Rendaku Lovers, Rendaku Haters & The Logistic Curve

Mark Irwin
Yamagata University
Rendaku – or sequential voicing – is a productive morphophonemic process found in Japanese.

It is a process found only in compounds, and restricted to non-initial elements (hereafter NIEs) which begin in a voiceless obstruent.

In an NIE, the initial voiceless obstruent may voice.

Some examples are shown in the following slide.
<table>
<thead>
<tr>
<th>Word 1</th>
<th>Word 2</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>kawa</code></td>
<td><code>soko</code></td>
<td><code>kawa.zoko</code></td>
</tr>
<tr>
<td>‘river’</td>
<td>‘bottom’</td>
<td>‘riverbed’</td>
</tr>
<tr>
<td><code>cuke</code></td>
<td><code>cume</code></td>
<td><code>cuke.zume</code></td>
</tr>
<tr>
<td>‘attach’</td>
<td>‘fingernail’</td>
<td>‘false nail’</td>
</tr>
<tr>
<td><code>hana</code></td>
<td><code>kami</code></td>
<td><code>hana.gami</code></td>
</tr>
<tr>
<td>‘nose’</td>
<td>‘paper’</td>
<td>‘tissue’</td>
</tr>
<tr>
<td><code>uširo</code></td>
<td><code>tate</code></td>
<td><code>uširo.date</code></td>
</tr>
<tr>
<td>‘behind’</td>
<td>‘shield’</td>
<td>‘backer, supporter’</td>
</tr>
</tbody>
</table>
Rendaku is not restricted to Modern Tokyo Japanese

Synchronously, it is found in all modern Japanese dialects, though research has been insufficient for determining to what extent behaviour differs

Diachronically, it is found in texts throughout all stages of the history of the language

Some of you may have noticed that I stated earlier that ‘the initial voiceless obstruent [of an NIE] may voice’

This is because many decades of detailed research on rendaku have shown that the process is hedged with constraints and conditions

There are many instances where rendaku may occur but does not

Many of these cases can be explained in a principled manner by rules and conditions (systematic constraints), but some cannot (unsystematic variation)
Systematic Constraints

There are six types of systematic processes which block or inhibit rendaku. These are based on:

- Motoori’s/Lyman’s Law
- vocabulary strata
- coordinate compounds
- the ‘right-branching rule’
- bare numeral initial elements
- \([m] < \text{original} [b]\)
Motoori’s/Lyman’s Law

If an NIE already contains a voiced obstruent, then rendaku is blocked

\[ \text{aši} \quad + \quad \text{kubi} \quad > \quad \text{aši.kubi (*aši.gubi)} \]

‘leg’ \quad ‘neck’ \quad ‘ankle’

\[ \text{mamizu} \quad + \quad \text{kurage} \quad > \quad \text{mamizu.kurage (*mamizu.gurage)} \]

‘freshwater’ \quad ‘jellyfish’ \quad ‘freshwater jellyfish’

Motoori (1822), Lyman (1894), Martin (1952), Vance (1980a), Suzuki (2005)
According to generally accepted theory, Japanese has four vocabulary strata: native, mimetic, Sino-Japanese and foreign.

- Rendaku does not occur in the mimetic stratum.
- It occurs only occasionally (around 10% of the time) in the Sino-Japanese stratum.
- Rendaku does not occur in the foreign stratum (with only a very few exceptions: e.g. karuta, kappa, kanakiN, etc.).

Coordinate Compounds

Coordinate, or dvandva, compounds resist rendaku: \textit{goma.šio} ‘salt with sesame’ \textit{taN.seki} ‘coughing and phlegm’ \textit{cuki.hi} ‘time’

However, reduplicative coordinate compounds do undergo rendaku: \textit{kata.gata} ‘people’ \textit{hore.bore} ‘charming’ \textit{šina.jina} ‘goods’

That said, mimetic reduplicative coordinate compounds, as expected, resist: \textit{šiwa.šiwa} ‘wrinkled’ \textit{fusa.fusa} ‘bushy’ \textit{cuya.cuya} ‘glossy’

Rendaku is restricted to elements on the right branch of a constituent structure.

