国立国語研究所論集 (NINJAL Research Papers) 7: 93–109 (2014)
ISSN: 2186-134X print/2186-1358 online

Rendaku Across Duplicate Moras

Mark IRWIN
Yamagata University / Project Collaborator, NINJAL

Abstract
Persistent claims have been made in the rendaku literature concerning the occurrence of two identical, successive moras. When two identical fricative moras occur in succession in tautomorphemic position (e.g. *kare+**susu**ki* 'withered plume grass'), it is claimed rendaku is blocked. Further, if any two identical moras occur in succession in heteromorphemic position (e.g. *ki**zu**+**tu**keru* 'scar, wound'), then it is claimed rendaku is severely restricted. Neither of these claims has been corroborated by statistical evidence and must be placed in the category of 'linguistic urban myth'. In this paper, I will demonstrate by means of a large database that these claims are manifestly false and must be rejected.*

Key words: rendaku, mora, dissimilation, rendaku database, linguistic urban myths

1. Preliminaries

Rendaku, or sequential voicing, is a well-known morphophonological phenomenon found in Japanese, whereby the initial voiceless obstruent (*k*, *s*, *t*, *h*)[1] of a non-initial element (E_2) in a compound may be voiced,[2] as in:

(1) k ~ g: *usu* + ***k**urai* > *usu.**g**urai*
thin, faint — dark — dim, gloomy

A number of different factors have been put forward claiming to block, dampen, constrain, or otherwise restrict rendaku. The most important of these—and the only one that deserves the appellation 'law'—is Motoori-Lyman's Law (Motoori 1822; Lyman 1894), whereby rendaku is blocked (though see Martin 1987: 115; Suzuki 2005) if E_2 contains a voiced obstruent (*g*, *z*, *d*, *b*):

(2) k ~ k: *usu* + ***k**ura**g**ari* > *usu.**k**uragari*
thin, faint — darkness — twilight, dusk

Other major restricting factors include:

(3) the **right branch condition**: Otsu (1980), Itô and Mester (1986). See Vance (1980a: 234; 1987: 138–139) and Kubozono (2005: 12–15) for criticism.

---

* This paper is part of the NINJAL 'Japanese Lexicon: A Rendaku Encyclopaedia' collaborative research project (project leader: Prof. Timothy Vance).

1 In this paper, I employ a conservative phonemic analysis. Although this is not an analysis I believe to be appropriate for Japanese in general (Irwin 2011: 71–74), the reason for its adoption will become clear below.

2 Henceforth, E_1 will be used as an abbreviation for the initial element. In spite of my use of a subscript 2 in E_2, the element undergoing rendaku does not have to be the second in a compound, merely non-initial (E_2 is thus shorthand for E_{2+}). Third, or even fourth, elements in longer compounds may undergo rendaku: e.g. *buQ+komi+**z**uri+**b**ari* 'ledger fishing hook'. That rendaku occurs in the 'second element' of compounds (and therefore, by implication, in second elements only) is an error found distressingly frequently in the literature.

coordinate or dvandva compounds: Martin (1952), Okumura (1955), Kindaichi (1976), Vance (2007a) *inter alia*. Reduplicated coordinate compounds show no constraints on rendaku (Martin 1952: 49; Oyakawa 1989: 10–17), provided they are not verbs (Itô & Mester in press).

lexical stratum: rendaku is extremely rare in the foreign stratum (T. Takayama 1999, 2005; Irwin 2011: 150–153) and uncommon in the Sino-Japanese (Martin 1952; Okumura 1955; Vance 1996, 2007b; Irwin 2005).

m* < original *b: rendaku does not occur when a word-internal *m* derives from an earlier *b*: e.g. ***hi**mo* 'string', ***ka**N**mu**ri* 'crown', ***ke**mu**ri*** 'smoke', ***koo**mo**ri*** 'bat (animal)', etc. See Nakagawa (1966), Martin (1987: 31–32).

For more detailed general discussion of these factors see Okumura (1980), Vance (1987: 133–148; in press), Labrune (2006: 116–127, 2012: 112–128) or Irwin (2012: 28–29). What is of great interest for rendaku as a morphophonological phenomenon, however, is its unsystematicity. Despite the fact that none of the restricting factors outlined in (3) apply, E_2 may still fail to voice, even though it 'should':

(4)	k ~ k:	*usu*	+	***k**uti*	>	*usu.**k**uti*
		thin, faint		mouth, flavour		delicately-flavoured

In Japanese the mora (μ) 'functions as the unit of length… [and] the length of a phrase [is] roughly proportional to the number of moras it contains' (McCawley 1968: 131). There are three types of mora:

(5) (i) *(C)(G)V* structures, where the optional onset *C* is a consonant, the optional glide *G* an approximant, and *V* a vowel
(ii) the mora nasal *N*
(iii) the mora obstruent *Q*

With a very small number of exceptions restricted to particles (*Qte* 'quotative particle') and foreign names (*Njamena* 'Ndjamena (capital of Chad)'), types (ii) and (iii) do not occur word-initially in the standard language. Since this paper concerns itself with rendaku, an allomorphy restricted to element-initial moras, the duplicate moras treated here belong to type (i) structures only.

This paper seeks to debunk claims in the literature regarding the interaction of rendaku and these duplicate moras. Such claims fall into two camps: tautomorphemic duplicate moras (e.g. compounds of the form μ(…)+**μμ**(…), where the first two moras of E_2 are identical) and heteromorphemic duplicate moras (e.g. compounds of the form (…)**μ**+**μ**(…), where the final mora of E_1 and the initial mora of E_2 are identical). I treat the first of these in §3, the second in §4. Before embarking on these analyses, however, I outline the source for the empirical data employed to test these claims: the rendaku database.

2. Rendaku database

The rendaku database (Irwin & Miyashita 2013) is a database of all rendaku candidate compounds collated from two major dictionaries, Shinmura (2008) and Watanabe et al. (2008), comprising 34,201 entries. As noted in §1, a number of major factors, listed in (3), restrict rendaku. These factors operate in an overwhelmingly systematic fashion. Since I aim to demonstrate the lack of any significant statistical trace left by the interaction between rendaku and duplicate moras in an environment devoid of systematicity (what Irwin (in press) calls 'chaotic'), those compounds to which apply one or more of these major restricting factors listed in (3) are excised. Also excised from the rendaku database are the three minor, yet also systematic, factors[3] listed in (6), as well as two types of entry presenting idiosyncratic difficulties, listed in (7).

(6) **abbreviated E_2**: The very few abbreviated E_2 overwhelmingly resist rendaku: e.g. *seseri* < *seseri+tyoo* 'skipper (butterfly)' in *itimoNzi+seseri* '*Parnara guttata*', or *sasi* < *sasimi* 'sashimi' in *hugu+sasi* 'blowfish sashimi'.

E_2 suru compounds: These are restricted to cases where E_1 is a Sino-Japanese mononom (i.e. written with one sinograph). It has been well established that these undergo rendaku in an almost completely systematic fashion, governed by the phonology of E_1 (Lyman 1894: 6; Martin 1952: 49–52; Okumura 1952, 1955, 1964).

potential E_2 suffixes: What constitutes an affix in Japanese is a problem with no easy resolution (see, for example, the discussion in Irwin 2012: 32–35), but compounds where E_2 is potentially a suffix have been excised: e.g. *-sama* (as a term of address only), *-kata* (in the sense of 'way, means of' only), *-katai* (in the sense of 'difficult to do' only). These show an acute tendency either towards (*-katai*) or away from (*-kata*, *-sama*) rendaku.

