The Old Japanese Rendaku Database and the Strong Version of Lyman’s Law

Timothy J. Vance

17th International Conference of JSLS
B-Con Plaza
Beppu, Japan
18 July 2015
The Old Japanese Rendaku Database and the Strong Version of Lyman’s Law

Timothy J. Vance
Introduction
In modern Tokyo ("standard") Japanese, many morphemes have one allomorph that begins with a voiceless obstruent and another allomorph that begins with a voiced obstruent.
For example:

寺 /tera/ ‘(Buddhist) temple’
寺男 /tera+otoko/ ‘temple assistant’
山寺 /yama+dera/ ‘mountain temple’

The initial voiced obstruent in the /dera/ allomorph is an instance of *rendaku* (連濁) ‘sequential voicing’.
A constraint on *rendaku* known as Lyman’s Law says that rendaku does not occur in a 2-element compound if the second element contains a medial voiced obstruent.
For example:

亀 /kame/ ‘turtle’
海亀 /umi+game/ ‘sea turtle’
風 /kaze/ ‘wind’
海風 /umi+kaze/ ‘sea breeze’
The Strong Version of Lyman’s Law
It is often claimed that *rendaku* did not occur in 8th-century Old Japanese (OJ) if either the first element (E1) or the second element (E2) of a 2-element compound contained a medial voiced obstruent.

This stricter constraint is known as the **strong version of Lyman’s Law**.
The strong version of Lyman’s Law clearly does not hold in modern Tokyo Japanese.
For example:

袖 /sode/ ‘sleeve’
口 /kuči/ ‘mouth’
袖口 /sode+guči/ ‘cuff’
The Domain of the Strong Version
I use the Frellesvig and Whitman romanization for OJ forms.

The 3-element OJ compound $sode + tuke + goromo$ ‘sleeved garment’ raises an interesting question.
As independent words, the three elements are:

{sode} ‘sleeve’
{tuke} ‘attaching’
{koromo} ‘garment’

And the constituent structure is clearly \{\{sode+tuке\}+goromo\}. 
Thus, in relation to the consonant that shows *rendaku* (the *g* in *goromo*), the earlier voiced obstruent *d* is not in the immediately preceding syllable and is not in the immediately preceding morph:

\[ \{\text{so} \text{de} + \text{tuke}\} + \text{goromo} \]
If the strong version of Lyman’s Law was a local effect in OJ, then \textit{sode+tuke+goromo} is not a counter-example.

It is not clear, however, whether the relevant domain included the entire preceding morpheme or just the preceding syllable.
Examples that would help to determine the domain of the strong version of Lyman’s Law are extremely rare.

We would like to find 2-element compounds in which $E_1$ contains a voiced obstruent in a non-final syllable.
For the most part, however, OJ did not allow word-initial voiced obstruents, so the search is limited to compounds with E1s longer than two syllables.

But very few monomorphemic elements in OJ were that long.
Relevant compounds would have the form:

\[
\cdot \cdot \cdot V D V C V + X V \cdot \cdot \cdot
\]

\[
\underline{E1} \quad \underline{E2}
\]

D = any voiced obstruent
C = any other kind of consonant
X = potential *rendaku* site
If X shows *rendaku* in such examples, the inhibiting domain is the immediately preceding syllable.
If X consistently lacks *rendaku* in such examples, the inhibiting domain is the immediately preceding morph.
The Old Japanese Rendaku Database
The Old Japanese Rendaku Database (OJRDB) is the product of a collaboration between the Rendaku Project at NINJAL and the Oxford Corpus of Old Japanese (OCOJ) project at the University of Oxford in the UK.
Welcome to the Oxford Corpus of Old Japanese!

The Oxford Corpus of Old Japanese (abbreviated OCOJ) is a long-term research project which aims to develop a comprehensive annotated digital corpus of all extant texts in Japanese from the Old Japanese period. Old Japanese is the earliest attested stage of the Japanese language, largely the Japanese language of the Asuka and Nara periods of Japanese history (7th and 8th century AD). This is the formative literate period upon which the development of Japanese civilization is based, and these texts are of paramount importance for the study and understanding of the origins and development of civilization of Japan, including language, writing, literature, religion, history, and culture.
My collaborator at Oxford is Kerri Russell, who did a Ph.D. in Japanese historical linguistics at the University of Hawai‘i.
The first preliminary version of the OJRDB went online in April of this year.

