vowels, are generally described as "weak" or "deficient" moras that are adjoined to regular moras. In contrast, the latter view proposes that syllabic structure—that is, "light" monomoraic (e.g., *ka*) versus "heavy" bimoraic (e.g., *kan*) syllables—can undertake the explanatory purposes of "regular" versus "special" moras, and that the syllable as a unit itself has import to Japanese phonology in addition to the mora. If there are syllables, whether moras are nested within syllables in the prosodic hierarchy is also open for debate: nesting, as shown in (1), is dominant in the literature (Kubozono 1989), but other accounts have argued for a three-dimensional prosodic structure in which syllabic structures coexist with a separate, non-nested moraic structures (Ohta 1989, 1991).

Accounts of Japanese prosodic structure must consider potential differences among the strata of the Japanese lexicon. Itô and Mester (1995) distinguish four distinct Japanese lexical strata of diverse historical origins (Table 1). The four lexical strata are characterized not only by distinct writing systems, but also by different phonological constraints, ranging from prosodic to phonotactic. Sino-Japanese stems, for example, are restricted in length to two moras, while two voiced obstruents are disallowed in Yamato stems (Itô & Mester 1996: 19; Kawahara et al. 2006: 32).[3] The influence of foreign phonological systems in certain Japanese lexical strata raises the question of the potential impact of foreign prosodic systems. While the mora is considered the primary prosodic unit of Japanese, Chinese and European languages are syllable-based. It seems plausible, then, that the influence of the syllabic unit would be stronger in Sino-Japanese and Foreign words than in the Yamato stratum. This is particularly likely in the case of the Foreign stratum for two reasons: first, these words are relatively recently borrowed and tend to be less assimilated into the Japanese phonological system, with a stronger resemblance to their cognates in the original language (see Itô & Mester 1993), and second, many Japanese speakers are familiar with the English words corresponding to their loanwords in Japanese.[4] Although previous research has argued that Japanese speakers' understanding of English syllabification is imperfect compared to native speakers, advanced Japanese learners of English are found to variably use the syllable as the salient prosodic unit for segmentation in English, indicating that they have some awareness of the prominence of the syllable in

Table 1: Four lexical strata of Japanese.

Stratum	Origin	Writing system	Example
Yamato	native Japanese	kanji and hiragana	大きい *ookii* 'big'
Sino-Japanese	Chinese	kanji	温泉 *onsen* 'hot spring'
Foreign	primarily English	katakana	オンライン *onrain* 'online'
Mimetic	native Japanese	katakana and hiragana	ピカピカ *pikapika* 'shine'

---

[2] See Otake (2015) for a historical overview of the study of Japanese prosodic structure, including the traditional Japanese linguistics notion of *haku*.

[3] See Itô & Mester (1999) for a full discussion of phonological constraints on Japanese lexical strata.

[4] English loanwords account for approximately 85% of the foreign stratum in Japanese (Irwin 2011: 27).

English relative to Japanese (Ueyama 2003; Tajima 2011). The proposal that the syllable is a more salient prosodic unit in certain lexical strata of Japanese supports and extends an argument from Hyman (2011) that languages exhibit gradient differences in the activation of properties of the syllable: just as the prominence of the syllable may vary across languages, it may also vary within strata of a language that are characterized by different phonological patterns.

If the syllable is a more salient prosodic unit in the case of Sino-Japanese and Foreign strata words, we would expect these prosodic differences to be reflected in certain linguistic tasks. The present analysis focuses on one such task: text-setting, or the pairing of language and music in song. Debates over the role of the syllable in Japanese commonly draw upon evidence from linguistic art forms, including text-setting (Poser 1990; Inaba 1998; Labrune 2012). The standard assumption from linguistic art form evidence in Japanese is that "the mora is the metric unit of Japanese verse in poetry and singing" (Labrune 2012: 116). The well-known poetic form of *haiku* or *senryu*, for example, is generally described as requiring a certain number of moras per line (e.g., 5-7-5) (Labrune 2012), as demonstrated in (2) from Tanaka (2012).[5] (Moras in the example are separated by a '.'; word-internal morphemes are separated by a '+'.)

(2) a. 5 moras: i.tsu. + ma.de. + mo
             when + until + also
             'forever'

     b. 7 moras: a.ru. + to.   o.mo.u. na
             exist + COMP think   IMP.NEG
             'don't expect to exist'

     c. 5 moras: o.ya. + to.   ka.ne
             parents + and   money
             'parents and money'

             'Don't expect your parents and money to be there forever.'

While both the prescriptive and descriptive norms of Japanese linguistic art forms stress the role of the mora, analyses of poetry and song data have consistently uncovered evidence of syllabic units playing a metrical role. Tanaka (2012) observes that multimoraic syllables that count as a single metrical unit are in fact common in Japanese poetry, particularly at the ends of lines, where the mora count can exceed what is licensed by the template. The final two lines in the *senryu* in (3) demonstrate extra moras (example from Tanaka 2012):

(3) a. 5 moras: pu.ro.po.o.zu
             proposal
             'proposal'

     b. **8 moras:** a.no.   hi. + ni.   ka.e. + t.te
             that   day + to   return + and
             'return to that day'

     c. **6 moras:** ko.to.wa.ri. + ta.i
             decline + want
             'want to decline'

             'A proposal of marriage, if I were to be at that time again, I would have declined it surely.'

---

[5] *Senryu* have the same 5-7-5 template as *haiku* but do not have the same semantic restrictions, such as the requirement of seasonal references.

Some authors have also noted that syllable-based segmentation occurs in Japanese songs (Kubozono 1999b; Tanaka 2000). No previous work, however, has examined the patterning of prosodic units in Japanese text-setting from a statistical perspective, or proposed an account of when and why syllables are used.

The present work examines the phonological, lexical, stylistic, and information density-based factors that condition the use of moraic versus syllabic settings in Japanese songs. One roadblock in statistically evaluating the factors predicting text-setting style is the relative infrequency of syllabic settings in conventional songs. The studies reported herein take advantage of the higher rate of syllabic settings in songs translated into Japanese, contrasting one corpus of songs originally written in Japanese with two translated corpora. The higher rate of syllabic settings in translated songs provides an opportunity to examine the linguistic and extralinguistic conditions that may favor or disfavor the syllable in Japanese prosody.

The evidence found in these corpora indicates that the syllable is a crucial phonological unit that is referenced in setting Japanese to music. That the syllable is accessible in Japanese for linguistic art forms demonstrates its reality and necessity in describing prosodic intuitions in the language. Moreover, the syllable is present in all lexical strata of Japanese, regardless of historical origin, although it is found to play a more prominent role in the Foreign stratum. We argue that familiarity with the prosodic structure of English and other syllable-based languages, rather than a difference in underlying prosodic structure, is responsible for a higher frequency of syllabic settings for Foreign words. Furthermore, differences in text-setting style across time and genre, along with other idiosyncratic phonological features of sung Japanese, suggest that the dominance of moraic text-setting in certain genres of Japanese song is a prescriptive, stylistic norm rather than a reflection of a lack of syllabic structure.

The following section introduces relevant work on prosodic units in Japanese, text-setting, and differences in translated versus native texts. Section 3 presents the linguistic variables analyzed in each study. Two studies are then presented, comparing translated Disney songs to native cartoon theme songs (Study 1) and subsequently to translated Christmas songs (Study 2).

2 Background

2.1 Prosodic units in Japanese

Kubozono (1996a: 78) identifies four functions of the mora in Japanese: (1) phonological weight, (2) speech timing, (3) speech segmentation in production, and (4) speech segmentation in perception. In the realm of speech timing, there is considerable evidence that Japanese exhibits mora timing, meaning that speakers produce each mora with approximately the same duration (Han 1962; Port et al. 1987; Warner & Arai 2000). In segmentation tasks, the mora rather than the syllable is consistently found to be the preferred segmentation unit of Japanese listeners (Otake et al. 1993).

There is a lack of consensus, however, concerning the extent to which the mora functions as an active prosodic unit. In a series of corpus studies on naturally-occurring speech, Arai and colleagues find mora timing to be less evident in spontaneous speech and suggest that the mora influences timing only indirectly through its role in the phonological structure of Japanese (Arai & Greenberg 1997; Arai 1999; Warner & Arai 2001; etc.). Beckman (1982) argues similarly that the mora is perceived by native speakers as a salient unit only due to the mora-based orthography of Japanese, and does not function as a real prosodic unit in spontaneous speech. Some evidence from Japanese children's perception supports this view, indicating that children shift to exclusively mora-based segmentation once they become literate in the mora-based *kana* writing system (Inagaki et al. 2000).

One problematic aspect of frameworks in which the mora is the basic prosodic unit is the behavior of certain moraic units, such as coda nasal, which cannot occur in isolation and are dependent on their adjoining moras in various other respects. The proposed distinction between regular and special moras has been used to account for these phenomena. Scholars who use this framework, however, disagree over precisely which moras are special, and whether moras can be divided into two uniform classes or are more properly organized along a scale. The core set of special moras identified by most analyses are the first portion of geminate consonants (often written as /Q/), as in the first *t* in *zutto* ('throughout'); the coda nasal (written as /N/), as in *san* ('three'); and the second portion of long vowels (written as /R/), as in the second *i* in *hoshii* ('to want'). Additional proposed special moras are those containing high vowels that are devoiced between voiceless consonants (e.g., *su* in *suki* 'like'), certain second vowels in two-vowel sequences (e.g., *i* in *mai* 'every'), and onsetless vowels (Labrune 2012: 116, 135). Most phonological analyses of Japanese propose a simple distinction between regular and special moras; experimental studies, however, consistently find varying degrees of independence among the set of special moras in both production and perception. Most studies have found that /R/ and /N/ are most closely connected to the preceding mora in segmentation tasks while /Q/ and second vowels function more independently (Otake 1992; Matsuzaki 1994; Tamaoka & Terao 2004); Machida (1988), on the other hand, argues for the opposite order. Labrune (2012: 141) proposes an alternative analysis in which all moras are arranged along a sonority-based scale starting with C*a* and ending with the least sonorous unit, /Q/, with moras decreasing in independence as they decrease in sonority.

Aside from the argument that syllables are a linguistic universal and therefore must exist in Japanese, advocates of the syllable in Japanese point to several phenomena that are argued to depend upon the syllabic unit, including avoidance of trimoraic syllables and accentuation patterns of loanwords (Ohta 1991; Kubozono 1996b; Kubozono et al. 2008).[6] Critics have contested the claim that only the syllabic unit may account for these patterns; they also point to the fact that the syllable was an unknown unit in traditional Japanese phonology until it was introduced by Western linguists (Inaba 1998; Labrune 2012). As discussed above, psycholinguistic research comparing the mora and the syllable has consistently pointed to the mora, and not the syllable, as the primary unit in production and segmentation (e.g., Otake et al. 1993). The fact that the mora is a more prominent unit does not, however, constitute negative evidence against the existence of the syllable. Indeed, psycholinguistic evidence does indicate that the syllable is a psychological reality in Japanese (Tamaoka & Terao 2004; Nakamura & Kolinsky 2014; see Kawahara 2016 for a summary).

The foot, traditionally defined as a prosodic unit consisting of two syllables, may also play a key role in Japanese. Labrune (2012) and Inaba (1998), extending proposals from Poser (1990) and Bekku (1977), argue that, contrary to the usual structure of a foot, the Japanese foot consists of two moras, the first of which must be a regular CV mora. This proposal thus eliminates the need for the syllabic unit by proposing that two moras in a word such as *san* (*sa.N*, 'three') should be analyzed as one foot containing one regular and one special mora, rather than as a syllable.

The foot-based versus syllable-based analyses of multimoraic sequences such as *san* make crucially different predictions with regard to text-setting: while a multimoraic syllable may be assigned a single note through a syllable-based text-setting, a bimoraic foot generally corresponds to two distinct beats in poetry and song. In fact, a bimoraic foot

---

[6] See Labrune (2012) and Kawahara (2016) for more complete examinations of the evidence for and against the existence of the syllable in Japanese.
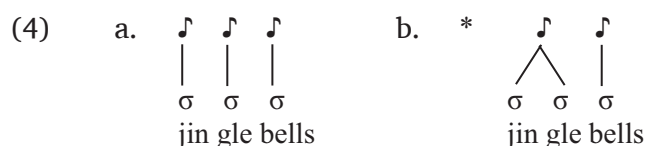
consisting of two regular CV moras must necessarily correspond to two distinct notes because it contains two separate attack points. Evidence that multimoraic sequences are frequently treated as single units in text-setting would therefore constitute a significant challenge to foot-based theories that reject a role for the syllable, as these accounts cannot explain why only certain feet may be treated as a single unit without recreating the notion of a syllabic unit corresponding to a subset of bimoraic feet.