(7) **anthroponyms and toponyms**: A number of anthroponyms and toponyms are written in *ateji* (Chinese characters employed for their phonetic rather than semantic value), while not a few toponyms in northern Tōhoku and Hokkaidō are borrowings from Ainu. In addition, cross-linguistically, names frequently exhibit aberrant behaviour.

E_2 with unbound voiced allomorphs: These include *kawa* ~ *gawa* 'side' or *sama* ~ *zama* 'state'. The voiced allomorph sometimes possesses a negative meaning. When the voiced allomorph of such a morpheme appears as an E_2, it is impossible to determine whether we are dealing with this voiced allomorph itself or with the rendaku form of the voiceless allomorph.

When all applicable ((3) + (6) + (7)) compounds are removed, there remains a residue of 27,900 compounds. This will henceforth be referred to as the DATABASE.

The mean rendaku rates (MRRs)[4] for this DATABASE, calculated by E_2 part of speech, are

[3] Other minor systematic factors, on whose validity the jury is still largely out, include semantics (Masuda 1979; Irwin in press), element length and prosody (Rosen 2003; Irwin 2009; Tamaoka et al. 2009), prefixes (Nakagawa 1966; Itō 2008; Irwin 2012), as well as, possibly, E_1 vocabulary stratum (Ihara & Murata 2006; Ihara et al. 2009; Tamaoka et al. 2009), frequency (van de Weijer et al. 2013) and accent (Kida 1979; Unger 2000; Yamaguchi 2011; Yamaguchi & Tanaka 2013; *inter alia*). Rendaku database entries subject to any of these factors have not been excised. In addition, dialect borrowing cannot be discounted (Tamaoka & Ikeda 2010; Ōta 2010, 2011; Vance, Miyashita & Irwin in press).

[4] The MRR for a given compound is calculated by assigning a score of 1 if it listed in a DATABASE dictionary in rendaku form, 0 if it listed in a DATABASE dictionary without rendaku, and 0.5 if it is listed with both forms in the same DATABASE dictionary. The compound's total score is then divided by its number of dictionary appearances (max: 2). A score of 0.5 can thus have two sources: a given compound is listed in both dictionaries, one with and one without rendaku; or a given compound is listed in only one dictionary but in both forms.

shown in Fig. 1. The higher the MRR, the more prone to rendaku a given part of speech is. As pointed out by Okumura (1955), Vance (2005a) and others, rendaku is heavily dampened when E_2 is a verb. The 2,899 verbs in the DATABASE have an MRR of .184, while MRRs for other parts of speech are broadly similar, ranging from .639 for adjectives to .840 for deadjectival nouns. Since both adjectives and deadjectivals comprise only a small proportion of the DATABASE (together, only 2.2%), I will lump them together with the more frequent deverbals and nouns to create a 'non-verb' category with 25,001 entries and an MRR of .768. It is these two STANDARD MRRs, .184 for verbs and .768 for non-verbs, against which candidate DATABASE compounds containing tautomorphemic and heteromorphemic duplicate moras will be measured.

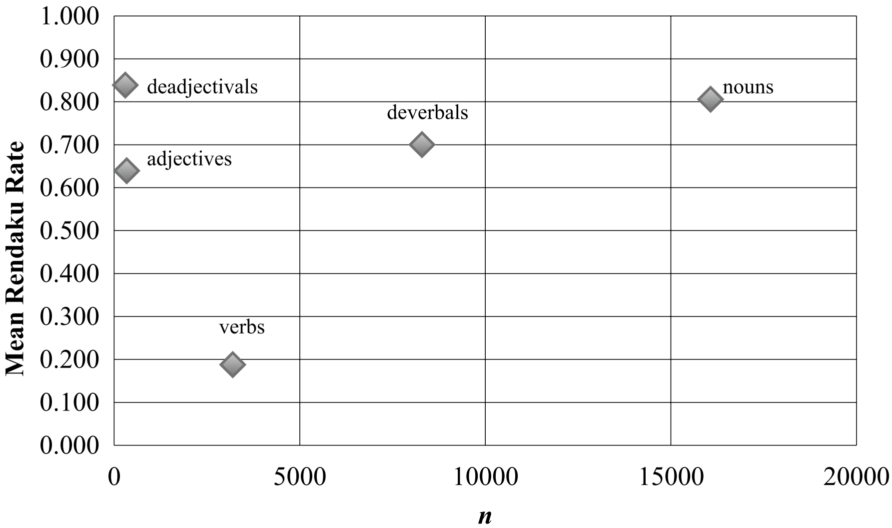

Figure 1 DATABASE Mean Rendaku Rates by part of speech (data from Irwin & Miyashita 2013)

Since the E_2-initial mora figures in the analyses of both tautomorphemic and heteromorphemic duplicate moras, it is worthwhile looking briefly at its MRR distribution. Figs. 2 and 3 show the MRRs of possible E_2-initial moras grouped by full mora (CV structures) and by mora type (C_ and _V structures), respectively. Data is for non-verbs only. Fig. 2 shows data for 20 full moras, while Fig. 3 limits itself to the 10 major mora types: the four consonant types (*k* = *ka ki ku ke ko*, etc.), in bold, and the five vowel types (*a* = *ka sa ta ha*, etc.).[5] With the full mora analysis in Fig. 2, we see that *ka* is by far the most commonly found E_2-initial mora (n = 4,785), while the two farthest outliers from the non-verb MRR are both *e*-type moras: *he* (.938) and *te* (.571). With the mora type analysis in Fig. 3, *k*- and *a*-type moras occupy a near identical position on the graph, and are by a large margin the two most common mora types (n = 10,020 and 10,058, respectively). The highest MRR is exhibited by *h*-type moras (.835). A one-way ANOVA test was run on the Fig. 2 data: there were no statistically significant differences between group means ($F(19,20) = 1.380$, $p = .24$).

[5] In Fig. 2, the 19 DATABASE entries whose E_2 begins in *kya*, *sya* and *syu* are omitted due to paucity of data. Similarly, Fig. 3 omits both *ky*-type and *sy*-type moras. Both Figs. 2 and 3 thus represent data for 24,982 non-verbs.

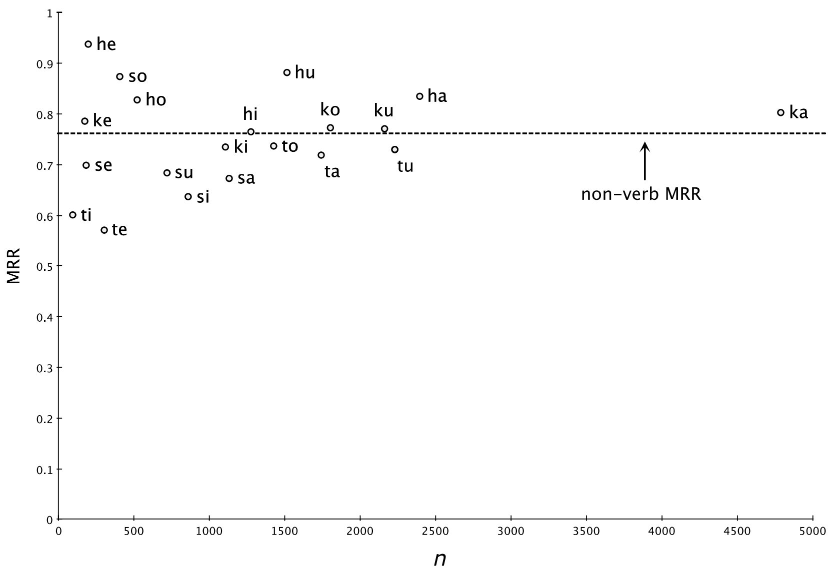

Figure 2 DATABASE Mean Rendaku Rates for non-verb E_2-initial moras (data from Irwin & Miyashita 2013)