It has since been revised twice, and the last preliminary version (version 0.3) can be accessed at:

http://pj.ninjal.ac.jp/rendaku/database/
Welcome to The Japanese Lexicon: A Rendaku Encyclopedia

The Japanese Lexicon: A Rendaku Encyclopedia

Research field: linguistics
Keywords: rendaku, morphophonemic alternations, history of Japanese

What's New

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015.05.22</td>
<td>A new version (version 0.2) of the Old Japanese Rendaku Database (OJRDB) is now available.</td>
</tr>
<tr>
<td>2015.04.21</td>
<td>A preliminary version (version 0.1) of the Old Japanese Rendaku Database (OJRDB) is now available. The OJRDB is the outcome of a collaboration between Timothy J. Vance at NINJAL and Kerri M. Russell at the University of Oxford. Improved versions of the OJRDB will be uploaded frequently as checking and correcting proceed.</td>
</tr>
<tr>
<td>2015.04.17</td>
<td>Website is now available!</td>
</tr>
</tbody>
</table>

Project leader

Timothy J. VANCE
The OJRDB is an Excel file, and the compounds it lists as entries were taken from the standard dictionary of OJ (published in 1967).
時代別国語大辞典
上代編

三省堂
The first step in compiling the OJRDB was to examine the headwords in this dictionary and extract all the compounds with an E2 that begins with a voiceless obstruent as an independent word.
Then, for every such compound, each attestation in the OCOJ was labeled as either logographic or phonographic.

The current version of the OJRDB includes only compounds that have at least one phonographic attestation in the OCOJ.
For example, *yama*+*bito* ‘mountain person’ is attested 3 times, twice logographically and once phonographically.

(1) 山人 [logographic]
(2) 夜麻纻等 [phonographic]
(3) 山人 [logographic]
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
<th>P</th>
<th>Q</th>
<th>R</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PH</td>
<td>PH</td>
<td>E1</td>
<td>K1</td>
<td>E2</td>
<td>K2</td>
<td>C(-3) V(-3) C(-2) V(-1) C(+1) V(+2) C(+3) V(+4)</td>
<td>+R</td>
<td>-R</td>
<td>GLOSS</td>
<td>OCOJ LINK</td>
<td>notes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>mara</td>
<td>稀</td>
<td>pilo</td>
<td>mara</td>
<td>r</td>
<td>a</td>
<td>p</td>
<td>t</td>
<td>t</td>
<td>0</td>
<td>1</td>
<td>rare person</td>
<td>mara+pilo</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>yama</td>
<td>山</td>
<td>pilo</td>
<td>ma</td>
<td>y</td>
<td>a</td>
<td>m</td>
<td>a</td>
<td>p</td>
<td>t</td>
<td>t</td>
<td>0</td>
<td>0</td>
<td>mountain person</td>
<td>yama+bito</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>two</td>
<td>外</td>
<td>pilo</td>
<td>t</td>
<td>o</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>outsider</td>
<td>two+pilo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The JAWA-base entry for this compound is...
<table>
<thead>
<tr>
<th>phonographic attestations</th>
<th>logographic attestations</th>
</tr>
</thead>
<tbody>
<tr>
<td>phonographic with <strong>rendaku</strong></td>
<td>phonographic without <strong>rendaku</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>+PH</th>
<th>-PH</th>
<th>E1</th>
<th>K1</th>
<th>E2</th>
<th>K2</th>
<th>C(-4)</th>
<th>V(-3)</th>
<th>C(-2)</th>
<th>V(-1)</th>
<th>C(+1)</th>
<th>V(+2)</th>
<th>C(+3)</th>
<th>V(+4)</th>
<th>+R</th>
<th>-R</th>
<th>GLOSS</th>
<th>OCOJ LINK</th>
<th>notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>231</td>
<td>0</td>
<td>mara</td>
<td>pito</td>
<td>mara</td>
<td>pito</td>
<td>m a r a</td>
<td>p i t o</td>
<td>0 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rare person</td>
<td>mara+pito</td>
<td>mara=mare (父音助)</td>
</tr>
<tr>
<td>232</td>
<td>0</td>
<td>miyakokekata</td>
<td>pito</td>
<td>miyakokekata</td>
<td>pito</td>
<td>k a t a</td>
<td>p i t o</td>
<td>0 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>capital-area person</td>
<td>miyakokekata+pito</td>
<td></td>
</tr>
<tr>
<td>233</td>
<td>1</td>
<td>yama</td>
<td>pito</td>
<td>yama</td>
<td>pito</td>
<td>y a m a</td>
<td>p i t o</td>
<td>0 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>mountain person</td>
<td>yama+pito</td>
<td></td>
</tr>
<tr>
<td>234</td>
<td>1</td>
<td>ta</td>
<td>pito</td>
<td>ta</td>
<td>pito</td>
<td>t a</td>
<td>p i t o</td>
<td>0 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>current</td>
<td>ta+pito</td>
<td></td>
</tr>
<tr>
<td>235</td>
<td>2</td>
<td>two</td>
<td>pito</td>
<td>two</td>
<td>pito</td>
<td>t o</td>
<td>p i t o</td>
<td>0 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>outsider</td>
<td>two+pito</td>
<td></td>
</tr>
</tbody>
</table>

