Title

Arisaka's Law and the Phonemic Analysis of Old Japanese

Journal

Japanese/Korean Linguistics, 31(1)

Author

Vance, Timothy

Publication Date

2025-07-07

DOI

10.5070/J7.49004

Arisaka's Law and the Phonemic Analysis of Old Japanese*

TIMOTHY J. VANCE
National Institute for Japanese Language and Linguistics

1 Introduction

Old Japanese (OJ) is the language of eighth-century documents, the earliest stage of Japanese for which we have substantial written records. These documents are somewhat heterogeneous, both geographically and temporally, and OJ orthography (Section 2) makes it very challenging to infer the OJ phonological system (Section 3). Some phonemic analyses include contrasts between consonant+vowel (CV) syllables and consonant+glide+vowel (CGV) syllables, where G is ^{OJ}/y/ or ^{OJ}/w/. For example, both Lange (1973: 124) and Frellesvig and Whitman (2008: 2–5) transcribe OJ with a contrast between ^{OJ}/Co/ and ^{OJ}/Cwo/. Arisaka's Law (Section 4) is a set of putative cooccurrence restrictions on OJ syllables within a root, and in an analysis that incorporates this contrast between ^{OJ}/Co/ and ^{OJ}/Cwo/, Arisaka's Law prohibits roots containing both ^{OJ}/Co/ and ^{OJ}/Cwo/. OJ also allowed glide+vowel (GV) syllables, including ^{OJ}/wo/, and it is natural to assume that ^{OJ}/wo/ in a GV syllable is phonemically the same as ^{OJ}/wo/ in a CGV syllable. If so, Arisaka's Law predicts that the GV syllable ^{OJ}/wo/ should not occur in a root that also contains a CV syllable of the form ^{OJ}/Co/, but this prediction is false (Section 5). This contradiction prompts consideration of alternative phonemic analyses that do not include ^{OJ}/Cwo/ syllables (Section 6).

* Many thanks to Jim Unger and John Whitman for their input

Japanese/Korean Linguistics 31.
Edited by William Giang, Lucien Brown, Shimako Iwasaki, Satoshi Nambu, and Daniel Pieper.

2 Old Japanese Orthography

Much of OJ is recorded phonographically, with Chinese characters used as phonograms. Such characters are called *man'yōgana* (万葉仮名), and each OJ syllable could be written with any one of several *man'yōgana*. For example, 侈, 多, 太, 大, 他, 陁, 哆, 駄, 党, 丹, 田, and 手 are all attested as *man'yōgana* for ^{OJ}/ta/, although not all in a single text.

In the late eighteenth century, 'national learning' scholars (*kokugakusha* 国学者) discovered 'spelling' distinctions in OJ texts. For some modern Japanese syllables, the corresponding OJ syllable was written with one or the other of two mutually exclusive sets of *man'yōgana*, depending on the lexical item involved. For example, the OJ counterpart of modern /no/ 野 'field' is attested with the *man'yōgana* 努, 怒, and 奴. In contrast, the OJ counterpart of the modern genitive particle /no/ is attested with the *man'yōgana* 乃 and 能. Note the lack of overlap between the two sets of *man'yōgana*.

3 Phonological Distinctions

In the early twentieth century, Hashimoto Shinkichi 橋本進吉 realized that the OJ orthographic distinctions described in Section 2 must have represented phonological distinctions: Some modern C/i/, C/e/, and C/o/ syllables correspond to two different OJ syllables (Hashimoto 1922). OJ is sometimes romanized using i_1 versus i_2, e_1 versus e_2, and o_1 versus o_2, and the majority view has been that these phonological distinctions were not in the initial consonants of the relevant syllables, but Lange (1973: 55–6) cautions against jumping to this conclusion. For any OJ CV_1 syllable and OJ CV_2 syllable corresponding to the same modern CV syllable, the difference could have been in the vocalic portions (including monophthongs, diphthongs, glide+vowel sequences, and vowel+ glide sequences), but it could just as well have been in the initial consonants or in both.