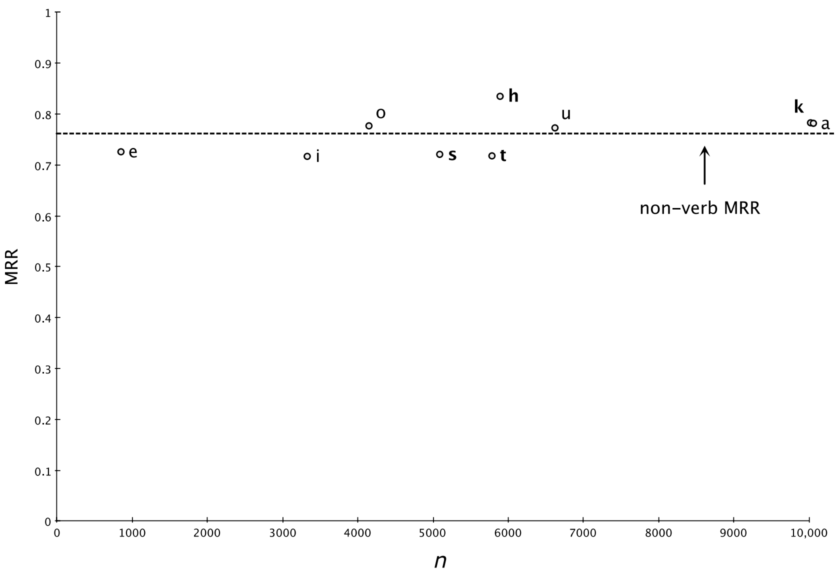

Figure 3 DATABASE Mean Rendaku Rates for non-verb E_2-initial moras, grouped by type (data from Irwin & Miyashita 2013)

3. Tautomorphemic moras

As far as the author is aware, the first claim regarding rendaku and tautomorphemic duplicate moras was made by Satō:

> 無声摩擦音音節が連続する語は連濁を起こさない。この規則は、「はは」「すす」「ひひ」などではじまる語は濁らないことを述べている。
> [Rendaku does not occur in words where two voiceless fricative syllables succeed each other. This rule means that words beginning with *haha*, *susu*, *hihi* or the like do not undergo rendaku.] Satō (1989: 256)

He then goes on to cite the example of *susuki* 薄 'plume grass'.

Satō's (1989) claim is repeated twice in subsequent work by Labrune, although in both cases she notes that *sasa* 笹 'bamboo grass' may undergo rendaku and is thus an exception to Satō's rule:

> Conditions du blocage du rendaku:… lorsque le second composé comporte une syllabe rédupliquée dont la consonne initiale est une fricative (Satô 1989).
> [Conditions under which rendaku is blocked:… when the second element contains a reduplicated syllable whose initial consonant is a fricative (Satō 1989).]
> Labrune (1999: 124–125)

> … selon Satô (1989), il arrive que l'apparition du *rendaku* soit bloquée lorsque le second composé comporte une succession de deux mores identiques impliquant les fricatives sourdes /h/ ou /s/. Les mots commençant par *haha*, *hihi*, *susu*, etc. ne subiraient donc jamais le *rendaku*.
> [According to Satō (1989), rendaku is blocked when the second element contains two identical moras with the voiceless fricatives /h/ or /s/. Thus, words beginning in *haha*, *hihi*, *susu*, etc. never undergo rendaku.] Labrune (2006: 124)

From the examples cited by Satō, I am presuming he means 'two *identical* voiceless fricative syllables', i.e. he does not intend words such as *susi* 'sushi' or *husa* 'bunch' to be covered by his 'rule'. Moreover, since rendaku is frequently blocked in the Sino-Japanese stratum and regularly in the foreign stratum (see (3) for references), I am also presuming Satō intended his 'rule' to apply to the native stratum only. Further, tautomorphemic non-light (heavy or superheavy) syllables are relatively rare in the native stratum, still rarer when the field is reduced to rendaku candidate morphemes, and non-existent when further shrunk to rendaku candidate morphemes beginning in two identical 'voiceless fricative syllables'. In other words, since there exist no rendaku candidate morphemes beginning with, for example, **suusuu*- or **sinsin*-, where the reduplicated syllable is heavy (bimoraic) or longer, Satō's 'syllable' can only refer to a light syllable. As all light syllables in Japanese are monomoraic, Satō's 'syllable' may be replaced with 'mora', as in fact Labrune (2006) does. Satō's putative rule may thus be reworded, employing the terminology I use in this paper, as:

(8) Rendaku is blocked in compounds where a native stratum E_2 begins in two duplicate voiceless fricative moras[6]

[6] A series of sound changes known collectively as labial lenition (*hagyō tenkō'on* in the Japanese tradition), beginning in Old Japanese and still incomplete, includes the change of intervocalic $p > f$ [ɸ ~ β] > *w* (before *a*) or ø (elsewhere) in the native Japanese stratum. Morpheme-internal *h* is thus exceptionally highly marked and confined to analogical restorations, dialect borrowings and other unusual, sporadic processes (e.g. *ahureru* 'overflow', *ahoo* 'idiot', *ahiru* 'duck' and examples in Table 1). The rule in (8) is therefore over-

Neither Satō (1989) nor Labrune (1999, 2006) cite any statistical evidence to back up their claims.

Of the 1,333 different E_2 in the DATABASE,[7] only 15 (1.1% of the total) begin in two duplicate voiceless fricative moras, as defined in (8). These appear in 74 compounds (a mere 0.3% of the DATABASE) and are listed in Table 1 below, along with their English gloss, part of speech, the number of compounds in which they appear (*n*), and their MRR.

Table 1 Mean Rendaku Rates for E_2 beginning in two duplicate voiceless fricative moras (data from Irwin & Miyashita 2013)

CANDIDATE MORPHEME	ENGLISH GLOSS	PART OF SPEECH	*n*	MRR
sasa 笹	bamboo grass	noun	20	.950
sisi 肉	flesh, meat	noun	6	1.000
susuki 薄	plume grass (*Miscanthus sinensis*)	noun	16	.000
haha 母	mother	noun	4	.000
hahaki 帚	broom	noun	2	.375
huhuki 蕗	butterbur (*Petasites japonicus*)	noun	1	.000
ALL NOUNS			**49**	**.526**
sasae 支え	supporting	deverbal noun	3	.667
susumi 進み	proceeding	deverbal noun	1	.000
susuri 啜り	sipping, slurping, sucking	deverbal noun	1	.000
seseri 挵り	picking at	deverbal noun	9	.556
ALL DEVERBAL NOUNS			**14**	**.500**
ALL NON-VERBS			**63**	**.520**
sasaeru 支える	support	verb	1	.000
sasaru 刺さる	embed itself	verb	1	.000
susumu 進む	proceed	verb	6	.000
susumeru 勧める	recommend	verb	1	.000
susumeru 進める	advance, promote	verb	2	.000
ALL VERBS			**11**	**.000**
TOTAL			**74**	**.443**

Although the MRRs for non-verbs and verbs of .520 and .000, respectively, are both lower than the comparable STANDARD MRRs of .768 and .184, respectively, the number of compounds involved is too small to come to a conclusion that holds any statistical significance. The putative rule in (8) is manifestly false: it is not the case that rendaku 'is blocked'. Not only does rendaku

whelmingly applicable to 'two identical *s*-type moras'. See Kiyose (1985), Martin (1987: 10–13) or Frellesvig (2010: 201–210) for historical detail.