The **rendaku** entry for this compound is...
link to the Oxford Corpus
<table>
<thead>
<tr>
<th>Simple Forms</th>
<th>Text</th>
<th>Trees</th>
</tr>
</thead>
<tbody>
<tr>
<td>yamabito</td>
<td>MYS.20.4293</td>
<td>MYS.20.4293</td>
</tr>
<tr>
<td>yamabito</td>
<td>MYS.20.4294</td>
<td>MYS.20.4294</td>
</tr>
<tr>
<td>yamabito</td>
<td>MYS.20.4294</td>
<td>MYS.20.4294</td>
</tr>
</tbody>
</table>
山人 [logographic]
<table>
<thead>
<tr>
<th>Simple Forms</th>
<th>Text</th>
<th>Trees</th>
</tr>
</thead>
<tbody>
<tr>
<td>yamabito</td>
<td>MYS.20.4293</td>
<td>MYS.20.4293</td>
</tr>
<tr>
<td>yamabito</td>
<td>MYS.20.4294</td>
<td>MYS.20.4294</td>
</tr>
<tr>
<td>yamabito</td>
<td>MYS.20.4294</td>
<td>MYS.20.4294</td>
</tr>
<tr>
<td></td>
<td>安之比奇能</td>
<td>asipikwi no</td>
</tr>
<tr>
<td>---</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td>2</td>
<td>山爾由伎家牟</td>
<td>yama ni yukikye-mu</td>
</tr>
<tr>
<td>3</td>
<td>夜麻妣等能</td>
<td>yamabito no</td>
</tr>
<tr>
<td>4</td>
<td>情母之良受</td>
<td>kokoro mo sirazu</td>
</tr>
<tr>
<td>5</td>
<td>山人夜多礼</td>
<td>yamabito ya tare</td>
</tr>
</tbody>
</table>

夜麻妣等 [phonographic]
**L050488 yamabito** mountain person

<table>
<thead>
<tr>
<th>Simple Forms</th>
<th>Text</th>
<th>Trees</th>
</tr>
</thead>
<tbody>
<tr>
<td>yamabito</td>
<td>MYS.20.4293</td>
<td>MYS.20.4293</td>
</tr>
<tr>
<td>yamabito</td>
<td>MYS.20.4294</td>
<td>MYS.20.4294</td>
</tr>
<tr>
<td>yamabito</td>
<td>MYS.20.4294</td>
<td>MYS.20.4294</td>
</tr>
</tbody>
</table>
山人 [logographic]
Testing Lyman’s Law
To test the strong version of Lyman’s Law, the first step is to sort the compounds in the OJRDB according to the last consonant in E1 — an easy operation in Excel.
For some OJ compounds, only E2 is written phonographically in the OCOJ attestations, whereas E1 is written logographically.

For example, $asi + gani$ ‘reed crab’ appears only once, written 蕲河尔, where 河尔 is phonographic for $hani$ but 蕲 is logographic for $asi$. 
Compounds like $asi + gani$, attested only with phonographic+logographic hybrid spellings, can be used in investigating the behavior of E2, but they should not be used to assess the influence of E1.
When examples with logographically written E1s are excluded, there are only 21 compounds in the OJRDB with a voiced obstruent in the last syllable of E1.
Among these 21, 3 also have a medial voiced obstruent in E2 and thus would be expected to resist *rendaku* anyway.

One of the 3 compounds with a voiced obstruent both in the last syllable of E1 and medially in E2 is *yorodu*+*tabi* ‘myriad times’.
Thus, there are only 18 compounds in the OJRDB that are unquestionably relevant in testing the inhibiting effect of a voiced obstruent in the last syllable of E1.

None has *rendaku*, as the strong version of Lyman’s Law predicts, but the sample is uncomfortably small.
The next step in testing the strong version of Lyman’s Law is to sort the compounds without a voiced obstruent in the last syllable of E1 by the second-to-last consonant in E1.

In many cases, there is no such consonant, as in *opo+kimi* ‘great monarch’.
Incidentally, the only OJRDB compound with a voiced obstruent in the antepenultimate syllable of E1 is 3-element \{\textcolor{orange}{sode}+\textcolor{yellow}{tuke}+\textcolor{green}{goromo}\} ‘sleevded garment’, which was cited earlier.
There are 4 compounds in the OJRDB with a voiced obstruent in the penultimate syllable of E1.

One of these 4 can be excluded from consideration because it has a medial voiced obstruent in E2.
The remaining 3 examples are $wagimo + kwo$ ‘my beloved’, $mugura + pu$ ‘weed patch’, and $madara + busuma$ ‘multicolored bedding’.

The first example is problematic because $wagimo$ is etymologically a contracted phrase (\(< wagimo < wa ga imo)$. 
The E2 *pu* ‘patch’ does not show *rendaku* consistently in other compounds.

Thus, the absence of *rendaku* in *mugura*+*pu* cannot be taken as evidence that the domain of the strong version of Lyman’s Law was the immediately preceding morph.
In contrast, the presence of *rendaku* in *madara*+*busuma* can be construed as evidence that the domain was the immediately preceding syllable.

It is, of course, just a single example.
Thank you for your kind attention.

MYS.14.3354

伎倍比等乃  
万太良夫須麻爾  
和多佐波太  
伊利奈麻之母乃  
伊毛我乎抒許爾

kipepito no  
madarabusuma ni  
wata sapada  
iri-na-masi mono  
imo ga wo-doko ni

ご清聴ありがとうございました