Primarily, it is reconstructed Middle Chinese syllables that provide the basis for inferences about the phonetic values of OJ syllables represented by *man'yōgana* (Lange 1973, Miyake 2003), and this evidence leads Lange to conclude that Ci_1 was more palatal than Ci_2; that Ce_1 was more palatal than Ce_2; and that Co_1 was more labial than Co_2 (Lange 1973: 120–4).[1] Lange phonemicizes i_2 , e_2, and o_2 as monophthongs (i_2 = ^{OJ}/i/, e_2 = ^{OJ}/e/, o_2 = ^{OJ}/o/) and i_1, e_1, and o_1 as GV sequences (i_1 = ^{OJ}/yi/, e_1 = ^{OJ}/ye/, o_1 = ^{OJ}/wo/) (Lange 1973: 124).[2] As he is careful to explain, however, the primary motivation for these choices is economy, that is, minimizing the number of phonemes in the OJ inventory. Since ^{OJ}/y/ and ^{OJ}/w/ are necessary anyway, analyzing the CV_1 versus CV_2 distinctions as /CGV/ versus /CV/ keeps the number of phonemes the same. Also, if the i_1/i_2, e_1/e_2, and o_1/o_2 distinctions were phonologically just vowel differences, OJ had an eight-vowel system after certain onset consonants but only a five-vowel system in onsetless syllables. A CV/CGV analysis avoids this typologically oddity.[3]

---

[1] Miyake (2003) proposes phonetic reconstructions of the vowels in these syllables that are compatible with Lange's (1973) characterizations: i_1 = [i], i_2 = [ɨ]; e_1 = [e], e_2 = [əj]; o_1= [o], o_2 = [ə].

[2] Lange uses ⟨j⟩ rather than ⟨y⟩ as a phonemic symbol, but I have modified his notation to simplify the comparison with Frellesvig-Whitman phonemic transcriptions.

[3] I acknowledge that I am relying on my own intuition that a phonemic system with more vowel distinctions in CV syllables than in V syllables is exceptional, that is, I do not know of any uncontroversial example of such system in the small subset of the world's languages with which I have the requisite familiarity. It is possible to analyze Russian as having a phonemic distinction between /i/ and /ɨ/ only in syllables with onsets, but the standard analysis is that [i] and [i̯i] are allophones of a single phoneme, with the former occurring predictably in a syllable with a palatalized onset

The romanization system that Frellesvig and Whitman (2008: 2–5) adopt for OJ reflects their phonemic analysis, which treats i_1, e_2, and o_2 as monophthongs and i_2, e_1, and o_1 as GV sequences: i_1 = ^{OJ}/i/, i_2 = ^{OJ}/wi/; e_1 = ^{OJ}/ye/, e_2 = ^{OJ}/e/; o_1 = ^{OJ}/wo/, o_2 = ^{OJ}/o/. Using the examples cited above in Section 2 to illustrate, the counterpart of the modern genitive particle was ^{OJ}/no/, while the counterpart of the modern noun meaning 'field' was ^{OJ}/nwo/. The difference between this analysis and Lange's is just that Lange treats i_1 and i_2 as ^{OJ}/yi/ versus ^{OJ}/i/ instead of ^{OJ}/i/ versus ^{OJ}/wi/. In addition to CGV syllables like ^{OJ}/nwo/, OJ had GV syllables, and as Frellesvig (2010: 32) notes, ^{OJ}/wi/, ^{OJ}/ye/, and ^{OJ}/wo/ were all possible (as were ^{OJ}/yu/, ^{OJ}/yo/, ^{OJ}/ya/, ^{OJ}/we/, and ^{OJ}/wa/). OJ also allowed onsetless V syllables word-initially, but there were no *man'yōgana* distinctions for any of these, and they are analyzed phonemically as ^{OJ}/i u e o a/.

Table 1 lists the Frellesvig-Whitman phonemic transcriptions of all the Ci_1, Ci_2, Ce_1, Ce_2, Co_1 and Co_2 syllables that Frellesvig (2010: 33) recognizes.

Ci_1	Ci_2	Ce_2	Ce_1	Co_2	Co_1
C/i/	C/wi/	C/e/	C/ye/	C/o/	C/wo/
/pi/	/pwi/	/pe/	/pye/	/po/	
/bi/	/bwi/	/be/	/bye/	/bo/	
/mi/	/mwi/	/me/	/mye/	/mo/	/mwo/
/ti/		/te/		/to/	/two/
/di/		/de/		/do/	/dwo/
/si/		/se/		/so/	/swo/
/zi/		/ze/		/zo/	/zwo/
/ni/		/ne/		/no/	/nwo/
/ri/		/re/		/ro/	/rwo/
				/yo/	/ywo/
/ki/	/kwi/	/ke/	/kye/	/ko/	/kwo/
/gi/	/gwi/	/ge/	/gye/	/go/	/gwo/