2.2 Text-setting and Japanese song

2.2.1 Text-setting and prosodic structure

Linguistic art forms, including styles of poetry and song, are hypothesized to develop in relation to the phonological and prosodic structures of particular languages. A considerable literature in metrics and poetics has examined the role of prosodic structure in poetic forms (Kiparsky 1973, 1977, et seq.; Halle & Keyser 1971; Hayes 1988; Hanson & Kiparsky 1996). Text-setting, the arrangement of lyrics over a melody, also makes use of the salient prosodic units particular to a language. In the case of English, syllables and lexical and phrasal stress play a large role in text-setting (Halle & Lerdhal 1993; Shih 2008; Hayes 2009). For example, the word *prosody* would most naturally be set so that the most musical prominence fell on the initial stressed syllable. In tonal languages such as Cantonese and Tommo So, lexical tone influences text-setting so that there is often a correspondence between tone contour and melodic line (Yung 1991; McPherson 2014).[7]

Phonologists have made use of the relationship between text-setting and prosodic structure to investigate the structure of a variety of languages through evidence from song (Schuh 1994; Noel 2010; Calder 2013; Sui 2013). This area of research uses the distribution of segments over musical notes as evidence of prosodic units; if a group of segments can be set on a single note, this can be taken as evidence suggesting that these segments constitute a single salient prosodic unit, such as a syllable. A "note" here refers to the basic unit of tonal music representing a continuous sound that may vary in length. Notes are auditorily distinct from other notes as they each have an "attack point" or onset. A sequence of notes may maintain the same pitch but nonetheless be distinct if they have a perceptible disjunction.[8] For example, in the Christmas carol *Jingle Bells*, the beginning of the chorus, "jingle bells," consists of three separate notes on the same pitch, with each note corresponding to a syllable, in (4a).[9] Due to the prosodic structure of English, it would sound unnatural or even impossible to assign more than one syllabic unit to a single note: see (4b).

(4)　　a.　♪　♪　♪　　　b.　*　　♪　　♪
　　　　　　|　|　|　　　　　　　　/\　　|
　　　　　　σ　σ　σ　　　　　　　σ　σ　σ
　　　　　jin gle bells　　　　　　jin gle bells

2.2.2 Text-setting and prosodic structure in Japanese

The present study focuses on Japanese text-setting in modern Western-style music rather than traditional Japanese musical forms such as *shigin*, which often lack the regular

---

[7] A reviewer points out that certain musical and poetic traditions have been argued to not straightforwardly utilize phonological units that are active in a language's grammar (see Golston & Riad 2005; Katz 2015). Nonetheless, overwhelming evidence in the existing literature indicates that, at the very least, unnatural phonological units are not commonly used, nor are they perpetuated in artistic practices (Hanson & Kiparsky 1996: 288).

[8] See Lerdahl & Jackendoff (1983) for a more complete discussion of the structure of tonal music.

[9] In all examples, a musical note is represented as "♪", regardless of its temporal duration.

relationship between prosodic timing and musical rhythm that is characteristic of Western music (Manabe 2009). Composition of Western-style music became widespread in Japan following the Westernization movement of the Meiji Restoration in 1868 (Manabe 2009). Today, Western and Western-derived musical genres—including pop, rock, and hip-hop—dominate the popular music landscape in Japan.[10]

The majority of phonological descriptions of Japanese metrics in song and verse describe the text-setting system as moraic, meaning that one mora is assigned to one note (or at least one note) (Otake et al. 1996: 3831; Hayes & Swiger 2008: 2; Labrune 2012: 116). This includes the special moras: those which are connected with the previous mora in a single syllable (e.g., coda nasal). Moraic setting contrasts with syllabic setting, in which multimoraic syllables would receive only one note. Thus, the phrase *onsei gakkai* ('phonetics society') would receive eight notes in a moraic setting (5a), versus four notes in a syllabic setting (5b; example adapted from Tanaka 2000: 31).

(5)   a.   moraic setting

♪ ♪ ♪ ♪ ♪ ♪ ♪ ♪
| | | | | | | |
μ μ μ μ μ μ μ μ
o n se i ga k ka i

   b.   syllabic setting

♪   ♪   ♪   ♪
⋀   ⋀   ⋀   ⋀
μ μ μ μ μ μ μ
o n se i ga k ka i

Hayes & Swiger (2008) take the position that Japanese text-setting is moraic as a result of the mora-based prosodic structure of Japanese. They argue that long vowels and other moras are naturally heard by native speakers as distinct rhythmic beats, and therefore must be assigned two separate notes:

> "[…]Japanese speakers 'feel' a rhythmic beat on every mora of a word. Thus Tookyoo is felt to have four rhythmic beats (to o kyo o), whereas an English speaker listening to the very same word (in correct Japanese pronunciation) hears just two beats: too kyoo[…]. This has consequences for music. Because Japanese speakers are more closely attuned to moras than syllables, they tend to feel that every mora counts as (at least one) note of the music, just as English speakers feel that every syllable counts as at least one note." (Hayes & Swiger 2008: 2)

Analyses by native listeners, on the other hand, tell a different story. In her 2009 musicology dissertation on Western music in Japan, Noriko Manabe is critical of the moraic text-setting of a song from the early 20th century, arguing that a syllabic setting would sound better:
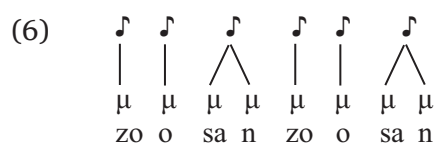
> "One curious aspect of the text setting [of the song *Issun Boshi*] is the insistence of setting each mora as one note, including special moras such as the nasal N, double consonant, and long vowels. As explained in chapter one, these moras are spoken with the duration of one mora but are heard as subsumed into the previous syllables. These words are more naturally set as syllables." (Manabe 2009: 149)

Surprisingly, Manabe describes the moraic setting of this song as "curious," as if such a style is unusual or unexpected, even though moraic setting is viewed in most phonological work as the default or universal Japanese text-setting style. Nonetheless, she finds the moraic

---

[10] For example, all of the top 100 singles on the Japanese Billboard chart as of May 2012 are Western-style genres (www.billboard-japan.com).

settings used in this song to be unnatural. Manabe specifically argues that words containing special moras are more appropriately set syllabically, rather than moraically. Elsewhere in her dissertation, she describes syllabic setting as optional: "when two moras form one syllable, one note may be used, preferably with longer duration" (Manabe 2009: 30).

Manabe's musicological analysis suggests that the claim in the phonological literature that Japanese text-setting is fundamentally moraic is an oversimplification. Indeed, a cursory look at Japanese song data reveals many examples of syllabic settings. For instance, in the children's melody *zoo san, zoo san* ('mister elephant, mister elephant'), as shown in (6), while the two moras in the long vowel in *zoo* are each assigned individual notes, the syllable *san* is treated as a single unit rather than the coda nasal mora receiving a separate note:

(6)

Previous studies of variation between moraic and syllabic setting confirm that this is a common phenomenon, although setting style differs by genre. Kubozono (1999b) found a 2:1 preference for moraic settings in traditional songs (i.e., Western-style songs composed following the Meiji Restoration) while Sugitou & Sakai's 1999 study of traditional children's songs found a strong preference for moraic settings. Tanaka (2000) examined moraic versus syllabic settings in nouns within Japanese popular music lyrics, comparing *enka* music, an older Western-style popular ballad genre, with other genres. His study revealed an increasing use of syllabic settings over the course of the 20th century, moving from an average of 1.13 notes per special mora in the 1930s to only 0.69 notes per special mora in the 1990s (Tanaka 2000: 154). Tanaka also observed that *enka* music used more moraic setting than other genres and that, for certain time periods and genres, syllabic settings were more common for Foreign words and least common for Yamato words (Tanaka 2000: 155).

Collectively, this research suggests a preference for moraic text-setting in traditional, older musical genres while popular music is moving toward a more syllabic style; with the exception of Tanaka's study, however, little is known about Japanese text-setting in modern popular music. Furthermore, previous work offers no statistical analysis controlling for interactions between factors, or addresses which linguistic contexts may be conducive to moraic versus syllabic settings.

2.2.3 Japanese singing style

When analyzing Japanese text-setting for evidence of prosodic structure, it is important to keep in mind that Japanese is generally sung in a style that differs from spoken Japanese in certain respects. For example, the devoicing of high vowels between voiceless consonants does not occur in traditional singing style, so that a word such as *suki* ('like') would be pronounced [sɯki], rather than [sɯ̥ki]. Also, geminate consonants are typically dropped or replaced by a long vowel, so that *itte* ('say') becomes [ite] or [iite] (Dell 2011: 191). These differences conspire to increase the sonority of Japanese by avoiding devoiced or silent segments, thus making it more appropriate for singing.

Overall, these differences mean that sung Japanese is a conservative site to measure the saliency of the syllable, as mora-based segmentation would generally be more sonorous and thus in that sense should be preferred in singing (e.g., su-ki moraic vs. ski syllabic). In other words, if there is evidence for syllable-based segmentation in sung Japanese,

this would suggest that the syllable is salient enough to override sonority concerns. One exception to this trend is coda nasal: the lower sonority of nasals as compared to vowels may cause a dispreference for the production of isolated nasals in the context of singing. Regardless of these sonority concerns, the fact that a solution is available in which the coda nasal groups with the previous mora would strongly suggest that the syllable is a salient prosodic unit in both sung and spoken Japanese.
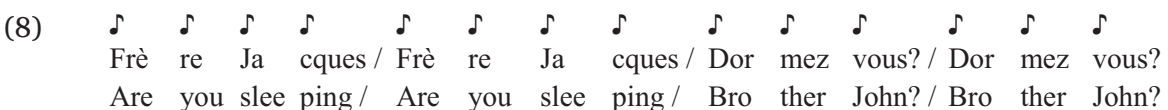
2.3 Translated versus native text

The following two studies contrast native versus translated text-setting: i.e., song lyrics originally written in Japanese versus songs that have been translated into Japanese. The field of translation studies boasts a considerable body of research indicating that translated text consistently differs from native text, to the extent that it is argued by some to be a "hybrid" form that is typical of neither the original nor translated language (Trosborg 1997; Schaffner & Adab 2001). One common finding is that translated texts are simpler and more explicit than native texts with respect to lexicon, discourse structure, and other features (Blum & Levenston 1978; Pápai 2004: 144). These differences between translated and native text are thought to occur primarily as a result of the constraints posed by the original text; the translator is not free to compose and arrange their own message, but must convey the thoughts of another author (Chesterman 2004: 44).

Song translation is a particularly challenging translation context. Songs frequently contain poetic language, puns, and other linguistic devices that are difficult to translate. Furthermore, the translator must work within the restrictions created by the melody, notes, and metrical structure. For example, let us consider the well-known French children's melody Frère Jacques:

(7)   ♪   ♪   ♪   ♪   ♪   ♪   ♪   ♪   ♪   ♪   ♪   ♪   ♪   ♪
      Frè re   Ja   cques / Frè re   Ja   cques / Dor mez vous? / Dor mez vous?

The most literal English translation would yield, "Brother Jacob, Brother Jacob, are you sleeping? Are you sleeping?" However, because there is no way to fit "are you sleeping" into the three notes provided by "dormez-vous," the English version of the song flips the lines and changes the wording as follows:

(8)   ♪   ♪   ♪   ♪   ♪   ♪   ♪   ♪   ♪   ♪   ♪   ♪   ♪   ♪
      Frè re   Ja   cques / Frè re   Ja   cques / Dor mez vous? / Dor mez vous?
      Are you slee ping / Are you slee ping / Bro ther John? / Bro ther John?

While there is some flexibility in adding or deleting notes during the translation process to fit extra syllables into a melodic line, this is considered suboptimal because it alters the original intentions of the composer. In the case of the songs examined in the following study of Disney films, such modifications must be kept to a minimum to maintain a realistic match between the audio and the visible mouth movements of the animated characters (a.k.a. "mouth flap") (Baker & Hochel 2001; Yu 2013; Warachananan & Roongrattanakool 2015).

The length restrictions inherent in song translation present a particular challenge for translation into Japanese. Due to its small phonological inventory, the syllable-based average information density of Japanese is much lower than in English and many other languages: in other words, it takes more syllables to say the same thing in Japanese (Pellegrino et al. 2011: 544). If we consider the mora rather than the syllable to be the primary prosodic unit of Japanese, the difference becomes magnified, with an even larger

gap in information density. Examples of equivalent expressions in English and Japanese to illustrate the information rate per syllable and mora are given in (9) and (10):

(9)	a.	I love cheese.					3 syllables
	b.	Watashi	wa	chiizu	ga	daisuki.	12 moras (9 syllables)
		I	TOP	cheese	NOM	love	

(10)	a.	Where is your umbrella?							6 syllables
	b.	Anata	no	kasa	wa	doko	desu	ka?	12 moras (11 syllables)
		2SG	GEN	umbrella	TOP	where	COP	Q	

In casual contexts, Japanese speakers often drop arguments, case-markers, and other material (e.g., *chiizu daisuki*, 'cheese love', 7 moras, for (9)), generally leading to less information being conveyed. Even this degree of elision, however, is not sufficient to close the gap with English. This difference in information density is partially compensated for by a faster speech rate in Japanese; Pellegrino et al. (2011: 544), for example, found an average of 7.84 syllables per second in Japanese versus 6.19 in English. In the context of song translation, however, an equivalent increase in production rate is not easily achievable; more syllables cannot be inserted into a given time period without the addition of notes. If the lyrics are arranged mono-moraically, with one mora per note rather than one syllable per note, this problem becomes more severe, since even less material can fit into a given line.[11] Moreover, while Japanese allows for considerable deletion, there is a limit to how much material can be elided while preserving meaning.