[7] Different polysemes are counted as independent E_2. See Irwin (2012).

occur in *sasa*, as pointed out by Labrune, with a higher than average MRR of .950, it also occurs in the now obsolete *sisi* and *hahaki*, as well as in the two deverbals *sasae* and *seseri*. The only two E_2 to occur in compounds more than 10 times in the DATABASE, *sasa* and *susuki*, either overwhelming favour or disfavour rendaku. Of the 74 compounds in Table 1, only 22 (29.7%) appear in both rendaku database dictionaries and are thus definable, to any extent, as 'frequent'. Of these, the non-verbs have an MRR of .574 (17/22), slightly higher than the overall non-verb MRR of .520 in Table 1.

If we examine the two fricatives *s* and *h* separately, *s* non-verbs exhibit an MRR of .571, while the MRR for *h* non-verbs is .107. The MRR for *s* non-verbs is thus lower than the STANDARD MRR of .768. While, at .107, the comparable figure for *h* is very low, there are only 7 examples, all of whose E_2 are highly marked.[8] The most that can be said about an E_2 beginning in two identical voiceless fricative moras is that there is—at a push—a very weak tendency for those beginning in *s* to resist rendaku, but that there exist insufficient examples for those beginning in *h* to provide any significant statistical corroboration. The rule stated in (8) must be rejected.

4. Heteromorphemic moras

The first claim regarding rendaku and heteromorphemic duplicate moras was also put forward by Satō:

> 連濁によって、同種または類似の音が連続するときは、濁音化が避けられ易い。
> [Rendaku frequently does not occur where two identical, or similar, sounds would occur in succession.] Satō (1989: 255)

He then cites, as example compounds with 'identical sounds,' *kizu+**tukeru*** 'scar, wound', as opposed to *ato+**zukeru*** 'follow up', *iti+**zukeru*** 'rank' and *na+**zukeru*** 'name', where E_2 is *tukeru ~ zukeru* 付ける 'attach'; as well as *tobi+**hi*** 'flying sparks', as against *taki+**bi*** 'bonfire', *kitune+**bi*** 'will-o'-the-wisp', *nokori+**bi*** 'ember' and *morai+**bi*** 'catching fire', where E_2 is *hi ~ bi* 火 'fire'. The rendaku forms **kizuzukeru* and **tobibi* are blocked, according to Satō, since the duplicate 'identical sounds' *zu* and *bi* would result. As example compounds with 'similar sounds', Satō cites *siage+**kaNna*** 'smoothing plane' as opposed to *dai+**gaNna*** 'block plane', *yari+**gaNna*** 'tipped plane' and *tuki+**gaNna*** 'Western-style plane', where E_2 is *kaNna ~ gaNna* 鉋 'plane'. Here, the rendaku form **siagegaNna* is blocked, he contends, due to the fact that the duplicate 'similar sounds' *ge* and *ga* (i.e. mora-initial *g*) would result.

M. Takayama (1992: 116), quoting Satō (1989), makes the same claim, though for 'identical sounds (同音)' only. Later, in M. Takayama (2012: 109–110), he reaffirms his position on 'identical sounds', although this time more circumspectly. Citing it as a classic example of haplology, he also lists a number of exceptions, including *yobi+bi* 'alternative date (for an event)' and *soba+batake* 'field of buckwheat'.

Labrune (1999: 125) repeats Satō's (1989) claim regarding both 'identical' and 'similar' sounds, quoting his *hi ~ bi* 'fire' examples *verbatim*, though using the word 'segment' rather than 'sound'. She repeats both claims once more in Labrune (2006: 123–124), this time citing both his *hi ~ bi* 'fire' and *tukeru ~ zukeru* 'make, attach' examples *verbatim* and, in place of 'sound', using 'élément phonologique qui peut être soit une more, soit un segment [phonological element, be it a mora or a segment]'. Finally, she repeats both Satō's claims a third time in Labrune (2012:

8 See footnote 6.

120–121), essentially an English translation of Labrune (2006: 123–124), citing Satō's *hi ~ bi* 'fire' and *tukeru ~ zukeru* 'make, attach' examples *verbatim* yet again and this time writing 'phonological element, which can be either a mora or a segment' for Satō's 'sound'. In both Labrune (2006) and (2012) she also quotes a single exception to Satō's claim: *tabi+bito* 'wayfarer'.

From his examples (and also reflected in M. Takayama (1992, 2012) and Labrune (2006, 2012)), it is clear that what Satō means by 'identical sound' is in fact 'identical mora', while by 'similar sound' he means 'mora beginning with an identical initial'. When possible rendaku candidate moras are taken into consideration, Satō's claim can be reworded as:

(9) Rendaku is frequently avoided when two duplicate moras, or two moras beginning with the same voiced obstruent (e.g. …..*da+do*…..), would occur across an element boundary.[9]

Once again, neither Satō (1989), M. Takayama (1992, 2012) nor Labrune (1999, 2006, 2012) cite any statistical evidence whatsoever to corroborate the claim in (9). Before examining this evidence in detail, however, it is worth considering the data for E_1-final moras in general. This is shown in Fig. 4 for the 25,001 DATABASE non-verb compounds.[10] Here, E_1-final moras are categorized in two different ways: by mora-initial consonant (V = no initial consonant; N = mora nasal) and by mora-final vowel. The former are shown in bold. Two separate one-way ANOVA tests were run, one by mora-initial consonant and one by mora-final vowel. In both cases, there were no statistically significant differences between group means (mora-initial consonant test: $F(15,16) = 1.411, p = .25$; mora-final vowel test: $F(5,6) = 0.914, p = .53$). In other words, it cannot be argued in the analyses to follow that a particular E_1-final mora is exerting an influence.

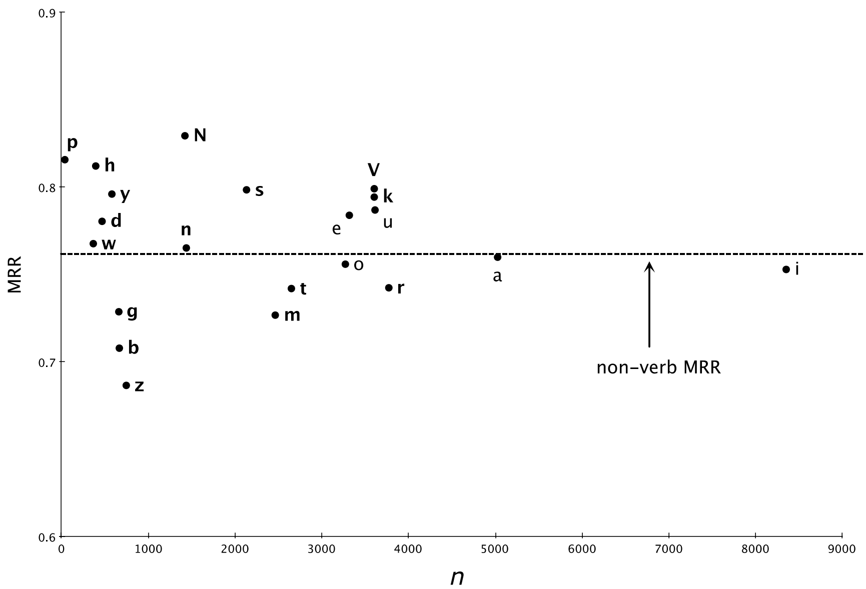

Figure 4 DATABASE Mean Rendaku Rates for non-verb E_1-final moras (data from Irwin & Miyashita 2013)

[9] When *di*, *du* are treated as *zi*, *zu*.