Thus, \text{nuri.gasa}.ire ‘case for lacquered umbrellas’ but \text{nuri.kasa.ire} ‘umbrella case which is lacquered’

In other words, rendaku is blocked when a compound is itself part of a larger compound: \text{fuyoo.teate} ‘family allowance’ \text{jidoo.furikomi} ‘automatic transfer’

When the initial element of a compound is a bare numeral, rendaku is typically resisted: *hito.te ‘single-handed’ yo.sumi ‘four corners’ nana.kusa ‘the seven spring plants’*

This is not the case with ‘full’ numerals or with Sino-Japanese numerals: *icucu.boši ‘5-star’ micu.gašiwa ‘buckbean’ seN.giri ‘thin strip’*

While other elements also have a dampening effect on rendaku, I will not consider their presence to constitute a systematic constraint. Such elements include: *kata ‘one of two’ hacu ‘first, initial’ kara ‘Cathay’*

Nakagawa (1966), Sato (1989), Ito (2008), Irwin (forthcoming)
A few NIEs contain an [m] derived from an original [b]: *kemuri* < *keburī* ‘smoke’ *himo* < *hibo* ‘string’

Here, rendaku is blocked

In the very few cases where an [m] alternates with a [b] in the modern language, rendaku may be found in NIEs with [m]: *samurai* ~ *saburai* ‘samurai’ *susamu* ~ *susabu* ‘grow wild’ *samu* ~ *sabu* ‘cold’

Nakamura (1966), Martin (1987)
Of the six systematic constraints just described, all exhibit exceptions, though usually to only a very minor extent: these will not be discussed further here.

Moreover, parts of speech undergo rendaku to varying degrees: resistance to rendaku is significantly high when an NIE is a verb – and particularly conspicuous when all elements in a compound are verbs.

It has also been claimed that rendaku is dampened when the NIE is a deverbal noun in an argument-type compound.

Okumura (1955), Vance (2005), Yamaguchi (2011)
Logic thus tells us that, in a non-dvandva compound which does not violate the right-branching rule and whose initial element is not a bare numeral, a native Japanese NIE which is not a verb and which does not contain a voiced obstruent...

[actually, despite all these constraints and conditions, there are thousands of such examples]

...should undergo rendaku

This is not, however, the case
Clearly there are other processes at play

It is known that rendaku varies by dialect (though little research has been done). Dialect borrowing cannot be ignored either

Sociolinguistic factors must also be at play: rendaku may vary by socio-economic group, by age cohort, by gender, etc., but there has been little research carried out here either

The only data we can usefully apply are headwords in dictionaries

Although individual Japanese speakers in the audience may disagree with the rendaku status of some of the headwords to be introduced, it is on dictionary headwords that the bulk of previous research – and the research I will outline today – must rely

Individual NIEs undergo rendaku to varying degrees

Some NIEs are rendaku immune – they never undergo rendaku: susuki ‘plume grass’, saki ‘future, tip’, sumire ‘violet’, hama ‘beach’

Most undergo rendaku most of the time (kucuwa ‘horse bit’, same ‘shark’, kome ‘rice’), while a smaller number resist rendaku most of the time (take ‘mushroom’, ho ‘sail’, či ‘blood’)

For a few NIEs, however, the small number of compounds in which they appear means that it is difficult to assess to any acceptable level of significance which way they lean.

A few compounds even have alternative pronunciations, where the same NIE has a rendaku and a non-rendaku alteration: *mizu.tori* ~ *mizu.dori* ‘waterfowl’ *cuke.shiru* ~ *cuke.jiru* ‘dipping sauce’ *jimu.cukue* ~ *jimu.zukue* ‘office desk’
Rosen, in his 2001 doctoral thesis, coined the terms ‘rendaku lover’ for a morpheme which undergoes rendaku more than 67% of the time; and ‘rendaku hater’ for a morpheme which undergoes rendaku less than 33% of the time.

Provided NIEs are restricted to native nouns, Rosen also claimed there was ‘an apparent tendency for an [NIE] to either strongly prefer to voice, or else to resist voicing, with almost no [NIEs] occupying a middle ground between the two tendencies’ and that rendaku lovers far outnumber rendaku haters.

This ‘apparent tendency’ represents an inverse logistic curve, or inverse S-curve.
There are 2 crucial points to note about Rosen’s thesis

Firstly, it must be pointed out in the interests of fairness that setting out to demonstrate the various proportions of rendaku haters and rendaku lovers was not Rosen’s main programme – instead it was an OT analysis of rendaku within a prosodic size criterion.