In light of these factors, there is strong evidence to predict that songs translated into Japanese will contain more syllabic settings than songs originally written in Japanese. From an information density perspective, syllabic settings, which fit more linguistic material into a single note than moraic settings, will be needed to convey the necessary information from the original lyrics. More generally, evidence from translation studies suggests that translation contexts create restrictions on content and arrangement that will result in more non-traditional text-setting. Translated songs also may differ from native songs in lexical strata distribution, with a higher proportion of Foreign words due to their non-Japanese subject matter; if such words are indeed more likely to be set syllabically, this greater number of Foreign words would result in more syllabic settings in translated songs. The following studies will test whether translated and native songs do in fact differ with respect to text-setting style.

3 Linguistic variables

3.1 Introduction

Because there is no difference between moraic and syllabic text-setting in the case of mono-moraic syllables (i.e., CV syllables), this analysis must focus on multi-moraic syllables, and specifically on "special moras" which form a syllabic unit with the proceeding mora (Kubozono 1999a). The following sections detail each linguistic variable examined. Where relevant, details are given regarding how these variables are distributed among the lexical strata, as well as how they are realized in sung versus spoken Japanese.

---

[11] The low information density of Japanese presents a problem not just for translated songs, but for Japanese Western-style songs in general. As Manabe (2009: 37) observes, "the number of moras in a semantically complete block of text can far exceed the number of notes that can be sung in a four-measure phrase."

3.2 Coda nasal

Coda nasal, as in *san* ('three'), is a nasal phoneme that assimilates to the place of articulation of the following segment, yielding variants such as [m] before labials (e.g., *sampo* 'walk'), [n] before coronals (e.g., *bento* 'lunch box'), [ŋ] before velars (e.g., *genki* 'healthy'), and [N] utterance-finally (e.g., *hon* 'book') (Otake et al. 1996: 3932). Some analyses treat the coda nasal as unspecified for place of articulation, while others claim it is underlyingly uvular (Vance 1987). Most agree, however, that coda nasal is an independent mora that is distinct from syllable-initial phonemes /n/ and /m/; evidence from experimental work indicates that the coda nasal is processed differently from nasal initials, and as an independent moraic unit (Otake et al. 1996). A word such as *minna* ('everyone'), therefore, is analyzed as containing the three moras (*mi-n-na*) and two syllables (*min-na*).[12]

The coda nasal, which first appeared in Early Middle Japanese in approximately 800 AD, is believed to be the product of contact with Middle Chinese (Loveday 1996: 41; Tranter 2012: 214). As a result, coda nasals are rare in native Yamato words, but relatively common in Sino-Japanese and Foreign words. Certain Yamato words, such as *minna* ('everyone'), contain coda nasals due to phenomena such as the reduction of *nV* and *mV* sequences (Tranter 2012: 214). In the case of Sino-Japanese words, syllables ending with coda /n/ and /m/ in Middle Chinese retained the coda nasal in Japanese, yielding modern Sino-Japanese words such as *ningen* ('human'). Foreign words containing coda nasal include many English and other European-derived loanwords with syllable-final /n/, such as *guriin* ('green').

In moraic text-setting, because the coda nasal is a distinct mora, it receives its own separate note (see Table 2). We observe that the separation of the nasal from its preceding mora results in a reduction of coarticulation effects; specifically, the preceding vowel is usually not nasalized, as it would be in spoken Japanese. In certain cases, the nasal is so isolated that does not even assimilate to the place of articulation of the following segment (e.g., *sa-m-po* vs. *sa-N-po* 'walk'). In contrast, in syllabic settings, the nasal is joined with the preceding mora on a single note, and vowel nasalization is preserved. As a result of these differences, the two settings are quite perceptually distinct, with the syllabic setting sounding more similar to spoken Japanese in terms of phonological effects. The behavior of coda nasal in each setting is given in Table 2.

To control for potential effects of word-final lengthening and differences in coda nasal position among the lexical strata, word-final versus word-medial position will be incorporated as an additional factor in statistical models for coda nasal.

3.3 Long vowels

Japanese distinguishes between short vowels and long vowels, as in *obasan* ('aunt') vs. *obaasan* ('grandmother'), which differ primarily in duration (Vance 1987; Hirata 2004). These long vowels are traditionally analyzed as consisting of two shorter vowels placed

Table 2: Possible settings for coda nasal.

Word	Moraic	Syllabic
ningen	ni-n-ge-n	nin-gen
'human'	[ni N ge N] or [ni ŋ ge N]	[nĩŋ gẽN]
sampo	sa-n-po	san-po
'walk'	[sa N po] or [sa m po]	[sãm po]

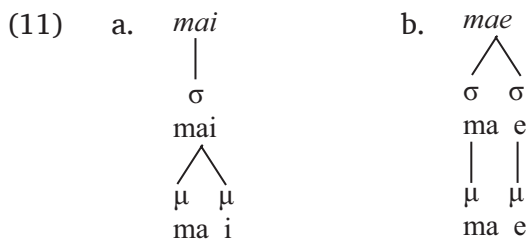[12] We use "–" to denote musical note boundaries.

together (e.g., as [aa] rather than [a:]) (Martin 1952: 13; Labrune 2012: 116). Thus, a syllable *baa* would consist of two moras, *ba* and *a*. Long vowels occur for all five of the main vowels in Japanese (*a, i, u, e, o*). Certain long vowels, however, do not occur with certain lexical strata. Long *a*, for example, does not occur with Sino-Japanese words (Martin 1952: 13; Moreton & Amano 1999).

Although they are considered to be composed of two short vowels, in spoken Japanese long vowels are produced as one continuous vowel. In moraic text-setting, however, the second mora is split from the preceding mora and sung as a distinct segment. The syllable *baa*, for example, would be produced as *ba-a*. In syllabic text-setting, as in spoken Japanese, there is no break between the two moras and they are placed on a single note. The two settings are detailed in Table 3.

As the five vowels differ in sonority, we will control for sonority in statistical models for long vowels, assuming a scale from open to closed of *a* > *o, e* > *u, i* (Labrune 2012: 141). Word final position and identity of the vowel will also be tested as potential factors.

3.4 *Vi sequences*

In traditional phonological analyses of Japanese, sequences of two consecutive vowels, as in *mae* ('front') are analyzed as two short vowels rather than as diphthongs (Martin 1952: 13; Vance 1987; Labrune 2012). Some scholars propose a distinction between Vi—those sequences which end in /i/ (*ai, oi, ei, ui*)—and those which do not, such that /i/ is analyzed as a special mora while other second vowels in sequences are regular mora (Tanaka 2000: 31; Labrune 2012: 116). In this view, Vi constitutes the rime of a single syllable, while other non-identical VV sequences are separate syllables, as illustrated in (11):

(11)   a.   *mai*   b.   *mae*

```
        |                          /\
        σ                         σ   σ
       mai                       ma   e
       /\                         |   |
      μ  μ                        μ   μ
      ma i                       ma   e
```

This analysis raises several questions about the status of /i/ and what distinguishes Vi from other VV sequences. If Vi is indeed the rime of a single syllable, it is unclear what phonetic details might exist in connected speech to distinguish between the analysis of Vi as two sequential monophthongs in which /i/ is a special mora versus an analysis in which /i/ is an offglide in a diphthong; indeed, Kubozono (2015) proposes that intra-morpheme Vi sequences ought to be classified as diphthongs. In the Sino-Japanese and Foreign strata the diphthong question is particularly relevant since the Vi sequence is equivalent to diphthongs in the original language (e.g., *naito* 'night'). It is also notable that no orthographic distinction exists between Vi sequences and other VV sequences, raising the possibility that other sequences of decreasing sonority, such as *ao*, might sometimes be treated as

Table 3: Possible settings for coda long vowels.

Word	Moraic	Syllabic
hoshii	ho-shi-i	ho-shii
'want'	[ho ɕi i]	[ho ɕi:]
keeki	ke-eki	kee-ki
'cake'	[ke e ki]	[ke: ki]

multi-moraic syllables as well; this seems most plausible in the Sino-Japanese and Foreign strata, which draw from languages containing a variety of diphthongs. Nonetheless, in the present text-setting corpora only sequences of Vi were found to receive syllabic settings; given that there were only 29 tokens of VV that were not Vi in the three corpora examined in the two studies, the lack of syllabic settings here may simply be due to low token count. Because no variation was observed in other VV sequences, this analysis will focus exclusively on Vi sequences. We will examine whether differences in sonority between the first vowel and the /i/ (e.g., *ai* vs. *ei*) have any impact on the likelihood of the sequence receiving a syllabic setting in which both vowels are placed on a single note.

As in the case of coda nasal, in a moraic setting in which the /i/ is sung on a separate note, the coarticulatory effects generally present in spoken Japanese are absent, making the initial vowel and the following /i/ more perceptually distinct. In a syllabic setting, the coarticulatory effects are comparable to spoken Japanese and the Vi is produced with a more gradual shift from V to /i/.

The behavior of Vi sequences in different text-setting styles is given in Table 4. As with long vowels, for Vi, differences in sonority may produce significant differences in text-setting behavior. In this case, we predict that Vi sequences with a larger drop in sonority, such as *ai*, will be more likely to be set syllabically, as they are perceptually more distinct and conform more closely to the preferred sonority structure of a syllable. Sonority drop, assuming the same sonority scale as for long vowel, will therefore be incorporated into the statistical model, in addition to identity of the first vowel and word final position.

3.5 High vowel devoicing

In spoken Japanese, the high vowels *i* and *u* generally devoice between two voiceless consonants and in certain word-final environments, so that *suki* ('like') is pronounced [su̥ki] and *desu* ('is') is [desu̥] (Vance 1987; Shibatani 1990; Tsuchida 2001). The phonetic reality of this phonological rule is complex: *i* and *u* devoice variably, other vowels sometimes also devoice, and devoicing can occur in voiced environments (Arai 1999; Maekawa & Kikuchi 2005). Phonetic studies of spontaneous speech data indicate that devoicing of these vowels may be realized as dramatic reduction or even deletion, resulting in consonant clusters (Arai 1999; Labrune 2012: 136). While most authors do not include moras containing devoiced vowels in the class of special moras, some do propose such an analysis (e.g., Warner & Arai 2001: 1144). Under these accounts, in the case of dramatic vowel reduction or deletion, words such as *suki* are generally analyzed as consisting of two moras that form a single multi-moraic syllable, such as in the realization [ski] (Kondo 2005). This syllable structure would differ from others proposed to exist in Japanese, as the special mora would precede the regular mora, rather than the reverse. Alternatively, some advocates of the Japanese syllable propose that, even in the case of vowel deletion, the remaining single consonant retains status as a separate syllable (Kawahara & Shaw 2016; Matsui, to appear); this proposal does not differ from the moraic account in terms of

Table 4: Possible settings for Vi sequences.

Word	Moraic	Syllabic
eigo	e-i-go	ei-go
'English'	[e i go]	[ei go]
hai	ha-i	hai
'yes'	[ha i]	[hai]

its predictions for segmentation in text-setting, as, in either case, the remaining consonant would be placed on a separate note.

Due to variable devoicing of *i* and *u,* there are three possible text-setting outcomes for this phonological context. As discussed in §2.2.3, in traditional Japanese singing style, vowel devoicing does not occur; this is understandable, since a devoiced vowel does not make for great singing. The word *suki* ('like'), for example, would be realized as [sɯki]. In such a case, the mora containing the full vowel must be treated as an independent note, since it cannot form a multi-moraic syllable without reduction or deletion of the vowel. However, in singing styles that come closer to spoken Japanese, the *i* and *u* are devoiced as they are in speech. This outcome yields two possible text-settings: in a moraic setting, the mora containing the devoiced vowel remains on a separate note, yielding a setting such as [sɯ̥-ki]. In a syllabic setting, the vowel is reduced or entirely deleted and the remaining mora joins the following mora on a single note as [sɯ̥ki] or [ski].[13]

The present analysis will focus on devoicing of *i* and *u* between voiceless consonants; devoicing of other vowels and devoicing of word-final vowels, as in *desu,* will not be examined. The possible outcomes for this variable are listed in Table 5.