[10] The single E_1-final *tse*-mora, problematic under a conservative phonemicization, is treated as a *t*-mora.

I turn first to the strong version of the claim, i.e. that the two moras straddling the element boundary are identical. As with the tautomorphemic analysis above, non-verb and verb data are examined separately. Data for the non-verb strong claim is thin on the ground, with only 89 compounds (a scant 0.4% of all non-verb DATABASE compounds) possessing a potential rendaku site for a duplicate heteromorphemic mora. These are spread across 14 different E_1-final moras and shown in Fig. 5 (there are no relevant E_1-final *be*, *de*, *ge* or *ze* moras). If the strong version of the claim in (9) were true, these moras should all show severely depressed MRRs. This is clearly not the case. Exactly half the E_1-final moras exhibit an MRR higher than the non-verb MRR of .768, although in many cases *n* is small. Overall, the average MRR for non-verb DATABASE compounds possessing a potential rendaku site for a duplicate heteromorphemic mora is .567, which although somewhat dampened, can by no means be construed as statistically representative of a situation where rendaku is 'frequently avoided.'

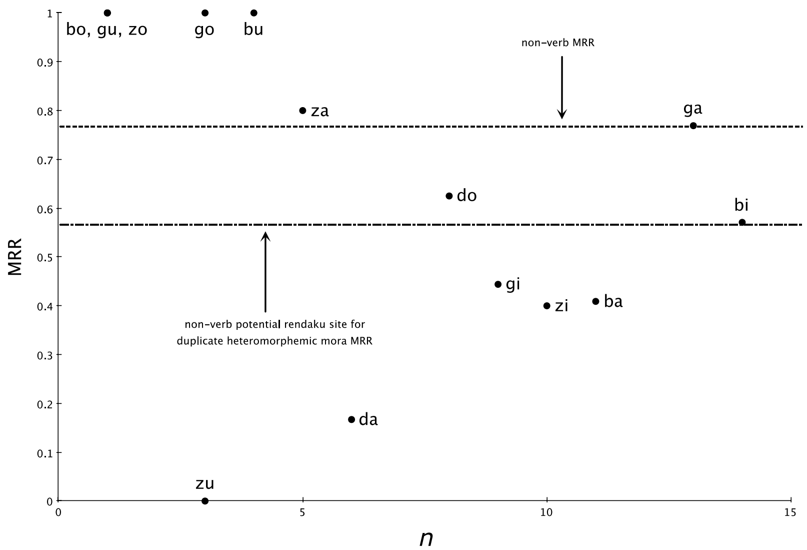

Figure 5 Mean Rendaku Rates for non-verb DATABASE compounds possessing a potential rendaku site for a duplicate heteromorphemic mora (data from Irwin & Miyashita 2013)

Data for the strong claim for verbs are necessarily even thinner, with only six candidates in the entire DATABASE:[11] two exhibiting rendaku (*mizu+zuku* 'soak' and *suga+gaku* 'perform on the *wagon* (a stringed instrument)', both now obsolete) and four not (*kizu+tukeru* 'scar, wound' as cited by Satō, its intransitive companion *kizu+tuku* 'be scarred, wounded', *guzu+tuku* 'dawdle', and the obsolete *roozi+simu* 'occupy, possess'). Although this yields an overall verb MRR of .200 (on a close par with the STANDARD MRR for verbs of .184), the data are insufficient for any meaningful conclusions to be drawn.

[11] This even with an extremely conservative approach, which allows E_1-final *zu* and *zi* two voiceless E_2-initial counterparts each: *su*, *tu* and *si*, *ti*, respectively.

Turning now to the weak version of the claim in (9), that the two moras straddling the element boundary begin only with the same obstruent, the volume of data is considerably larger: some 425 non-verb compounds (1.7% of all non-verb DATABASE compounds) possess a potential rendaku site. These are spread across all the possible 18 different E_1-final moras and shown in Fig. 6. As with its strong version, if the weak version of the claim in (9) were true, these moras should show severely depressed MRRs. Once again, as with the strong version, this is clearly not the case: in fact, even less so. Although only three E_1-final moras have an MRR higher than the non-verb MRR of .768, the average MRR here is .652, scarcely one point below the non-verb MRR. The weak version of the claim in (9) holds no water whatsoever and must be rejected.[12]

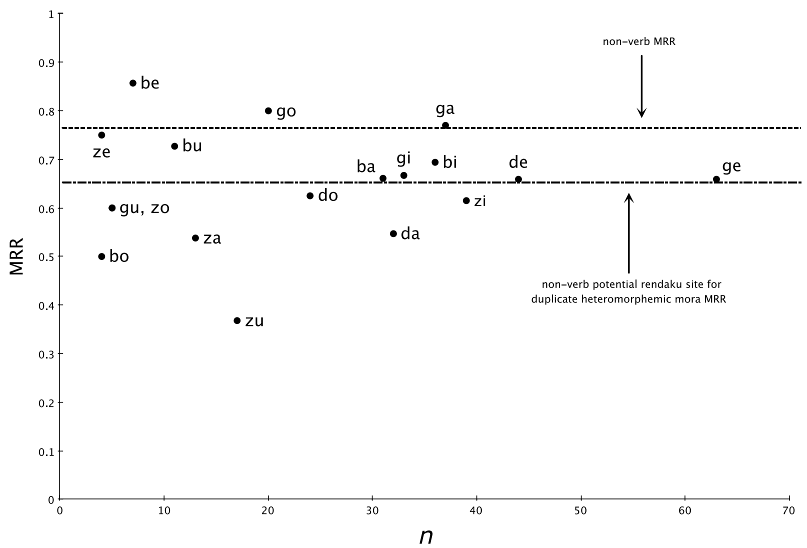

Figure 6 Mean Rendaku Rates for non-verb DATABASE compounds possessing a potential rendaku site for a duplicate mora beginning in the same voiced obstruent (data from Irwin & Miyashita 2013)

As with the strong version of (9), with the weak version also data for verbs is extremely limited. There are only 58 such rendaku candidate compounds (just 2% of all verb DATABASE compounds)[13] and together they show a MRR of .129, a little lower than the verb STANDARD MRR of .184. Here, then, rendaku is indeed 'frequently avoided' although, as we have seen in Fig. 1, for verbs this is generally the case anyway. Once again, however, the data are insufficient for

[12] Although not mentioned in (9) and, to my knowledge not advocated by Satō, M. Takayama, Labrune or anyone else, a corollary of the heteromorphemic mora claim is that an E_1 ending in a voiceless mora should be more likely to undergo rendaku, in order to avoid two identical moras (the 'strong corollary') or two moras beginning with the same voiced obstruent (the 'weak corollary'). An analysis of the DATABASE, for non-verbs only, shows that the MRRs for applicable compounds are only slightly elevated: the MRR is .812 for the strong corollary (n = 190) and a near identical .814 for the weak corollary (n = 1643).

[13] As with the corresponding strong claim, I adopt an extremely conservative approach and permit E_1-final *zu* and *zi* moras both *s* and *t* voiceless E_2-initial moras.

any meaningful conclusions to be drawn.