Secondly, the database Rosen employed was small: just over 800 compounds containing 143 different NIEs, drawn from the NHK accent dictionary, a Japanese-English dictionary, and a freeware programme.

Although he restricted his analysis to native nouns, with the exception of Motori/Lyman’s Law, he took no account of the various systematic constraints just outlined.
Irwin (2009) demonstrated that the prosodic size rule proposed by Rosen – the core of his thesis – holds up to detailed empirical scrutiny and appears robust.

Whether Rosen’s claim of ‘an apparent tendency for an [NIE] to either strongly prefer to voice, or else to resist voicing, with almost no [NIEs] occupying a middle ground between the two tendencies’ – i.e. the existence of a rendaku lover/hater inverse logistic curve and the lack of any ‘rendaku waverers’ – will also hold up to detailed empirical scrutiny is the goal of this presentation.
Polysemy

- Nearly all accounts of rendaku, including Rosen’s, fail to take into account polysemy.

- Polysemy = ‘a lexical item with a range of different meanings’

- Distinguishing a polyseme from a homophone can be theoretically problematic.

Croft & Cruse (2004), Murphy (2010), Cruse (2011)
‘program(me)’

‘a booklet giving details of items or performers at an event’

‘an item broadcast between stated times on TV or radio’

‘a series of coded software constructions’

‘a planned series of future events or performances’

‘a radio or TV channel’
‘手’

‘te’

‘possession’

‘someone who helps, care’

‘the upper limb’ (= hand, arm)

‘means, device, ruse, move (in games)’

‘-er’
In English, polysemes are only infrequently distinguished orthographically: *programme* v. *program* in British English is a rare example.

In Japanese, differing orthographies are more commonly found than in English:

- šio 潮 ‘tide’ v. 塩 ‘salt’
- cucu 筒 ‘tube’ v. 銃 ‘gun’
- kawa 革 ‘leather’ v. 皮 ‘skin, film’

But identical orthographies are still frequent:

- hi 日 ‘day’ v. hi 日 ‘sun’
- ki 木 ‘tree’ v. ki 木 ‘wood, lumber’
- keta 柱 ‘beam, column’ v. keta 柱 ‘digit, figure’

Masuda (1979)
Two different Japanese polysemes may differ considerably in their rendaku behaviour when they appear as an NIE in a compound:

- **kawa** ‘leather’ 12% v. ‘skin, film’ 72%
- **kuči** ‘mouth’ 72% v. ‘entrance’ 87%

Or they may not:

Such behaviour is independent of orthography.

Polysemy cannot be ignored in any statistical analysis of NIEs.
Homophony

As a brief aside, homophones (when accent is ignored) may also display broadly similar rendaku rates when NIEs in a compound:

- **hana** 花 ‘flower’ 90% v. **鼻** ‘nose’ 89%
- **kaki** 柿 ‘persimmon’ 100% v. **牡蠣** ‘oyster’ 100%
- **kami** 神 ‘god’ 92% v. **紙** ‘paper’ 98%

Or they may not:

- **koi** 恋 ‘love’ 44% v. **鯉** ‘carp’ 100%
- **taka** 鷹 ‘hawk’ 39% v. **高** ‘amount’ 95%
- **su** 巣 ‘nest’ 0% v. **酢** ‘vinegar’ 90%

There is no evidence that rendaku operates to differentiate homophony
The Rendaku Database

- The rendaku database is part of NINJAL’s ‘The Japanese Lexicon: A Rendaku Encyclopaedia’ Collaborative Research Project, of which I am a member.

- The database currently contains 33,504 entries, all of which are compounds possessing a potential rendaku site.

- Entries are taken from two major dictionaries and tagged for a range of criteria, including vocabulary stratum, length, accent and part of speech.

- Although further dictionaries and tags are planned in the future, the rendaku database is already the largest ever compiled.
In order to obtain an accurate analysis of NIE rendaku rates, any compounds subject to the systematic constraints described earlier will be removed.

Additionally...

1. Since rendaku varies according to part of speech, and Rosen’s study dealt only with nouns, compounds which are not nouns will be excised.

2. Many NIEs appear in fewer than 10 compounds. This is too few to come to an accurate conclusion regarding their rendaku rate, and these NIEs will be removed.