We expect the syllabic option to occur more frequently when a consonant cluster is articulatorily easier to produce—specifically, when one of the consonants is a continuant and the other is a stop, as in *suki* → [ski], rather than *kite* → [kte]. This factor will be controlled for in the statistical model, in addition to manner of articulation of the first and second consonant, and identity of the high vowel (*i* versus *u*).

3.6 Excluded variables

Several additional variables, although interesting in relation to text-setting, are not included in the present study. Geminate consonants (e.g., *robotto* 'robot') are traditionally considered to contain a special mora, often written as /Q/, although analyses differ as to how this geminate mora functions in the Japanese phonological system (Tanaka 2000: 31; Labrune 2012: 115). In moraic text-setting, the geminate should theoretically receive its own note, resulting in *ro-bo-Q-to.* From a practical perspective, however, a geminate mora is simply not singable. This problem is most frequently resolved by extending the vowel from the preceding mora, so that *ro-bo-Q-to* becomes *ro-bo-o-to* (Dell 2011: 190). Another solution used to convey the geminate consonant is to add a coda to the preceding mora, as in *ro-bot-to.* In other cases, the geminate is simply deleted, yielding *ro-bo-to.* Because the relationship between these possible outcomes and a syllabic versus moraic setting is ambiguous, geminates will not be analyzed in the following corpus studies.

Table 5: Possible settings for high vowels between voiceless consonants.

Word	Moraic	Syllabic
suki	su-ki	suki
'like'	[sɯ ki] or [sɯ̥ ki]	[ski]
ashita	a-shi-ta	a-shita
'tomorrow'	[a ɕi ta] or [a ɕi̥ ta]	[a ɕta]

---

[13] While three outcomes are possible in the case of high vowels in a devoicing environment (sɯ-ki, sɯ̥-ki, and ski), it is worth noting that, generally speaking, only the moraic versus syllabic setting is decided by the composer. Given a moraic setting in which *su* is allotted its own note, the decision regarding whether to voice or devoice the vowel is determined by the performer, because there is no conventional way to indicate devoicing in written Japanese. Japanese sheet music does occasionally attempt to indicate a voiceless vowel by use of a "ghost note", meaning a note written with an "x" in place of the dot, used in Western music to indicate rhythmic notes with no pitch in a variety of contexts (Shigeto Kawahara, *p.c.*). Other factors, such as the duration of the note, may also influence the realization of potentially voiceless vowels.

Also excluded from analysis are word-final high vowels which are generally devoiced, as in *desu*, as these contexts were rare and variation was not observed in the initial stages of data analysis. Additionally, while pitch accent is thought to play a role in Japanese text-setting, it will not be considered here, as it is a complex factor that may interact with melody (cf. Manabe 2009).

3.7 Summary

Table 6 summarizes the four variables and the contextual linguistic factors examined for each variable.

4 Study 1

4.1 Introduction

Study 1 explores the linguistic and stylistic factors that coincide with different text-settings in translated and native Japanese songs. Translated and native songs are contrasted here in order to test the prediction that, in the restrictive context of translation, syllabic settings will become more frequent. This study examines a particularly difficult translation context, songs from animated Disney films, and contrasts that corpus with a contemporary native corpus of anime theme songs.

4.2 Corpora

As discussed in §2.3, we predict that translated songs will contain a higher proportion of syllabic settings than native songs. Data yielding more syllabic settings allow us to better understand which linguistic factors predict syllabic versus moraic setting; it is therefore desirable to select a relatively difficult translation context, which will be most likely to yield syllabic settings. The songs in Disney animated films present particular difficulty for several reasons. First, the translator is accountable to Disney and must produce a translation as close as possible to the original. Disney songs contain considerable semantic content that must be preserved to facilitate the storyline, including many foreign personal and place names. Finally, in the context of animation, adding or deleting notes to accommodate the translation is strongly discouraged because the resulting audio will no longer match the mouth movements visible on screen. Disney songs also contain quite a lot of text with little repetition compared to typical pop songs, making them well-suited to the present analysis.

The translated corpus includes 17 songs from Disney films spanning *The Little Mermaid* (1989) to *The Lion King* (1994), translated into Japanese between 1989 and 1997 (see Appendix A). The films of this period, and their songs, form a cohesive genre written in a musical theater style. This era of films was wildly successful and marked a revival in the

Table 6: Variables, possible realizations, and linguistic factors.

Variable name	Description	Example	Moraic setting	Syllabic setting	Linguistic factors
Coda N	syllable coda nasal	*hon* 'book'	ho-n	hon	word final
Long V	long vowel	*koohii* 'coffee'	ko-o-hi-i	koo-hii	sonority, vowel, word final
Vi	vowel followed by /i/	*mai* 'every'	ma-i	mai	sonority drop, vowels, word final
Inter-voiceless IU	*i, u* between voiceless consonants	*tsuki* 'month'	tsu-ki or tsu̥-ki	tski	i vs. u, manner of preceding and following consonant, consonants form a cluster of a stop and continuant

fortunes of Disney; their international success prompted Disney to move from third-party localization companies to in-house teams. As a result, while *The Little Mermaid* was translated into Japanese by Pony Canyon in 1989, subsequent films were all localized by Disney. Evidently unhappy with the third-party translations for *The Little Mermaid*, the film was redubbed in 1997, using the same voice cast but different translations for many of the songs. The following analysis includes both the original 1989 translation and 1997 translation; differences in the text-setting of the two versions will be discussed in §4.5.1.

To contrast native and translated songs while controlling for time period and genre as much as possible, native songs were selected from Japanese anime (animated) children's films of the same time period. Because anime films do not contain the same quantity of songs as Disney films, additional songs were drawn from contemporary television shows. The native anime corpus includes a total of 11 songs from a period of 1988 to 1996 (see Appendix A).

Examples of consecutive lines from each corpus containing moraic and syllabic settings are given in (12) and (13). In the Disney corpus, the bolded instances in (12a) are examples of syllabic setting while the instance in (12b) is moraic. In the anime corpus, the bolded instances in (13a) are syllabic and (13b) are moraic. Hyphens indicate word-internal musical note boundaries; for ease of reading, morpheme boundaries are not shown in the Romanized Japanese.

(12)   Translated Disney corpus, "Belle" (1991)[14]
   a.   i-tsu-mo-to   o-na-ji pa**n**-ya   sa**n**-ga
        when+also+with   same   bread+shop HON+NOM
        'the same baker as always'
        Original English lyrics: "There goes the baker with his tray like always."

   b.   pa-**n**   wo   u-ri   ni   ku-ru
        bread   ACC   sell   DAT   come
        'comes to sell bread'
        Original English lyrics: "The same old bread and loaves to sell."

(13)   Native anime corpus, "Give a Reason" (1996)
   a.   da-re-ni-mo   to-me-ra-re   wa   shi-na**i**
        who+DAT+also   stop+PASS   TOP   do+NEG
        'I won't be stopped by anyone'

   b.   mi-ra-**i**   no   ji-bu-**n**   e-to
        future   GEN   self   to
        'to my future self'

4.3 Methodology

All occurrences of coda nasals, long vowels, Vi, and inter-voiceless *i/u* were coded perceptually by the first author using audio data;[15] tokens were coded as either receiving a moraic setting or a syllabic setting. Each token was also assigned a lexical stratum and coded for the contextual linguistic factors indicated in Table 6.[16]

Certain tokens were excluded from analysis. Words from the Mimetic stratum containing the linguistic features analyzed in this study (e.g., *pyon pyon* 'hop') were extremely

---

[14] Lyrics in (12), (15), (16) and (17) are excerpted from songs published by the Walt Disney Music Company in various years. Lyrics in (13) are excerpted from the song "Give a Reason", published in 1996 by King Records. Lyrics in (18b) are excerpted from the song サンタ・クロースがやってくる ("Santa Claus is Coming to Town") published in 1959 by Victor. Lyrics in (18c) are excerpted from the song ママがサンタにキッスした ("I Saw Mommy Kissing Santa Claus") published in 1962 by Nippon Columbia. All rights reserved. Further details on the songs used in the corpora are available in Appendices A and B.

[15] The primary coder for these data is a trained musician who began learning Japanese at age eight.

[16] Lexical strata were assigned according to the WWWJDIC database developed at Monash University (Breen 2000).

rare in the corpus and were therefore excluded, leaving only the three strata Yamato, Sino-Japanese, and Foreign. Instances of code-switching, in which complex phrases or entire lines were produced in English or other languages, were also excluded. Code-switching was defined in the present analysis as consisting of any non-Japanese lyrics that contained words from more than one lexical category (e.g., *it's my day* versus *shiro hige Santa Claus*: only the former would be considered code-switching).

In total, 845 tokens were coded in the translated Disney corpus and 469 in the native anime corpus.

4.4 *Results*

Overall frequencies for moraic versus syllabic settings for each of the four variables in the translated and native corpora are given in Figure 1. Each of the variables were found in both moraic and syllabic settings, ranging from coda nasal, which occurred more frequently in syllabic than moraic settings in the Disney corpus, to inter-voiceless *i/u*, which has the fewest syllabic settings. This is an expected result: unlike the other variables examined, a syllabic setting for inter-voiceless *i/u* is not possible in traditional singing style, in which the vowel remains voiced.

As predicted, for each of the variables, moraic settings are more frequent in the native corpus. Also as predicted, Foreign words are more likely to receive syllabic settings than Sino-Japanese and Yamato words, as shown in Figure 2.
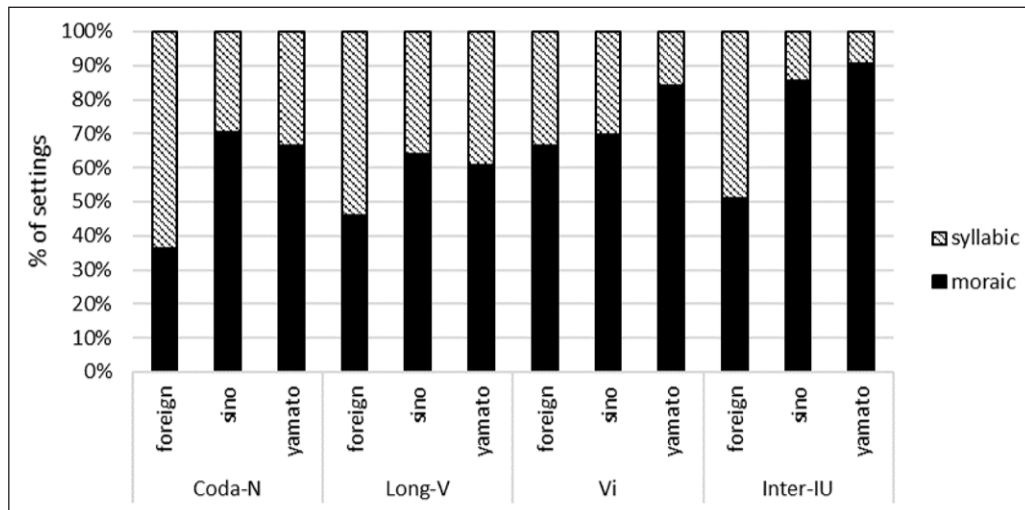
4.4.1 Modeling results

The contribution of the factors of corpus and lexical stratum to predicting syllabic versus moraic setting was evaluated for each variable using multivariate logistic regression modeling, which allows us to include control factors and to test for the independent effects of each factor of interest while controlling for the others. Models were fitted using the bayesglm() function in the arm R package (Gelman et al. 2013). No random effects were included because most of the data were represented by only one or two tokens of each word type.[17]

Figure 1: Proportion of syllabic and moraic settings for four variables in translated Disney versus native anime corpus.

---

[17] For the Coda N variable, 81% of word types have only one or two tokens in the data. For Long V, 82%; Vi, 85%; and inter-voiceless *i/u*, 88%. We tested a random effect of word in the Coda N data because it featured more clustering of word tokens (using lmer () in the lme4 R package), but found no changes in the direction of the main effects and little change in size of main effects between a model with random effect of word and the model reported in the main text.

Figure 2: Proportion of syllabic and moraic settings for four variables by lexical stratum for both corpora.

Model results predicting syllabic settings over moraic settings for each variable are given in Tables 7–10.