5. Discussion

As M. Takayama (2012: 110) states, it is likely Satō's (1989) claim in (9) for duplicate heteromorphemic moras was motivated by an appeal to haplology or, as Labrune (2012: 120) puts it, a 'dissimilatory principle'. The same may be said for the claim in (8), although why it should be restricted to fricatives is unclear. Motoori-Lyman's Law, the most important of the major rendaku restricting factors in (3) is also a process of dissimilation, albeit tautomorphemic. Ramsey & Unger (1972: 287–289), however, proposed a more strongly dissimilatory—because heteromorphemic—'strong clause revision' of the Law, under which rendaku is blocked when E_1 contains a voiced obstruent.[14] This strong version of the Law appears to have been a reality in Old Japanese (OJ), if E_1 are restricted to nouns of 2 moras or less, which accounts for the vast bulk of attested compounds (Vance 2005b: 28–33; Vance & Irwin 2012, 2013).[15] However, this strong version has since broken down: and as early as the 17th century, according to the MRR[16] garnered by Toda (1988: 87) from the Japanese-Portuguese dictionary *Nippō Jisho*, published in 1603. Compounds such as *nodo+botoke* 'Adam's apple', *rezi+bukuro* 'supermarket checkout bag' or *yubi+zumoo* 'thumb wrestling' are rife in the modern language. Satō's (1989) claim in (9) may thus be regarded as a restricted variant of Ramsey & Unger's (1972) 'strong clause revision', although purporting to still apply in Modern Japanese (ModJ): restricted to the E_1-final mora—though, since no OJ morpheme could begin in a voiced obstruent, it is tantamount to Vance's (2005b) *de facto* restriction—and restricted yet further to either completely identical or initial-identical moras. For a schematization of these dissimilatory processes, see Fig. 7. Here, the *y*-axis indicates time, while bold moras indicate the object of the dissimilatory process in question.

Kawahara & Sano (2013), in an experiment employing nonce words, claim that rendaku is observed less often (and to a statistically significant extent) when it results in adjacent but heteromorphemic identical CV moras. Rendaku blockage due to what they term 'identity avoidance' would appear to corroborate the putative rule in (9) and thus offer counter-evidence for one of the central claims of this paper. Indeed, in the history of rendaku research, this would not be the first time apparent psychological reality has not squared with the statistical facts: see Vance (1979, 1980b) for the case of Motoori-Lyman's Law.[17]

Another claim similar to that in (9) was made by Sugitō (1965), though severely restricted to a miniscule proportion of vocabulary: surnames ending in -*ta*, written 田. Her analysis showed that the *ta* morpheme underwent rendaku only 35% of the time when the E_1-final mora contained a voiced obstruent, a liquid or a semi-vowel; and that the same *ta* morpheme underwent

[14] Later, as pointed out by Vance (in press), Unger (1975: 9) credits this to Ishizuka (1801). See also Miyake (1932: 136).

[15] Miller (1984) proposed an even stronger version of the Law in which rendaku was blocked in OJ when the initial element in a compound contained any voiced consonant, whether obstruent or sonorant. See Vance & Irwin (2012, 2013) for a refutation.

[16] Toda's MRR for compounds listed in the *Nippō Jisho* whose E_1 ended in a voiced obstruent was .480 (n = 368).

[17] Kawahara & Sano (2013) conducted their survey online and thus this author awaits data replicated by more rigorously controlled experiments. Further, many of their nonce words were not readily interpretable as native stratum nouns, but as either verbs (with expected low MRRs—see Fig. 1) or as compounds (expected to conform to the right branching condition—see (3)).

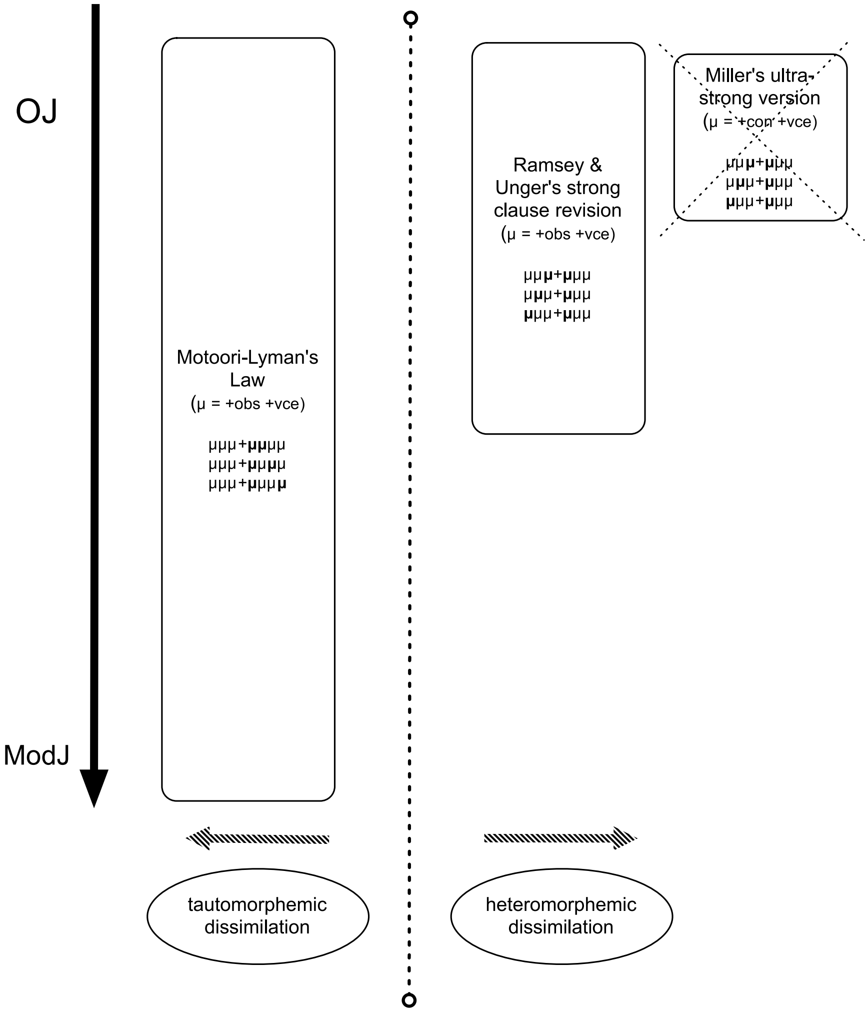

Figure 7 Dissimilatory processes proposed for Japanese compounds

rendaku 98% of the time when the E_1-final mora contained a voiceless obstruent or a nasal.[18] Similar analyses by Hirata (2010, 2011) produced more nuanced, though broadly similar, results to Sugitō's earlier findings. Zamma (2005) found that comparable tendencies were apparent for other (but by no means all) common surname-final morphemes such as *-sawa* 沢/澤, *-sima* 島/嶋, *-tuka* 塚 and *-saki* 崎.

That said, the analyses in §3 and §4 have demonstrated conclusively that any putative influence exerted by duplicate moras in the direction of restricting rendaku must be rejected. This conclusion has been drawn from statistical analyses of empirical data: it is most unfortunate that neither Satō in his original claims in (8) and (9), nor M. Takayama and Labrune in their subsequent restatements of these claims, chose to collate sufficient data to corroborate their asser-

[18] Accent patterns also played a role. Ohta (2013) concludes that only accent, and not the E_1-final mora, plays a role in the rendaku patterns found in E_2 *kawa* 'river', when restricted to river names.

tions.[19] It is to be hoped that, unlike the Eskimo-words-for-snow 'hoax' (Cichocki & Kilarski 2010; Krupnik & Müller-Willer 2010), a rendaku-blocked-by-duplicate-mora 'hoax' will now cease to perpetuate itself in the literature, and die a swift and well-warranted death.