Table 1: Old Japanese CV versus CGV (Frellesvig-Whitman Phonemic Analysis)

Lange (1973: 99–104), however, adopts more stringent criteria for recognizing each CV_1 versus CV_2 distinction and does not accept the six shaded in gray in Table 1, transcribing them all as ^{OJ}/Co/. The presence or absence of the six shaded ^{OJ}/Cwo/ syllables has no significant effect on the arguments in the remainder of this paper, but it should be noted that ^{OJ}/ywo/ is particularly suspect. A syllable of this form is not impossible, perhaps (Wonderly 1951: 107, 109), but it is certainly highly unusual. Furthermore, Unger (2018) argues convincingly for an etymology that depends crucially on the absence of ^{OJ}/ywo/.

or with no onset (Sussex 2003: 482). Broad typological surveys of phonological systems seldom if ever provide enough detail about phonotactics to compare the phonemic vowel distinctions in syllables with and without onsets.

4 Arisaka's Law

In the early 1930s, Arisaka Hideyo 有坂秀世 noted what he took to be cooccurrence restrictions on Co_2 syllables in OJ roots (Ōno 1977: 203–6). His three putative generalizations are stated below in (1).

(1) a. Co_2 and Co_1 do not cooccur within a root.
b. Co_2 and C*u* do not cooccur within a root.
c. Co_2 and C*a* seldom cooccur within a root.

Although sometimes described as three 'laws', I follow Martin (1987: 58) and Frellesvig (2010: 44) and treat the generalizations in (1) as a single 'law' with three subcases. Arisaka suggested that these cooccurrence restrictions are the vestiges of some kind of prehistoric vowel harmony system (Lange 1973: 53, Frellesvig 2010: 44), and Whitman (2023) has proposed that this system might have required agreement in [±retracted tongue root] (i.e. pharyngeal constriction; Morén-Duolljá 2011: 450–2). Francis-Ratte (2016: 64–73) argues, however, that the proto-Japanese ancestor of o_2 (pronounced [ə]) lowered to [a] (proto-Japanese */a/) in the vicinity of proto-Japanese */a/, */o/, or */u/ within a word, resulting in the Arisaka's Law cooccurrence restrictions in (1).

5 Word-Initial Syllables

Of the three putative subcases of Arisaka's Law in (1), Lange (1973: 57) accepts only (1a) as valid in OJ, and the remainder of this paper will leave (1b) and (1c) aside. It seems reasonable to broaden (1a) slightly and assume that it prohibits o_1 and o_2 within the same OJ root regardless of whether or not either is preceded by an onset consonant. If so, a simple-minded interpretation of the Lange and Frellesvig-Whitman phonemic analyses leads to some testable predictions about roots containing a V syllable (i.e. ^{OJ}/o/) and/or a GV syllable (i.e. ^{OJ}/wo/). (As noted above in Section 3, V syllables occur only word-initially.) Since very few OJ roots were longer than two syllables (Unger 1975: 8), I consider only two-syllable roots below. The predictions are that there should be no OJ roots of the four types in (2).

(2) a. ⊘ ^{OJ}/woCo/ o_1Co_2
b. ⊘ ^{OJ}/Cowo/ Co_2o_1
c. ⊘ ^{OJ}/oCwo/ o_2Co_1
d. ⊘ ^{OJ}/owo/ o_2o_1

The *Oxford-NINJAL Corpus of Old Japanese* (Frellesvig, Horn et al. 2023) can be used to test these predictions against the attested OJ lexicon. Fortunately, the corpus dictionary includes roots that are attested only as elements of longer words. Following Lange (1973: 121), the only /Cwo/ syllables considered are ^{OJ}/swo/, ^{OJ}/nwo/, ^{OJ}/kwo/, and ^{OJ}/gwo/ (see Table 1 in Section 3). While the number of relevant roots is small, predictions (2c) and (2d) seem to hold true, but predictions (2a) and (2b) do not, as (3) shows.