Controlling for all other factors, the effect of corpus (native anime versus translated Disney) is significant for all four dependent variables. As shown in Tables 7–10, Disney corpus tokens are much more likely to receive syllabic settings. Lexical stratum is also a reliable predictor of syllabic settings in all four models. For all four dependent variables investigated, Foreign words are significantly more likely to receive syllabic settings than native Yamato words. For three of the four variables—coda nasal, long vowel, and inter-voiceless *i/u*—we find no significant differences between the Sino-Japanese and Yamato words: special moras in Sino-Japanese words are not any more likely to be in a syllabic setting than Yamato special moras. The exception here is the case of Vi moras, for which we find that Sino-Japanese words pattern with Foreign words, making Yamato words significantly less likely than words from the other two strata to receive syllabic settings. A test of Foreign versus Sino-Japanese strata for Vi moras demonstrates no significant difference between the two non-Yamato strata ($\beta = -0.4677$, Std. error $= 0.5027$, $p = 0.352183$). The overall pattern here is consistent: foreign words are most likely to receive syllabic settings and Yamato words are most likely to receive moraic settings. In most cases, Sino-Japanese words do not differ from Yamato. We return to the issue of Vi cases in §6.

Some of the contextual linguistic factors tested for each variable were found to be significant predictors of moraic versus syllabic setting. For Vi sequences, occurring at the end of a word makes the sequence more likely to be set syllabically (e.g., *ippai* 'full'). A word-final location was not a significant factor, however, in the case of coda nasal or long vowels. Initial vowel was somewhat predictive of setting style for Vi, with *ai* significantly more likely to be set syllabically (25% syllabic) than *ei* (15% syllabic), but this difference was not significant for other Vi sequences when controlling for other factors, possibly due to small token counts for Vi sequences other than *ai*. The preference for *ai* as a syllable may be due to the large sonority gap between *a* and *i*, which makes it easier to perceive as a diphthong. It also may be preferred as a syllable due to its high frequency in Japanese, or perhaps its high frequency in Foreign words from English, in which /ay/ is the most frequent diphthong (Gimson 1980).

For *i/u* occurring between voiceless stops, syllabic setting was significantly more likely when the surrounding consonants consisted of one continuant and one stop, thus more

Table 7: Coda N logistic regression results (*** = *p* < 0.001, ** = *p* < 0.01, * = *p* < 0.05, . = *p* < 0.1).

Factor	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	−1.8852	0.3454	−5.46	<0.0001***
Corpus = Disney	1.6358	0.2991	5.47	<0.0001***
Strata = Foreign	1.2112	0.3619	3.35	0.0008***
Strata = Sino	−0.0771	0.3156	−0.24	0.8069
Word final = Yes	0.0410	0.2677	0.15	0.8783
N = 314; syllabic = 122; moraic = 192				

Table 8: Long V logistic regression results.

Factor	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	−1.6564	0.4807	−3.446	0.0005***
Corpus = Disney	1.3177	0.2724	4.83	<0.0001***
Strata = Foreign	0.7013	0.3155	2.223	0.026216*
Strata = Sino	−0.2815	0.2664	−1.057	0.290556
Word final = Yes	0.1513	0.2551	0.593	0.5531
Sonority	−0.1091	0.1873	−0.583	0.5601
N = 346; syllabic = 143; moraic = 203				

Table 9: Vi logistic regression results.

Factor	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	−3.272	0.528	−6.19	<0.0001***
Corpus = Disney	0.846	0.322	2.63	0.0086**
Strata = Foreign	1.229	0.523	2.35	0.0188*
Strata = Sino	0.713	0.304	2.34	0.0192*
Word final = Yes	0.251	0.319	0.79	0.4311
Is *ai*? = Yes	1.108	0.400	2.77	0.0056**
N = 328; syllabic = 71; moraic = 257				

Table 10: Inter-voiceless *i/u* logistic regression results.

Factor	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	−6.2567	0.8510	−7.35	<0.0001***
Corpus = Disney	1.5018	0.4760	3.16	0.0016**
Strata = Foreign	2.7436	0.5593	4.91	<0.0001***
Strata = Sino	0.3221	0.4995	0.64	0.5191
Good cluster? = Yes	3.6258	0.7193	5.04	<0.0001***
Devoiced vowel = [u]	−0.0956	0.4321	−0.22	0.8248
N = 326; syllabic = 53; moraic = 273				

easily forming a consonant cluster upon deletion of the intervening vowel (e.g., *suk*i → [ski] versus *kite* → [kte]). No difference was found in the behavior of the two vowels *i* and *u*. In addition to the syllabic versus moraic setting differences for inter-voiceless consonant *i/u*, we also examined the two options of moraic text-setting for the *i/u* vowels: voiced versus devoiced (e.g., [sɯ ki] or [sɯ̥ ki]). As discussed in §2.2.3, *i* and *u* are traditionally not devoiced in singing, as they are in speech. When an independent musical note is assigned to a mora containing *i/u* in a devoicing environment in the lyrics, a singer has the option of either maintaining voiced vowels as per the traditional singing style, or devoicing the *i/u* as in speech; this choice may be shaped by factors such as duration of the note in

addition to genre. In our data, we see a difference in devoicing between the two corpora of translated and native music. In the anime corpus, among the 131 *i/u* tokens that were set moraically, only 4.5% were devoiced, while in the Disney corpus, 17.6% of the 142 moraic tokens were devoiced ($\chi^2 = 10.227$, df = 1, $p = .0014$). Thus, even among those tokens set moraically, the performers in the Disney corpus use significantly less traditional singing style.

4.5 Discussion

4.5.1 Initial conclusions

This analysis has confirmed three key predictions: (1) syllabic settings occur to some degree in all strata of the Japanese lexicon, for all four variables examined (2) syllabic settings occur most frequently in the Foreign stratum and least frequently in the Yamato stratum, and (3) controlling for other linguistic factors, syllabic settings occur more frequently in the translated corpus than in the native corpus.

While no previous research has investigated translated songs, the data from the native corpus appears to be consistent with previous work: the native anime corpus contained approximately 80% moraic settings, in line with Tanaka's survey of popular songs, in which he found .79 and .69 notes per special mora in songs of the 80s and 90s, respectively (Tanaka 2000: 154). The data also support Tamaoka & Terao (2004) and others who have argued that coda nasal and long vowels are more likely to attach to the preceding mora than other special moras including the second vowel in two-vowel sequences; in these corpora, VV sequences other than Vi received no syllabic settings at all, and Vi sequences were less likely to receive syllabic settings than the former two variables.

Unlike previous studies of Japanese text-setting, which have focused exclusively on those moras traditionally identified as special moras, our study has also examined *i/u* in devoicing contexts. The occurrence of syllabic settings for this variable supports the claim that *i/u* in this context may be entirely deleted, resulting in the formation of a consonant cluster on a single syllable. Furthermore, this phenomenon was observed not only in Foreign words (in which case it might be dismissed as code-switching), but also in Sino-Japanese and Yamato words, as in (14a–b).

(14) a. *Yamato*
          ki-su **sh [i] te**
          kiss   do + IMP
          'kiss!'
          Original English lyrics: "Kiss the girl!"
          ("Kiss the Girl", 1989)

      b. *Sino-Japanese*
          do-ko ga   **s [u] te**-ki ka na-n-te sa
          where NOM   wonderful Q   like   even
          'even which part is wonderful'
          Original English lyrics: "Everyone's awed and inspired by you"
          ("Gaston", 1991)

The appearance of these consonant clusters in text-setting, as opposed to only in rapid, connected speech, supports the proposal that such clusters are not merely a phonetic byproduct of rapid speech but may be represented at some level in the phonological system (Kondo 2005).

The results of Study 1 undermine the claim that text-setting constitutes evidence for the mora being the only necessary prosodic unit in Japanese (cf. Labrune 2012). While

it is true that syllabic settings are relatively rare in native songs, when the text-setting system is stressed due to a restrictive translation context, Japanese lyricists are willing to use syllabic settings for as many as half the instances of special moras. This suggests that the use of predominantly moraic setting in Japanese songs is a *prescriptive norm* rather than reflective of underlying lack of syllabic structure. This places moraic text-setting in the same category as the idiosyncratic phonological behavior of Japanese singing style; in other words, just as *i/u* are traditionally not devoiced in singing, individual moras are traditionally assigned their own note in text-setting, but this does not then mean that *i/u* are not devoiced in spoken Japanese or that the syllabic unit does not exist.

   A comparison of two sets of translated lyrics further supports the characterization of moraic setting as a prescriptive norm. In 1997, Disney retranslated *The Little Mermaid* and rerecorded the songs using the same performers. While we do not know exactly what their objections were to the old translations, we can observe that the retranslated songs all either maintain or significantly increase the proportion of moraic settings in the lyrics. Most dramatically, the song "Part of Your World" jumps from 39% to 62% moraic settings. As illustrated in (15), while the first line of the original 1989 version managed to include syllabic settings for all four of our variables in a single line, the 1997 translation contains only moraic settings.

(15)   Comparison of the first lines of "Part of Your World" from *The Little Mermaid*
        (1998) versus (1997) translations (syllabic settings noted in bold)
        Original English lyric: "Look at this stuff! Isn't it neat?"
        a.   1989 translation
              go-**ran**-na-**sai**,   **s[u] te**-ki   de-**shoo**?
              see+HON+IMP   wonderful   COP+PRT
              'Look! Isn't it wonderful?'

        b.   1997 translation
              yo-ku   mi-te,   su-te-ki-ne?
              well   see+IMP   wonderful+PRT
              'Take a good look! Isn't it wonderful?'

One result of these new translations with higher proportions of moraic settings is that the songs have a more traditional feel, while the older versions seem more "speech-like" in style. Indeed, this stylistic difference raises some questions about the nature of the two corpora examined in Study 1, as discussed in the following section.

4.5.2 Stylistic concerns

We have proposed that, due to the operation of phonological processes that also operate in speech such as devoicing and coarticulation, syllabic text-setting results in a more speech-like style of Japanese than moraic setting. This creates a problem for the claim that the difference in moraicity between the translated Disney and native anime corpora stems from the pressures of translation itself. While the selection of songs in the anime corpus controlled for time period, which we know to be relevant in text-setting style, it failed to control entirely for genre. Specifically, the Disney songs are written in a modern musical theater style that is characterized by speech-like singing (Deer & Dal Vera 2008: 232; Spivey 2008: 483). Many of the songs in the corpus are speech-like to the extent that they actually alternate between speaking (which was excluded from the analysis) and singing, as in the song "Friend Like Me" from *Aladdin* (16).

(16)   Excerpt from "Friend Like Me" (1992) as performed in Japanese
       Sung:
       soo a-ri-ba-ba ni   wa yon-juu-nin   mo-no   to-u-zo-ku ga   i-ta
       yes Ali Baba   DAT TOP four+ten+person even+GEN thief   NOM exist+PST
       'yes Ali Baba had even 40 thieves'
       Original English lyrics: "Well Ali Baba had them forty thieves, Scheherazade had
       a thousand tales"

       Spoken:
       dakedo masutaa anta wa   motto rakkii
       but   master   2SG TOP more   lucky
       'but, master, you're luckier'
       Original English lyrics: "But master you in luck 'cause up your sleeves"

       Sung:
       da-re mo ka-na-i   wa shi-na-i
       who   also come true TOP do+NEG
       'for no one, [their dreams] come true'
       Original English lyrics: "You got a brand of magic never fails"

In contrast, although they are also songs drawn from children's animated films and television programs, the theme songs that comprise the anime corpus come from a range of popular music genres, none of which resemble musical theater. None of the songs in the anime corpus involved speech alternating with singing or use a speech-like singing style. Therefore, another possible account for the difference in moraicity between the Disney and anime corpora is that the Disney corpus employs syllabic text-setting as a stylistic device to give the songs a speech-like quality. Study 2 further investigates the role of style versus translation in syllabic text-setting.

5 Study 2

5.1 Introduction

Study 2 introduces a third corpus of translated Christmas songs, with the goal of teasing apart the roles of genre and translation in the frequency of syllabic and moraic settings. Christmas songs were selected for this second study because they offer a large corpus of translated songs in a relatively consistent, traditional musical genre with a range of lyricists. If the Christmas corpus contains the same rate of syllabic settings as the Disney corpus, this would be consistent with the increase in syllabic settings resulting from the restrictive translation context. On the other hand, if the Christmas corpus contains the same rate of syllabic settings as the native anime corpus, this would make the translation account less likely, and would support the proposal that the Disney corpus makes use of syllabic settings for stylistic purposes, to convey a speech-like quality.

5.2 Corpus and methodology

A corpus of 10 Christmas songs was selected (see Appendix B). When more than one translation was available for a song, only one translation was included based upon which audio version was most readily available. Unlike the two corpora of Study 1, which were restricted to songs of the late 1980s to mid-1990s, the Christmas corpus features a range of composition and translation dates, with dates of translation ranging from the late 19th to early 21st century. This allows us to compare the corpus to the native popular song corpus of Tanaka (2000), which found an increase in syllabic settings over the course of the 20th century.