References

Cichocki, Piotr & Marcin Kilarski (2010) On 'Eskimo Words for Snow': The life cycle of a linguistic misconception. *Historiographia Linguistica* 37(3): 341–377.

Frellesvig, Bjarke (2010) *A history of the Japanese language*. Cambridge: Cambridge University Press.

Hirata, Junko (2010) Rendaku shiron. *Kōbe Kaisei Joshi Gakuin Daigaku Kenkyū Kiyō* 49: 31–37.

Hirata, Junko (2011) Rendaku shiron (sono 2). *Kōbe Kaisei Joshi Gakuin Daigaku Kenkyū Kiyō* 50: 89–93.

Ihara, Mutsuko & Tadao Murata (2006) Nihongo no rendaku ni kansuru ikutsuka no jikken. *On'in Kenkyū* 9: 17–24.

Ihara, Mutsuko, Katsuo Tamaoka & Tadao Murata (2009) Lyman's Law effect in Japanese sequential voicing: Questionnaire-based nonword experiments. In: The Linguistic Society of Korea (ed.) *Current issues in unity and diversity of languages*, 1007–1018. Seoul: Dongnam Publishing.

Irwin, Mark (2005) Rendaku-based lexical hierarchies in Japanese: The behaviour of Sino-Japanese mononoms in hybrid noun compounds. *Journal of East Asian Linguistics* 14: 121–153.

Irwin, Mark (2009) Prosodic size and rendaku immunity. *Journal of East Asian Linguistics* 18: 179–196.

Irwin, Mark (2011) *Loanwords in Japanese*. Amsterdam: John Benjamins.

Irwin, Mark (2012) Rendaku dampening and prefixes. *NINJAL Research Papers* 4: 27–36.

Irwin, Mark (in press) Rendaku lovers, rendaku haters and the logistic curve. In: Mikio Giriko, Kyoko Kanzaki, Naonori Nagaya & Akiko Takemura (eds.) *Japanese/Korean linguistics 22*, Stanford: CSLI Publications.

Irwin, Mark & Mizuki Miyashita (2013) The rendaku database version 2.0. Available online at http://www-h.yamagata-u.ac.jp/~irwin/site/Rendaku_Database.html.

Ishizuka, Tatsumaro (1801) *Kogen seidaku kō*. Kyoto: Hishiya Magobei, Zeniya Rihei, Hayashi Ihei.

Itô, Junko & Armin Mester (1986) The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17: 49–73.

Itô, Junko & Armin Mester (in press) Word formation and phonological processes. In: Haruo Kubozono (ed.).

Itō, Mitsu (2008) Rendaku ni tsuite. *Kyūshū Kokusai Daigaku Kyōyō Kenkyū* 15(2): 83–102.

Kawahara, Shigeto & Shin'ichiro Sano (2013) *Testing rendaku experimentally: Rosen's Rule, (strong) Lyman's Law and identity avoidance*. Paper presented at the Tokyo Circle of Phonologists, October Meeting, Tokyo University.

Kida, Akiyoshi (1979) Rendaku to akusento. *Kokugo Kokubun* 48(3): 51–64.

Kindaichi, Haruhiko (1976) Rendaku no kai. *Sophia Linguistica* 2: 1–22.

Kiyose, Gisaburō (1985) Heianchō hagyōshiin-p onron. *Onsei no Kenkyū* 21: 73–87.

Krupnik, Igor & Ludger Müller-Willer (2010) Franz Boas and Inuktitut terminology for ice and snow: From the emergence of the field to the "Great Eskimo Vocabulary Hoax". In: Igor Krupnik, Claudio Aporta, Shari Gearheard, Gita Laidler & Lene Kielsen Holm (eds.) *SIKU: Knowing our ice*, 377–400. Dordrecht: Springer.

Kubozono, Haruo (2005) Rendaku: Its domain and linguistic conditions. In: Jeroen van de Weijer, Kensuke Nanjo & Tetsuo Nishihara (eds.), 5–20.

Kubozono, Haruo (ed.) (in press) *The handbook of Japanese phonetics and phonology*. Berlin: Mouton de Gruyter.

Labrune, Laurence (1999) Variation intra et inter-langue morpho-phonologie du rendaku en Japonais et du sai-sios en Coréen. *Cahiers de Grammaire* 24: 117–152.

Labrune, Laurence (2006) *La phonologie du Japonais*. Leuven and Paris: Peeters.

Labrune, Laurence (2012) *The phonology of Japanese*. Oxford: Oxford University Press.

[19] Labrune (2012: 121) does concede, however, that a 'thorough statistical analysis is called for'.

Lyman, Benjamin (1894) Change from surd to sonant in Japanese compounds. *Oriental Club of Philadelphia*, 1–17.
Martin, Samuel (1952) Morphophonemics of standard colloquial Japanese. *Language* 28: 3 (Part 2, Supplement).
Martin, Samuel (1987) *The Japanese language through time*. New Haven & London: Yale University Press.
Masuda, Yutaka (1979) Nihongo ren'onhenka no on'inkisoku. In: Shigeru Suzuki (ed.) *Ningen to shakai no shomondai*, 1123–1146. Matsuyama: Matsuyama Shōka Daigaku.
McCawley, James (1968) *The phonological component of a grammar of Japanese*. The Hague: Mouton.
Miller, Roy Andrew (1984) The first statement fo Lyman's Law. *Bochumer Jahrbuch zur Ostasienforschung* 7: 37–56.
Miyake, Takeo (1932) Dakuonkō. *Onsei no kenkyū* 5: 135–190.
Motoori, Norinaga (1822) *Kojiki den*. Edo: Maegawa Rokuzaemon; Osaka: Matsumura Kyūbei; Kyoto: Imai Kihei; Owari: Katano Tōshirō.
Nakagawa, Yoshio (1966) Rendaku, rensei (kashō) no keifu. *Kokugo Kokubun* 35: 302–314.
Ohta, Satoshi (2013) On the relationship between rendaku and accent. In: Jeroen van de Weijer & Tetsuo Nishihara (eds.), 63–87.
Okumura, Mitsuo (1952) Jion no rendaku ni tsuite. *Kokugo Kokubun* 21(6): 8–22.
Okumura, Mitsuo (1955) Rendaku. In: Kokugogakkai (ed.) *Kokugogaku jiten*, 961–962. Tokyo: Tōkyōdō.
Okumura, Mitsuo (1964) Shōzuru to shōsuru: Rendaku no yōin ni tsuite. *Kokugo Kokubungaku* 3: 1–5.
Okumura, Mitsuo (1980) Rendaku. In: Kokugogakkai (ed.) *Kokugogaku jiten*, 925–926. Tokyo: Tōkyōdō.
Ōta, Makie (2010) Wakamono ni ooi 'wakashiraga', kōnensō ni nokoru 'wakajiraga'. *Hōsō Kenkyū to Chōsa* 2010(11): 50–70.
Ōta, Makie (2011) Onna wa otoko yori mo 'tsumitsukuri'. *Hōsō Kenkyū to Chōsa* 2011(11): 26–37.
Otsu, Yukio (1980) Some aspects of rendaku in Japanese and related problems. In: Yukio Otsu & Ann Farmer (eds.), 207–227.
Otsu, Yukio & Ann Farmer (eds.) (1980) *Theoretical issues in Japanese linguistics*. MIT Working Papers in Linguistics.
Oyakawa, Takatsugu (1989) Some remarks on Japanese rendaku. *Ryudai Review of Language and Literature* 34: 1–24.
Ramsey, S. Robert & J. Marshall Unger (1972) Evidence of a consonant shift in 7th century Japanese. *Papers in Japanese Linguistics* 1(2): 278–295.
Rosen, Eric Robert (2003) Systematic irregularity in Japanese rendaku: How the grammar mediates patterned lexical exceptions. *Canadian Journal of Linguistics / Revue Canadienne de Linguistique* 48: 1–37.
Satō, Hirokazu (1989) Fukugōgo ni okeru akusentokisoku to rendakukisoku. In: Miyoko Sugitō & Yutaka Miyaji (eds.) *Nihongo no onsei on'in Vol. 1*, 233–265. Tokyo: Meiji Shoin.
Shinmura, Izuru (ed.) (2008) *Kōjien dai-6 han*. Tokyo: Iwanami Shoten.
Sugito, Miyoko (1965) Shibata-san to Imada-san: Tango no chōkakuteki benbetsu ni tsuite no ichi kōsatsu. *Gengo Seikatsu* 165: 64–72.
Suzuki, Yutaka (2005) Raiman no hōsoku no reigai ni tsuite. *Bunkyō Gakuin Daigaku Gaikokugogakubu Bunkyō Gakuin Tankidaigaku Kiyō* 4: 249–265.
Takayama, Michiaki (1992) Rendaku to renjōdaku. *Kuntengo to Kuntenshiryō* 88: 115–124.
Takayama, Michiaki (2012) *Nihongo on'inshi no kenkyū*. Tokyo: Hituzi Syobo.
Takayama, Tomoaki (1999) Shakuyōgo no rendaku/kōonka ni tsuite. *Report of the Special Research Project for the Typological Investigation of Languages and Cultures of the East and West* 1: 375–385.
Takayama, Tomoaki (2005) A survey of rendaku in loanwords. In: Jeroen van de Weijer, Kensuke Nanjo & Tetsuo Nishihara (eds.), 177–190.
Tamaoka, Katsuo, Mutsuko Ihara, Tadao Murata & Hyunjung Lim (2009) Effects of first-element phonological length and etymological-type features on sequential voicing (*rendaku*) of second elements. *Journal of Japanese Linguistics* 25: 17–38.
Tamaoka, Katsuo & Fumiko Ikeda (2010) Whiskey or bhiskey?: Influence of first element and dialect region on sequential voicing of *shoochuu*. *Gengo Kenkyū* 137: 65–79.
Toda, Ayako (1988) Wago no hirendaku kisoku to rendaku keikō. *Dōshisha Kokubungaku* 30: 80–98.