(3) a. ^{OJ}/woCo/: 6 roots
^{OJ}/woko/ 'foolish'
^{OJ}/woso/ 'lie'
^{OJ}/woto/ 'young'
^{OJ}/woto/ 'distant'
^{OJ}/wodor/ 'to jump'
^{OJ}/wono/ 'axe'

b. ^{OJ}/Cowo/: 2 roots
^{OJ}/towo/ 'ten'
^{OJ}/towo/ 'full'

c. ^{OJ}/oCwo/: 0 roots

d. ^{OJ}/owo/: 0 roots

As noted in Section 3, Lange (1973: 99–104) does not accept several of the Co_1/Co_2 distinctions that most other scholars recognize, including ^{OJ}/two/ versus ^{OJ}/to/ and ^{OJ}/dwo/ versus ^{OJ}/do/, but he would transcribe the relevant syllables in (3) as /to/ and /do/. The authoritative print dictionary of OJ (Jōdai-go Jiten Henshū Iin-kai 1967) and the *Oxford-NINJAL Corpus of Old Japanese* both recognize ^{OJ}/two/ versus ^{OJ}/to/ and ^{OJ}/dwo/ versus ^{OJ}/do/, and both have ^{OJ}/o/ (i.e., o_2) following ^{OJ}/t/ or ^{OJ}/d/ in the roots in (3).

In contrast to the examples in (3), the *Oxford-NINJAL Corpus of Old Japanese* contains no exceptions to the first subcase of Arisaka's Law as originally stated. That is, (1a) above in Section 4 holds for Co_1 and Co_2 syllables: There are no roots of the form ^{OJ}/CwoCo/ or ^{OJ}/CoCwo/. The absence of roots of the form ^{OJ}/oCwo/ (3c) and ^{OJ}/owo/ (3d) is consistent with the idea that root-initial ^{OJ}/o/ syllables behave like ^{OJ}/Co/ syllables with respect to Arisaka's Law, which means that identifying word-initial ^{OJ}/o/ with o_2 is unproblematic. On the other hand, the existence of roots of the form ^{OJ}/woCo/ (3a) and ^{OJ}/Cowo/ (3b) contradicts the idea that ^{OJ}/wo/ syllables behave like ^{OJ}/Cwo/ syllables with respect to Arisaka's Law. It thus appears that ^{OJ}/wo/ syllables cannot be identified as instances of o_1. The upshot is that, in the phonemic transcriptions adopted by Lange (1973) and by Frellesvig and Whitman (2008), ^{OJ}/wo/ without an immediately preceding consonant does not represent the same phonological entity as ^{OJ}/wo/ in ^{OJ}/Cwo/ syllables.

6 Summary and Conclusions

This paper has tested the predictions that follow from the two premises in (4).

(4) a. Within an OJ root, o_1 and o_2 do not cooccur.
b. Phonemically, o_1 is ^{OJ}/wo/ and o_2 is ^{OJ}/o/.

The first premise (4a) is a natural interpretation of the first subcase of Arisaka's Law (1a). The second premise (4b) is a natural interpretation of the phonemic analyses adopted by Lange (1973: 124) and by Frellesvig and Whitman (2008: 2–5). As the examples cited in Section 5 show, the predictions that follow from (4a) and (4b) are false, which means that there is a problem with one or both of these premises. I do not see any reason to abandon (4a), and if we accept (4a), (4b) must be faulty. The remainder of this section briefly considers phonological analyses that do not treat ^{OJ}/wo/ syllables as instances of o_1.

One possibility is some version of the traditional eight-vowel analysis. For example, in line with the phonetic values reconstructed by Miyake (2003) (see Footnote 1 above in Section 3), we could phonemicize o_1 as ^{OJ}/o/ (not ^{OJ}/wo/) and o_2 as ^{OJ}/ə/. On this analysis, the roots in (3a) and

(3b) are not counterexamples to Arisaka's Law because there is no o_1/o_2 distinction in the GV syllables, allowing them to be treated as ^{OJ}/wə/. The argument I gave in Section 3 against eight-vowel analyses was that it is typologically odd to have a larger vowel inventory following certain onset consonants than in onsetless syllables. On the other hand, as also noted in Section 3, the Frellesvig-Whitman analysis of i_2, e_1, and o_1 as GV sequences means that, of the eight GV sequences that occur as syllables (^{OJ}/yu/, ^{OJ}/ye/, ^{OJ}/yo/, ^{OJ}/ya/, ^{OJ}/wi/, ^{OJ}/we/, ^{OJ}/wo/, ^{OJ}/wa/), only three occur in CGV syllables (^{OJ}/Cye/, ^{OJ}/Cwi/, ^{OJ}/Cwo/). Perhaps an eight-vowel analysis is not so bad—typologically unusual but not impossible. After all, the OJ vowel system was unstable, and the CV_1/CV_2 distinctions had collapsed by 800, leaving Early Middle Japanese (800–1200) with essentially the same five short vowels as modern Japanese.