Like the Disney songs, these Christmas songs consist of translations from English into Japanese.[18] One difference between the Disney and Christmas corpora, however, is that the translations in the Christmas corpus are not quite as close, thus presenting a less restrictive text-setting context. In the case of Disney songs, the translators were accountable to the Disney corporation and subject to the restrictions of audio dubbing, making it necessary to compose a very close translation. When translating Christmas songs, while there is a desire to convey the original sense of the song, the translators are working independently and have more leeway to alter the lyrics. Although we attempted to select songs that were closely translated, it was impossible to restrict the corpus to songs that were completely faithful to the original versions. Given the lower pressure to convey all of the information from the original songs, we therefore expect fewer syllabic settings compared to the Disney corpus. Examples (17a–c) provide instances of text-setting in the Christmas corpus.

(17)   a.   "O Holy Night" (1847, trans. 1909)
            i-za   ki-ke,   mi-**tsu**-ka-**i** u-ta-u
            now   listen + IMP angel   sing
            'Listen now! The angels sing'
            Original English lyrics: "Fall on your knees! Oh hear the angel voices!"

       b.   "Santa Claus is Coming to Town" (1934, trans. 1959)
            mi-n-na   de   de-ka-ke-yo-o
            everyone INS go out + PRT
            'Let's all go out together.'
            Original English lyrics: "He's making a list, he's checking it twice."

       c.   "I Saw Mommy Kissing Santa Claus" (1952, trans. 1962)
            de-mo so-no **san**-ta wa   pa-pa
            but   that   santa TOP dad
            'But that Santa is dad.'
            Original English lyrics: "Mommy kissing Santa Claus last night."
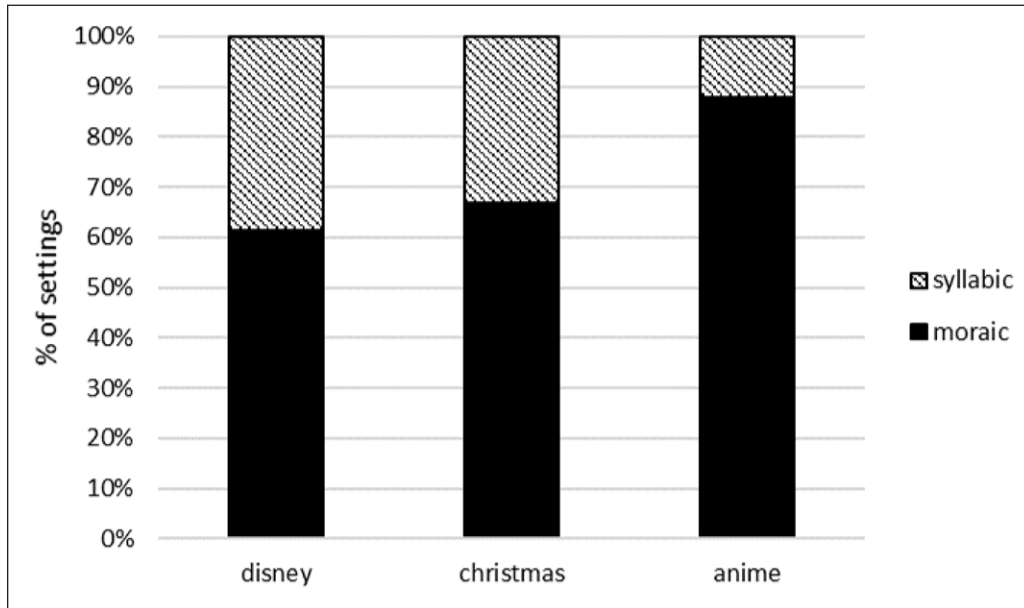
All tokens of the four variables described in §3 were coded by the first author in the same manner as in Study 1. Two hundred and fifteen tokens were coded in total.

5.3 *Results and discussion*

Figure 3 indicates the syllabic versus moraic text-setting rate for each of the three corpora. Based on these overall rates, it appears that the two translated corpora are patterning together, with more syllabic settings than in the native corpus. Further statistical analysis is needed, however, to control for various factors and examine the behavior of each variable.

Models for each of the four variables in all three corpora are given in Tables 11–14. As shown in Tables 11–14, for variables Coda N and Long V, the Christmas corpus is significantly different from the anime corpus and not significantly different from the Disney corpus. For factors Vi and inter-voiceless *i/u*, however, the Christmas corpus is significantly different from Disney and not significantly different from anime. In the case of Vi, the Christmas corpus is in fact marginally more moraic than the anime corpus, as shown in

---

[18] As specified in Appendix B, two of the Christmas corpus songs were not originally written in English: "Silent Night" was originally composed in German and "O Holy Night" was originally French. This point is of little importance to the analysis, but in all likelihood the songs were translated into Japanese based on the English versions as part of a larger set of English hymns.

Figure 3: Overall syllabic versus moraic text-setting rates for each corpus.

Table 11: Coda N logistic regression results (*** = *p* < 0.001, ** = *p* < 0.01, * = *p* < 0.05, . = *p* < 0.1).

Factor	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	−0.204	0.343	−0.59	0.55
Corpus = Anime	−1.944	0.141	−4.69	<0.0001***
Corpus = Disney	−0.232	0.355	−0.65	0.51
Strata = Foreign	1.784	0.323	5.51	<0.0001***
Strata = Sino	0.210	0.299	0.70	0.48
Word final = Yes	−0.135	0.253	−0.53	0.59
N = 368; syllabic = 158; moraic = 210				

Table 12: Long V logistic regression results.

Factor	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	−0.579	0.509	−1.136	0.26
Corpus = Anime	−1.271	0.356	−3.56	0.0003***
Corpus = Disney	−0.0939	0.306	−0.307	0.76
Strata = Foreign	1.108	0.288	3.84	0.0002***
Strata = Sino	−0.181	0.253	−0.714	0.48
Word final = Yes	0.260	0.241	1.08	0.28
Sonority	0.095	0.177	0.54	0.588
N = 414; syllabic = 179; moraic = 235				

Table 13: Vi logistic regression results.

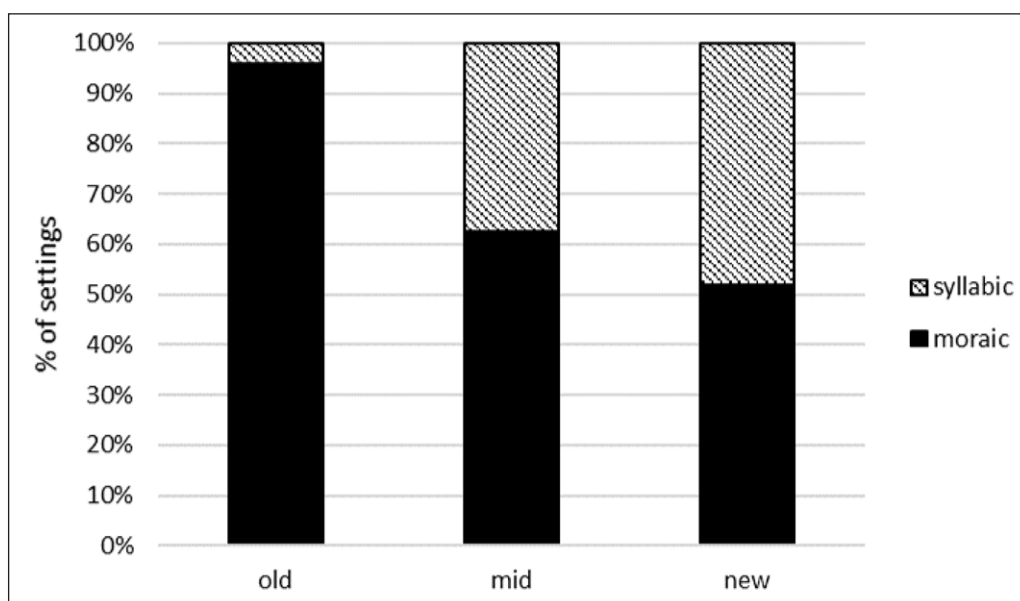| Factor	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	−4.385	0.791	−5.55	<0.0001***
Corpus = Anime	1.227	0.74	1.66	0.0971.
Corpus = Disney	2.103	0.713	2.95	0.0032**
Strata = Foreign	1.152	0.515	2.24	0.0253*
Strata = Sino	0.662	0.300	2.21	0.0272*
Word final = Yes	0.201	0.312	0.65	0.5182
Is *ai*? = Yes	1.002	0.382	2.63	0.0087**
N = 381; syllabic = 72; moraic = 309				

Table 13. The fixed effects that significantly predicted text-setting style in each model in Study 1, *ai* for Vi and good cluster for inter-voiceless *i/u*, remain significant in Study 2.

Interpreting these findings presents some difficulties, as two of the variables (Coda N and Long V) are consistent with the hypothesis that translation status accounts for increased syllabic text-setting, while the other two (Vi and Inter-voiceless i/u) point to genre. In light of Tanaka (2000), however, we must also consider the factor of text-setting changes over time. The songs of the Christmas corpus can be divided by translation date into an "old" group, consisting of the three songs translated in the late 19th and early 20th century, a "mid" group, the five songs translated in the mid-20th century, and a "new" group, the two songs translated in the late 20th and early 21st century (see Appendix B for translation dates for each song). Figure 4 indicates the proportion of moraic and syllabic settings for each group; consistent with Tanaka's findings, a significant shift toward syllabic settings can be observed between the old and mid group ($\chi^2 = 15.544$, df = 1, $p < .0001$). This shift corresponds to date of translation into Japanese, not to date of original composition: the song "We Wish You A Merry Christmas," for example, is a traditional carol dating back to the 16th century, but the version included in the corpus was translated in the 21st century and contains a high number of syllabic settings. This change in the Christmas corpus over time suggests that shifts in stylistic norms for text-setting, rather than musical genre *per se*, are responsible for the differences observed.

Table 14: Inter-voiceless *i/u* logistic regression results.

Factor	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	−6.5043	0.9753	−6.67	<0.0001***
Corpus = Anime	0.0291	0.7444	0.04	0.969
Corpus = Disney	1.5882	0.6677	2.38	0.017*
Strata = Foreign	2.9482	0.5501	5.36	<0.0001***
Strata = Sino	0.5330	0.4868	1.09	0.274
Good cluster? = Yes	3.7177	0.7254	5.12	<0.0001***
Devoiced vowel = [u]	−0.1214	0.4214	−0.29	0.773
N = 375; syllabic = 55; moraic = 320				

Figure 4: Proportion of moraic and syllabic settings for Christmas corpus songs by time period translated.

While the songs in the oldest translation group have a much higher rate of moraic settings, the Christmas songs translated from the mid-20th century onward have an overall syllabic setting rate comparable to the songs of the Disney corpus. When these older songs are removed from the analyses, the patterns for Coda N, Long V, and Vi remain the same, but for Inter-voiceless *i/u* the Christmas corpus no longer significantly differs from either the Disney or anime corpus. In other words, when changes over time are accounted for, the Christmas corpus does not significantly differ from the Disney corpus for three of the four variables. We therefore conclude that the higher rate of syllabic settings in the Disney corpus cannot entirely be attributed to its speech-like style; rather, controlling for genre and stylistic changes over time, the restrictions of translation consistently result in higher rates of syllabic settings.

6 Discussion

The two studies presented above provide considerable evidence that the syllable plays a key role in Japanese text-setting. We have also observed that syllables are most likely in translated texts and within Foreign words. Finally, in Study 2, evidence was found to support previous suggestions that syllabic settings have become more frequent over time.

The frequency of syllabic settings in translated text is accounted for by the restrictive nature of the translation task and by the large information density gap between Japanese and many other languages, including English. In other words, because it takes more prosodic units to express a given meaning in Japanese, it is often helpful to collapse moraic units into syllables. Indeed, as shown in the examples from the translated corpora, even when syllabic settings are used, many of the semantic details of the original English lyrics are lost in the Japanese versions due to space constraints. In "Friend Like Me" from *Aladdin*, for example, the tales of Scheherazade are entirely lost in the effort to convey the first line about Ali Baba's forty thieves (see ex. 16). This does not mean all translated songs must contain syllabic settings; when the prescriptive constraint discouraging the use of syllabic settings is strong, as it was with the earliest entries in the Christmas corpus, translated songs use more moraic settings at the expense of content.

These studies have shown that, controlling for other factors, the use of moraic versus syllabic setting is predicted by lexical stratum. For all four variables, Foreign words were more likely to use syllabic settings than Yamato words, and for all variables but Vi, Sino-Japanese words patterned together with Yamato words. In the case of Vi, however, Sino-Japanese patterned with Foreign words and were more likely than Yamato words to be set syllabically. This last result poses a hurdle to accounting for differences in the strata by appealing to individuals' knowledge of the original language; if only Foreign words patterned differently, the most likely explanation would be that the syllabic unit is more prominent in cases where a syllabic segmentation is similar to the word in the original language. For instance, the word *misuterii* ('mystery') is more faithful to the English when produced as "mi-ste-ri" with the inter-voiceless *u* deleted and the long vowel produced as a single unit, as opposed to the moraic setting "mi-su-te-ri-i." This explanation seems even more probable for songs in which code-switching from Japanese to English is frequent; if the lyricist is already using English for some lines of the song, it is not a big leap to use English-style segmentation for Foreign words in the Japanese portions of the song.