Unger, J. Marshall (1975) *Studies in early Japanese morphophonemics*. New Haven: Yale University dissertation.

Unger, J. Marshall (2000) *Rendaku* and proto-Japanese accent classes. In: Mineharu Nakayama & Charles Quinn (eds.) *Japanese/Korean Linguistics 9*, 17–30. Stanford: CSLI Publications.

van de Weijer, Jeroen, Kensuke Nanjo & Tetsuo Nishihara (eds.) (2005) *Voicing in Japanese*. Berlin & New York: Mouton de Gruyter.

van de Weijer, Jeroen & Tetsuo Nishihara (eds.) (2013) *Current issues in Japanese phonology: Segmental variation in Japanese*. Tokyo: Kaitakusha.

van de Weijer, Jeroen, Clemens Poppe & Marjoleine Sloos (2013) Family matters: Lexical aspects of Japanese rendaku. In: Jeroen van de Weijer & Tetsuo Nishihara (eds.), 129–148.

Vance, Timothy (1979) *Nonsense-word experiments in phonology and their application to rendaku in Japanese*. PhD dissertation, University of Chicago.

Vance, Timothy (1980a) Comments on Ōtsu (1980). In: Yukio Ōtsu & Ann Farmer (eds.), 229–236.

Vance, Timothy (1980b) The psychological status of a constraint on Japanese consonant alternation. *Linguistics* 18: 245–267.

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

Vance, Timothy (1996) Sequential voicing in Sino-Japanese. *Journal of the Association of Teachers of Japanese* 30: 22–43.

Vance, Timothy (2005a) Rendaku in inflected words. In: Jeroen van de Weijer, Kensuke Nanjo & Tetsuo Nishihara (eds.), 89–104.

Vance, Timothy (2005b) Sequential voicing and Lyman's Law in Old Japanese. In: Salikoko S. Mufwene, Elaine J. Francis & Rebecca S. Wheeler (eds.) *Polymorphous linguistics: Jim McCawley's legacy*, 27–43. Cambridge: MIT Press.

Vance, Timothy (2007a) *Right branch or head—What difference does it make?* In: Susumu Kuno, Seiichi Makino & Susan Strauss (eds.) *Aspects of linguistics: In honor of Noriko Akatsuka*, 221–240. Tokyo: Kurosio.

Vance, Timothy (2007b) Reduplication and the spread of rendaku into Sino-Japanese. *Proceedings of the 31st Annual Meeting*, 56–64. Ōsaka: Kansai Linguistic Society.

Vance, Timothy (in press) Rendaku. In: Haruo Kubozono (ed.).

Vance, Timothy & Mark Irwin (2012) *The first statement of Lyman's Law*. Paper presented at the 25th Paris Meeting on East Asian Linguistics, École des Hautes Études en Sciences Sociales, Paris.

Vance, Timothy & Mark Irwin (2013) *A rendaku database for Old Japanese*. Paper presented at the 21st International Conference on Historical Linguistics, University of Oslo.

Vance, Timothy, Mizuki Miyashita & Mark Irwin (in press) Rendaku in Japanese dialects that retain prenasalization. In: Seungho Nam, Heejeong Ko & Jongho Jun (eds.) *Japanese/Korean Linguistics 21*. Stanford: CSLI Publications.

Watanabe, Toshirō, Edmund Skrzypczak & Paul Snowden (eds.) (2008) *Kenkyūsha's new Japanese-English dictionary digital supplemented edition*. Tokyo: Kenkyūsha.

Yamaguchi, Kyoko (2011) Accentedness and rendaku in Japanese deverbal compounds. *Gengo Kenkyū* 140: 117–133.

Yamaguchi, Kyoko & Shin-ichi Tanaka (2013) Rendaku variation in deverbal compounds. In: Jeroen van de Weijer & Tetsuo Nishihara (eds.), 149–166.

Zamma, Hideki (2005) The correlation between accentuation and rendaku in Japanese surnames: A morphological account. In: Jeroen van de Weijer, Kensuke Nanjo & Tetsuo Nishihara (eds.), 157–176.

同一モーラの連続における連濁

アーウィン　マーク
山形大学／国立国語研究所 共同研究員

要旨

同一モーラが二つ連続する場合の連濁に関して，先行研究には次のような二つの主張が根強くある。「カレ＋ススキ（枯れ薄）」のように同一の摩擦音モーラが二つ同形態素で連続する場合，連濁は起こらないとされている。さらに，もし連濁が起こるとしても，「キズ＋ツケル（傷つける）」のように二つの同一モーラが異なる形態素にまたがって連続する場合は，連濁は非常に起こりにくいと主張されている。とはいえ，いずれの主張も統計的な証拠によっては一度も確証されておらず，「言語学的都市伝説」と見なさざるを得ない。この論文では，連濁データベースに基づいた統計的分析を通して，先行研究における「連続するモーラ」に関する主張への裏付けは極めて貧弱で，放棄すべきであると論じる。

キーワード：連濁，モーラ，異化，連濁データベース，言語学的都市伝説