Another possibility is that the CV_1/CV_2 distinctions were in the onset consonants. For example, if the glides in Frellesvig-Whitman CGV syllables are reanalyzed as secondary articulations, Ci_2 would be ^{OJ}/C^{w}i/ (instead of ^{OJ}/Cwi/), Ce_1 would be ^{OJ}/C^{j}e/ (instead of ^{OJ}/Cye/), and Co_1 would be ^{OJ}/C^{w}o/ (instead of ^{OJ}/Cwo/). This analysis gives OJ a five-vowel inventory across the board. At the price of increasing the inventory of consonant phonemes, it also simplifies OJ phonotactics by eliminating CGV syllables.

There are no doubt many other possible phonemic analyses of OJ that are worth considering, but the kind of evidence that might help to resolve many uncertainties will remain forever beyond our grasp. We can only infer the phonetic basis for the phonemic distinctions, and we will never have the psychological insights that sometimes come from phoneme transpositions, either accidental (as in speech errors) or deliberate (as in language games).

References

Arisaka, H. 1934. Kodai Nihongo ni okeru onsetsu ketsugō no hōsoku (Syllable cooccurrence rules in Old Japanese). *Kokugo to Kokubungaku* 11(1). Reprinted in 1955 in *Jōdai on'in kō* (Thoughts on Old Japanese phonology), 1–106. Tokyo: Sanseidō.

Francis-Ratte, A. T. 2016. *Proto-Korean-Japanese: A New Reconstruction of the Common Origin of the Japanese and Korean Languages*. Doctoral dissertation, The Ohio State University.

Frellesvig, B. 2010. *A History of the Japanese Language*. Cambridge: Cambridge University Press.

Frellesvig, B., S. W. Horn et al. (eds.). 2023. *Oxford-NINJAL Corpus of Old Japanese.* Accessed February 2025 from https://oncoj.ninjal.ac.jp/

Frellesvig, B. and J. Whitman. 2008. Introduction. *Proto-Japanese: Issues and prospects*, eds. B. Frellesvig and J. Whitman, 1–9. Amsterdam: John Benjamins.

Hashimoto, S. 1922. Kokugo kanazukai kenkyū-shi-jō no ichi-hakken (A discovery in the history of research on Japanese kana usage). *Teikoku Bungaku* 23(11). Reprinted in 1949 in *Moji oyobi kanazukai no kenkyū* (Research on writing and kana usage), 123–63. Tokyo: Iwanami Shoten.

Jōdai-go Jiten Henshū Iin-kai (ed.). 1967. *Jidai-betsu kokugo dai-jiten: Jōdai-hen* (Great dictionary of the Japanese language by era: Old Japanese). Tokyo: Sanseidō.

Lange, R. A. 1973. *The Phonology of Eighth-Century Japanese*. Tokyo: Sophia University.

Martin, S. E. 1987. *The Japanese Language Through Time*. New Haven: Yale University Press.

Miyake, M. H. 2003. *Old Japanese: A Phonetic Reconstruction*. London: RoutledgeCurzon.

Morén-Duolljá, B. 2011. Vowel Place. *The Blackwell Companion to Phonology*, eds. M. van Oostendorp, C. J. Ewen, E. Hume, and K. Rice, 440–64. Chichester: Wiley-Blackwell.

Ōno, S. 1977. On'in no hensen (1) (Phonological changes (1)). *Iwanami kōza Nihongo 5: On'in* (Iwanami course on the Japanese language 5: Phonology), eds. S. Ōno and T. Shibata, 148–219. Tokyo: Iwanami.

Sussex, R. 2003. Russian. *International Encyclopedia of Linguistics*, 2nd ed., ed. W. J. Frawley, 481–9. Oxford: Oxford University Press.

Unger, J. M. 1975. *Studies in Early Japanese Morphophonemics*. Doctoral dissertation, Yale University.

Unger, J. M. 2018. How Many OJ Syllables Are Reflected in EMJ *yo*? *Studies in Japanese and Korean Theoretical Linguistics and Beyond: Festschrift Presented to John B. Whitman*, eds. W. McClure and A. Vovin, 76–9. Leiden: Brill.

Whitman, J. 2023. Japanese and Korean Harmonic Vocalism Reconsidered. Paper presented at the *30th Japanese/Korean Linguistics Conference*, Vancouver, March 11, 2023.

Wonderly, W. L. 1951. Zoque II: Phonemes and Morphophonemes. *International Journal of American Linguistics* 17: 105–23.