In the case of Sino-Japanese words, however, knowledge of Chinese is not a reasonable explanation for syllabic segmentation. Although Chinese uses the syllable as its basic prosodic unit, most Japanese speakers are not familiar with Chinese, and Sino-Japanese words are generally quite different from their modern Mandarin Chinese cognates, as these borrowings took place many centuries ago. Moreover, while it is possible that the syllable is more prominent in the underlying prosodic structure of Sino-Japanese words, which

would account for differences without relying on individuals' knowledge of Chinese, such a model would predict consistent differences that were not observed in the other three variables; for example, coda nasal should have been syllabically segmented more often in Sino-Japanese words than in Yamato words. Instead, it is most likely that some additional linguistic contextual factor that was not accounted for in the model is responsible for the higher frequency of syllabic settings for Vi in the case of Sino-Japanese words. One culprit we might propose is a content-function word distinction: Vi only occurs within content morphemes in Sino-Japanese (e.g., *meirei* 'command'), while Vi frequently appears as part of inflectional Yamato endings (e.g., *ikitai* 'want to go'). However, removing the inflectional tokens from the Yamato data only increases the moraicity of tokens in this stratum, magnifying the difference between Yamato and Sino-Japanese. Another possible account is that Japanese in fact has two separate underlying representations for Vi: (1) a *Vi* diphthong and (2) a *V-i* sequence, the first of which is more likely to be set syllabically. If we assume that the Vi diphthong has developed due to language contact, it would make sense that it is represented more strongly among the Foreign and Sino-Japanese strata. While the Vi data is intriguing, given the lack of consistent differences between Sino-Japanese and Yamato words, there is no strong evidence here that the syllable is more salient in the prosodic structure of Foreign and Sino-Japanese lexical strata than in the Yamato stratum. The most likely explanation, given at the overall pattern in the data, is that individuals' knowledge of English is responsible for the strata differences observed in these corpora.

The linguistic factors found to encourage syllabic setting provide some evidence for avoidance of low-sonority units in text-setting. The mora of lowest sonority, coda nasal, was the most frequent variable to be joined to its preceding mora. Additionally, the fact that only Vi of the vowel sequences was found in a syllabic setting suggests a desire to avoid placing low-sonority *i* on its own. It was also found that instances of *i/u* devoicing when the mora containing the devoiced segment was assigned its own note were quite rare, with performers preferring to voice the vowel when presented with a moraic setting in 95.5%, 82.4%, and 97.8% of cases in the anime, Disney, and Christmas corpora respectively. Thus, the frequencies of the three possible realizations for inter-voiceless *i/u* follow the sonority scale: moraic voiced (*su-ki*) > syllabic (*ski*) > moraic devoiced (*s[u]-ki*). On the other hand, no effect of sonority was detected for the setting of long vowels. Unlike the other linguistic variables, the syllabic setting of long vowels creates ambiguity, as it is no longer possible to distinguish between long and short vowels when long vowels are set on a single note. Thus, it may be the case that the setting of long vowels is influenced by the predictability of the word in a particular context and whether there is an existing minimal pair containing a short vowel (e.g., *tori* 'bird' versus *toori* 'street'). We leave investigation of this possibility to future work.

These studies also examined the function of text-setting as a stylistic element; in particular, we evaluated the possibility that syllabic settings were used to create a speech-like quality in the songs of the Disney corpus. When broader changes in text-setting style over time were controlled for, however, there was no statistical difference between the text-setting of the Disney corpus and the Christmas corpus in three of the four variables. This finding indicates that the role of translation overwhelms genre differences, at least in the case of these two corpora. Study 2 also confirmed the finding of Tanaka (2000) that syllabic settings have increased over time, with a significant change between the earliest translated Christmas songs and those translated after the mid-20th century. It does not necessarily follow, however, that the syllable has become more prominent in Japanese prosodic structure over time. The change in text-setting style may instead be part of a broader shift in Japanese singing style; similarly, other idiosyncratic features of sung Japanese, such as voicing of *i/u,* appear to be gradually decreasing. These shifts may be

characterized as a move toward singing in a style that is more similar to speech, reflecting a general desire for more naturalistic singing styles in modern music.

At the same time, there are certain suggestions in the data that the syllable may be playing a broader role in Japanese. Specifically, the patterns of syllabic setting found in these data hint that the range of possible syllable structures may be in the process of expanding, perhaps due to contact with English. We suggest that CVCV sequences in which one vowel has the potential to be devoiced, such as for inter-voiceless *i/u* (but also for word-final vowel deletion, not examined here), have the potential to be reanalyzed as CCV or CVC and treated as syllables. This would represent a dramatic shift in the phonotactics of Japanese, but equally significant changes have occurred in Japanese historically due to contact, such as the development of coda nasals. Corresponding segmental phonological changes in Japanese due to English contact have already been observed, such as the development of coronal obstruents before *i* (Crawford 2008). As Japanese popular songs contain frequent code-switching into English, it may be that lyricists are among the early adopters of these new syllable forms. Future work might examine the extent to which units of this type are accepted among speakers of Japanese with varying levels of English contact.

This potential expansion of the syllable also calls into question the utility of the class of special moras. As observed by Itô and Mester, the notion of special versus regular mora appears to "recapitulate syllable theory within a network of assumptions that are specific to Japanese" (Itô & Mester 2015: 371). The core class of weak moras does not generally include CV moras containing high vowels in devoicing environments: while they could be simply added to the list of special moras, these syllables seem like an odd fit in this category, as they contain two underlying segments. Maintaining the special mora distinction in light of inter-voiceless *i/u* deletion also necessitates altering the proposal that heavy syllables contain a regular mora followed by a special mora, because the order of moras is reversed in this case. It increasingly appears that the gains of identifying a separate class of special moras are not worth the cost of positing a category unique to Japanese.

7 Conclusion

While previous analyses describe Japanese text-setting as moraic, data from translated and native songs confirm that the syllable plays a prominent role in Japanese text-setting. Coda nasals, long vowels, Vi sequences, and *i/u* in devoicing environments were all found to vary between moraic and syllabic settings. Setting choice was partially predicted by the avoidance of low sonority units as well as several other contextual factors such as preference for articulatory ease. Syllabic settings occurred in all strata of the Japanese lexicon, but were most frequent in Foreign words, most likely as a result of knowledge of the English source word rather than underlying differences in prosodic structure among the strata. Translated songs consistently used more syllabic settings than native songs; this difference stems from the restrictions of translation as well as the relatively low information density of Japanese.

The high frequency of syllabic settings in translated text suggests that the preference for moraic settings in Japanese stems from a prescriptive norm rather than a lack of syllabic structure in Japanese prosody. This prescriptive style is part of a set of idiosyncratic phonological practices in traditional sung Japanese that distinguish it from speech, such as the splitting of long vowels into two short segments and the voicing of *i/u* in devoicing environments. When the text-setting system is stressed due to translation, this style breaks down and more syllabic settings are introduced. There is no evidence to suggest that purely moraic settings sound better to Japanese listeners; in fact, there is evidence to the contrary, that such settings can be perceived as unnatural (e.g., Manabe 2009). Thus, the increase in syllabic settings over time observed in these data do not necessarily

indicate an increasing role of the syllable in Japanese prosodic structure but may simply reflect a change in text-setting style as sung Japanese shifts toward spoken Japanese, in which the syllable has always played a role. Nonetheless, the occurrence of syllabic settings in cases that have not been described as possible syllables in previous literature, as in the deletion of *i/u*, suggests that the salience and scope of the syllable in Japanese may be on the rise.

Abbreviations

ACC = accusative, COP = copula, DAT = dative, HON = honorific, INS = instrumental, NOM = nominative, PST = past, PRT = discourse particle, Q = question marker, SG = singular, TOP = topic

Additional Files

The additional files for this article can be found as follows:

- **Appendix A.** Disney and anime song corpora details. DOI: https://doi.org/10.5334/gjgl.355.s1
- **Appendix B.** Christmas song corpus details. DOI: https://doi.org/10.5334/gjgl.355.s2

Acknowledgements

Competing Interests

References

Arai, Takayuki. 1999. A case study of spontaneous speech in Japanese. In *Proceedings of 14th ICPhS*, 615–618. San Francisco, CA.

Arai, Takayuki & Steven Greenberg. 1997. The temporal properties of spoken Japanese are similar to those of English. In *Proceedings of Eurospeech*, 1011–1114. Rhodes.

Baker, Mona & Brano Hochel. 2001. Dubbing. In Mona Baker (ed.), *Routledge Encyclopedia of Translation Studies*, 74–76. New York: Routledge.

Beckman, Mary. 1982. Segment duration and the 'mora' in Japanese. *Phonetica* 39. 113–135. DOI: https://doi.org/10.1159/000261655

Bekku, Sadanori. 1977. 日本語のリズム:四拍子文化論 *[Japanese rhythm: The four-beat theory of culture]*. Tokyo: Kodansha.

Blum, Shoshana & E. A. Levenston. 1978. Universals of lexical simplification. *Language Learning* 28(2). 399–415. DOI: https://doi.org/10.1111/j.1467-1770.1978.tb00143.x

Breen, Jim. 2000. WWWJDIC: Online Japanese dictionary service. http://www.edrdg.org/cgi-bin/wwwjdic/wwwjdic?1C.

Calder, Jeremy. 2013. Theories of syllabification in Nuxalk: Hints from text-setting. *Paper presented at the annual meeting of the Linguistic Society of America*, Boston, MA.

Chesterman, Andrew. 2004. Beyond the particular. In *Translation Universals: Do they exist?*, 33–50. Amsterdam and Philadelphia, PA: John Benjamins. DOI: https://doi.org/10.1075/btl.48.04che

Crawford, Clifford. 2008. An evolutionary account of loanword-induced sound change in Japanese. *Proceedings of the 31st Annual Penn Linguistics Colloqium, University of Pennsylvania Working Papers in Linguistics* 14(1). 56–70.

Deer, Joe & Rocco Dal Vera. 2008. *Acting in musical theatre: A comprehensive course.* New York: Routledge.

Dell, François. 2011. Singing in Tashlhiyt Berber, a language that allows vowel-less syllables. In Charles E. Cairns & Eric Raimy (eds.), *Handbook of the Syllable*, 173–193. Leiden: Brill.

Gelman, Andrew, Yu-Sung Su, Masano Yajima, Jennifer Hill, Maria Grazia Pittau, Jouni Kerman & Tian Zheng. 2013. Package 'arm'. http://CRAN.R-project.org/package = arm.

Gimson, A. C. 1980. *An introduction to the pronunciation of English.* 3rd edition. London: Edward Arnold.

Golston, Chris & Tomas Riad. 2005. The phonology of Greek lyric meter. *Journal of Linguistics* 41. 77–155. DOI: https://doi.org/10.1017/S0022226704003068

Halle, John & Fred Lerdahl. 1993. A generative text-setting model. *Current Musicology* 55. 3–21.

Halle, Morris & S. J. Keyser. 1971. *English stress: Its form, its growth, and its role in verse.* New York: Harper & Row.

Han, Mieko S. 1962. The feature of duration in Japanese. *Onsei no kenkyuu* 10. 65–80.

Hanson, Kristin & Paul Kiparsky. 1996. A parametric theory of poetic meter. *Language* 72(2). 287–335. DOI: https://doi.org/10.2307/416652

Hayes, Bruce. 1988. Metrics and phonological theory. In *Linguistics: The Cambridge survey* 2. 220–249. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511621055.013

Hayes, Bruce. 2009. Textsetting as constraint conflict. In *Towards a typology of poetic forms: From language to metrics and beyond*, 43–62. Amsterdam and Philadelphia, PA: John Benjamins.

Hayes, Bruce & Tami Swiger. 2008. Two Japanese children's songs. Ms.: University of California, LA.

Hirata, Yukari. 2004. Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics* 32(4). 565–589. DOI: https://doi.org/10.1016/j.wocn.2004.02.004

Hyman, Larry M. 1983. Are there syllables in Gokana. In Jonathan Kaye, Hilda Koopman, Dominique Sportiche & Andre Dugas (eds.), *Current approaches to African linguistics*, 171–179. Dordrecht: Foris.

Hyman, Larry M. 2003. *A theory of phonological weight.* Stanford, CA: Center for the Study of Language and Information.

Hyman, Larry M. 2008. Universals in phonology. *The Linguistic Review* 25(1–2*)*. 83–137. DOI: https://doi.org/10.1515/TLIR.2008.003

Hyman, Larry M. 2011. Does Gokana really have no syllables? Or: what's so great about being universal? *Phonology* 28(1). 55–85. DOI: http://www.jstor.org/stable/41303154

Inaba, Seiichiro. 1998. Moras, syllables, and feet in Japanese. *Language, information and computation*, 106–117.