3. Since, cross-linguistically, names often exhibit aberrant behaviour, compounds which are names (but names only) will be excised: thus *yama.guchi* (a name but also a regular noun) will be retained, while *yama.zaki* (a name only) will be deleted.
Once all these trimmings are carried out, there remains the **NOUN DATABASE**

To recap: **NOUN DATABASE** compounds…

- are neither right-branching nor coordinate
- have native stratum noun NIEs containing no voiced obstruent (nor an /m/ < original /b/) which appear in 10 or more compounds
- do not contain bare numeral initial elements
- are not solely names
NOUN DATABASE Overview

- Size: 15,692 entries (Rosen = 800), being 47% of the original rendaku database
- Average rendaku rate: 79.7%
- Number of different NIEs: 300 (Rosen = 143)
- Average number of compounds in which an NIE appears: 52 (Rosen = 6)
- If the frequency (= number of compounds in which they appear) of these NIEs are plotted against their average rendaku rate, we obtain the following graph
Rendaku waverers include *kata* ‘shoulder’ *šimo* ‘frost’ *kaNna* ‘plane’ *kawa* ‘river’ *komo* ‘straw matting’ *kuki* ‘stem’ *seki* ‘dam, barrier’ *torii* ‘shrine archway’

If the rendaku rates of the 300 NIEs are combined into 5% blocks and stacked cumulatively, we do not find, as predicted by Rosen, an inverse logistic curve.

Instead, we find an exponential curve.
NOUN DATABASE NIE rendaku rates by 5% block, cumulative

Rosen’s prediction (inverse logistic curve)

2-period moving average

number of NIEs

Rendaku Rate <
Rendaku & the Mental Lexicon

- Not a few of the compounds in the NOUN DATABASE are obscure or obsolete and an obvious objection is that its content does not correspond closely to a native Japanese speaker’s mental lexicon.

- I have thus created a second smaller NOUN (FREQ) DATABASE.

- All compounds in the NOUN (FREQ) DATABASE appear in both dictionaries and all NIEs appear in 5 or more compounds.

- The NOUN (FREQ) DATABASE contains 240 NIEs appearing in 4,087 compounds (26% the size of the NOUN DATABASE).

- Its average rendaku rate is 78.7% (NOUN DATABASE = 79.7%) and the number of different compounds in which an NIE appears is 17 (NOUN DATABASE = 52).

- Exactly as before, if the frequency (= number of compounds in which they appear) of these 240 NIEs is plotted against their average rendaku rate, we obtain the following graph.
NOUN (FREQ) DATABASE NIE rendaku rates and frequencies

13% of NIEs
13% of NOUN (FREQ) DATABASE

12%
11%

10% of NIEs
10% of NOUN (FREQ) DATABASE

10%
6%

77% of NIEs
77% of NOUN (FREQ) DATABASE

77%
83%
Rendaku waverers include *kuči* ‘beginning of something’ *hata* ‘flag’ *cuyu* ‘rainy season’ *toši* ‘year’ *furui* ‘strainer’ *ki* ‘tree’ *ki* ‘lumber’

Once again, as before, if the rendaku rates of the 240 NIEs are combined into 5% blocks and stacked cumulatively, we do not find, as predicted by Rosen, an inverse logistic curve

Instead, we find an exponential curve
Conclusions

- Rosen’s hypothesis that rendaku rates follow an inverse logistic curve is incorrect – instead rendaku rates follow an exponential curve.

- As predicted by Rosen, rendaku lovers far outnumber rendaku haters – by a ratio of approximately 6:1.

- Rendaku ‘waverers’ make up approximately 10% of all NIEs and are only slightly fewer in number than rendaku haters – a far cry from ‘almost no [NIEs] occupying a middle ground between the two tendencies’.

- The frequency of a compound appears to play no role: after polysemy, as well as the various processes which constrain rendaku are taken into account, rendaku in nouns occurs approximately 80% of the time, with approximately 77% of all nouns being rendaku lovers.
Unresolved Issues with Polysemy

・ Division of dictionary entries by polyseme is unquestionably problematic

・ Although in many cases the various sub-entries of a major dictionary can be followed, in some cases there is scope for amalgamation, or for further division. Any such refinements are likely to be highly subjective

・ Moreover, it is not always clear if a given dictionary entry should be divided into polysemes at all, or if a polyseme is in fact really a homophone

・ In spite of these issues, it is unlikely that differing polysemic perspectives will, when analysing NIE rendaku rates overall, give rise to an inverse logistic, rather than an exponential, curve


Murphy, M. Lynne. 2010. Lexical meaning. Cambridge: CUP.