Inagaki, Kayoko, Giyoo Hatano & Takashi Otake. 2000. The effect of kana literacy acquisition on the speech segmentation unit used by Japanese young children. *Journal of Experimental Child Psychology* 75(1). 70–91. DOI: https://doi.org/10.1006/jecp.1999.2523

Irwin, Mark. 2011. *Loanwords in Japanese.* Amsterdam and Philadelphia, PA: John Benjamins. DOI: https://doi.org/10.1075/slcs.125

Itô, Junko & Armin Mester. 1995. Japanese phonology. In *Japanese phonology. The handbook of phonological theory*, 818–838. Oxford: Wiley-Blackwell.

Itô, Junko & Armin Mester. 1996. Stem and word in Sino-Japanese. In *Phonological structure and language processing: Cross-linguistic studies*, 13–44. Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110815825.13

Itô, Junko & Armin Mester. 1999. The structure of the phonological lexicon. In *The handbook of Japanese linguistics*, 62–100. Malden, MA and Oxford: Blackwell Publishers.

Itô, Junko & Armin Mester. 2015. Word formation and phonological processes. In Haruo Kubozono (ed.), *The handbook of Japanese phonetics and phonology*, 363–395. Berlin: Walter de Gruyter.

Katz, Jonah. 2015. Hip-hop rhymes reiterate phonological typology. *Lingua* 160. 54–74. DOI: https://doi.org/10.1016/j.lingua.2015.03.005

Kawahara, Shigeto. 2016. Japanese has syllables: A reply to Labrune (2012). *Phonology* 33(1). 169–194. DOI: https://doi.org/10.1017/S0952675716000063

Kawahara, Shigeto, Hajime Ono & Kiyoshi Sudo. 2006. Consonant co-occurrence restrictions in Yamato Japanese. *Japanese/Korean Linguistics* 14. 27–38.

Kawahara, Shigeto & Jason Shaw. 2016. Interaction of high vowel devoicing and syllabification. *Paper presented at Syllables and Prosody Workshop,* NINJAL, 13 October 2016. Tokyo, Japan.

Kiparsky, Paul. 1973. The role of linguistics in a theory of poetry. *Daedalus* 102(3). 231–244.

Kiparsky, Paul. 1977. The rhythmic struture of English verse. *Linguistic Inquiry* 8(2). 189–247. DOI: http://www.jstor.org/stable/4177986

Kondo, Mariko. 2005. Syllable structure and its acoustic effects on vowels in devoicing environments. In *Voicing in Japanese*, 229–246. Leiden: Walter de Gruyter. DOI: https://doi.org/10.1515/9783110197686.2.229

Kubozono, Haruo. 1989. The mora and syllable structure in Japanese: Evidence from speech errors. *Language and Speech* 32(3). 249–278. DOI: https://doi.org/10.1177/002383098903200304

Kubozono, Haruo. 1996a. Speech segmentation and phonological structure. In *Phonological structure and language processing: Cross-linguistic studies*, 1–12. Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110815825.77

Kubozono, Haruo. 1996b. Syllable and accent in Japanese: Evidence from loanword accentuation. *Onsei Kaihoo* 211. 71–82.

Kubozono, Haruo. 1999a. Mora and syllable. In *The handbook of Japanese linguistics*, 31–61. Oxford: Blackwell.

Kubozono, Haruo. 1999b. 歌謡におけるモーラと音節 [The syllable and mora in song.] In 文法と音声 [Speech and grammar] 2. 241–260. 音声文法研究会 [Speech and Grammar Study Group].

Kubozono, Haruo. 2003. The syllable as a unit of prosodic organization in Japanese. In *The syllable in Optimality Theory*, 99–122. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511497926.005

Kubozono, Haruo. 2015. Diphthongs and vowel coalescence. In *Handbook of Japanese phonetics and phonology*, 169–194. Berlin: Walter de Gruyter GmbH. DOI: https://doi.org/10.1515/9781614511984.215

Kubozono, Haruo, Junko Itô & Armin Mester. 2008. Consonant gemination in Japanese loanword phonology. In *Proceedings of the 18th International Congress of Linguistics*, 953–973. Seoul: The Linguistic Society of Korea.

Labrune, Laurence. 2012. Questioning the universality of the syllable: Evidence from Japanese. *Phonology* 29(1). 113–152. DOI: https://doi.org/10.1017/S095267571200005X

Lerdahl, Fred & Ray S. Jackendoff. 1983. *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Loveday, Leo J. 1996. *Language contact and Japan: A sociolinguistic history*. New York: Oxford University Press.

Machida, Shoichi. 1988. 日本語の基本的なリズムの設定方法とリズム・パターン [On a way of describing Japanese rhythm and the typical rhythmic pattern of Japanese]. *Sophia Linguistica* 25. 35–43.

Maekawa, Kikuo & Hideaki Kikuchi. 2005. Corpus-based analysis of vowel devoicing in spontaenous Japanese: An interim report. In *Voicing in Japanese*, 205–228. Berlin: Mouton de Gruyter.

Manabe, Noriko. 2009. *Western music in Japan: The evolution of styles in children's songs, hip-hop, and other genres*. New York: City University of New York dissertation.

Martin, Samuel E. 1952. Morphophonemics of standard colloquial Japanese. *Language* 28(3). 7–115. DOI: https://doi.org/10.2307/522176

Matsui, Michinao F. To appear. (Translated by Shigeto Kawahara). On the input information of the C/D model for vowel devoicing in Japanese. *Journal of the Phonetic Society of Japan* 21(1).

Matsuzaki, Hiroshi. 1994. 和語、漢語、外来語の語形と特殊拍の音配列上の制約 [Frequencies of the Japanese special sounds in the lexical categories of wago, kango and gairaigo.] 東北大．学文学部日本語学科論集 [Journal of the Department of Japanese, Tohoku University] 4. 75–86.

McPherson, Laura. 2014. A maxent model of tone-tune association in Tommo So songs. *Paper presented at the 88th meeting of the Linguistic Society of America*, Minneapolis, MN.

Moreton, Elliott & Shigeaki Amano. 1999. Phonotactics in the perception of Japanese vowel length: Evidence for long-distance dependencies. In *Proceedings of the 6th European Conference on Speech Communication and Technology*, 2679–2682. Budapest.

Nakamura, Miyoko & Regine Kolinsky. 2014. Multiple functional units in the preattentive segmentation of speech in Japanese: Evidence from word illusions. *Language and Speech* 57. 513–543. DOI: https://doi.org/10.1177/0023830913508077

Noel, Patrizia A. H. 2010. Layers of versification in Beowulf. *Anglia-Zeitschrift für englische Philologie* 127(2). 238–260.

Ohta, Satoshi. 1989. 日英語音節構造管見 [Remarks on English and Japanese syllable structure.] 英語と英米文学 [English and English-American Literature] 24. 1–11.

Ohta, Satoshi. 1991. Syllable and mora geometry in Japanese. *Tsukaba English Studies* 10. 157–181.

Otake, Takashi. 1992. 日本語音声の知覚上の分節の単位：音節とモーラ [Units of phonological segmentation in the perception of Japanese words—syllables and mora.] In 日本語のモーラと音節構造に関する総合的研究 [A systematic study of the phonological structure of mora and syllable] 1. 38–47.

Otake, Takashi. 2015. Mora and mora-timing. In Haruo Kubozono (ed.), *Handbook of Japanese phonetics and phonology*, 493–523. Berlin: Walter de Gruyter.

Otake, Takashi, Giyoo Hatano, Anne Cutler & Jacques Mehler. 1993. Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language* 32(2). 258–278. DOI: https://doi.org/10.1006/jmla.1993.1014

Otake, Takashi, Kiyoko Yoneyama, Anne Cutler & Arie van der Lugt. 1996. The representation of Japanese moraic nasals. *The Journal of the Acousical Society of America* 100(6). 3831–3842. DOI: https://doi.org/10.1121/1.417239

Pápai, Vilma. 2004. A universal of translated text? In *Translation universals: Do they exist?*, 143–164. Amsterdam and Philadelphia, PA: John Benjamins.

Pellegrino, François, Christophe Coupé & Egidio Marsico. 2011. Across-language perspective on speech information rate. *Language* 87(3). 539–558. DOI: https://doi.org/10.1353/lan.2011.0057

Port, Robert F., Johnathan Dalby & Michael O'Dell. 1987. Evidence for mora timing in Japanese. *The Journal of the Acoustical Society of America* 81. 1574–1585. DOI: https://doi.org/10.1121/1.394510

Poser, Bill. 1990. Evidence for foot structure in Japanese. *Language* 66. 78–105. DOI: https://doi.org/10.1353/lan.1990.0031

Schaffner, Christina & Beverly Adab. 2001. The idea of the hybrid text in translation: Contact as conflict. *Across Languages and Cultures* 2(2). 167–180. DOI: https://doi.org/10.1556/Acr.2.2001.2.1

Schuh, Russell G. 1994. Text and performance in Hausa metrics. In *25th Annual Conference on African Linguistics*. New Brunswick, NJ: Rutgers University.

Selkirk, Elizabeth. 1982. The syllable. In *The structure of Phonological Representations* 2. 337–383. Dordrecht: Foris.

Shibatani, Masayoshi. 1990. *The Languages of Japan*. Cambridge: Cambridge University Press.

Shih, Stephanie. 2008. Text-setting: A (musical) analogy to poetic meter. In *The linguistics of language arts: new research programs*. Berkeley, CA: University of California, Berkeley.

Spivey, Norman. 2008. Musical theater singing, let's talk part 1: On the relationship of speech and singing. *Journal of Singing* 64(4). 483–489.

Sugitou, Miyoko & Yakuso Sakai. 1999. わらべうたのリズムと音節 [Rhythm and syllables in children's songs.] In 文法と音声 [Speech and Grammar] 2. 291–306. Tokyo: Kurosio Publishers.

Sui, Yanyan. 2013. Phonological and phonetic evidence for trochaic metrical structure in Standard Chinese. Pennsylvania, PA: University of Pennsylvania Dissertation.

Tajima, Keiichi. 2011. Do native-language loanwords affect second-language speech perception? In *Proceedings of ICPhS XVII*, 1938–1941.

Tamaoka, Katsuo & Yasushi Terao. 2004. Mora or syllable? Which unit do Japanese use in naming visually presented stimuli? *Applied Psycholinguistics* 25. 1–27. DOI: https://doi.org/10.1017/S0142716404001018

Tanaka, Shin'ichi. 2000. 日本歌謡曲におけるリズムの変遷 [The changing rhythm of Japanese popular songs.] In *Proceedings of the 14th Meeting of the Japan Phonetics Society*, 153–158.

Tanaka, Shin'ichi. 2012. Syllable neutralization and prosodic unit in Japanese Senryu poems. *Paper presented at Metrics, Music and Mind in Rome*, Italy.

Tranter, Nicolas. 2012. Classical Japanese. In *The languages of Japan and Korea*, 212–245 New York: Routledge.

Trosborg, Anna. 1997. Translating hybrid political texts. In *Text typology and translation*, 145–158. Amsterdam and Philadelphia, PA: John Benjamins. DOI: https://doi.org/10.1075/btl.26.12tro

Tsuchida, Ayako. 2001. Japanese vowel devoicing: Cases of consecutive devoicing environments. *Journal of East Asian Linguistics* 10(3). 225–245. DOI: https://doi.org/10.1023/A:1011221225072

Ueyama, Motoko. 2003. Awareness of L2 syllable structures: The case of L2 Japanese and L2 English. *Journal of the Phonetic Society of Japan* 7(2). 84–100.

Vance, Timothy J. 1987. *An introduction to Japanese phonology*. Albany, NY: State University of New York Press.

Warachananan, Pannapa & Dutsadee Roongrattanakool. 2015. A study of translation strategies in the translation of songs in Walt Disney's animated feature films into Thai versions. In *National and international conferenceinterdisciplinary research for local development sustainability*, 75–84.

Warner, Natasha & Takayuki Arai. 2000. Japanese mora-timing: A review. *Phonetica* 58. 1–25. DOI: https://doi.org/10.1159/000028486

Warner, Natasha & Takayuki Arai. 2001. The role of the mora in the timing of spontaneous Japanese speech. *The Journal of the Acoustical Society of America* 109. 1144–1156. DOI: https://doi.org/10.1121/1.1344156

Yu, Haikuo. 2013. English-Chinese film translation in China. *Journal of Translation* 9(2). 55–65.

Yung, Bell. 1991. The relationship of text and tune in Chinese opera. In *Musical, language, speech, and brain*, 408–418. London: Macmillan. DOI: https://doi.org/10.1007/978-1-349-12670-5_37