
Vowel sequences in Old Japanese: from a corpus-based approach

Chihkai Lin

National Taiwan University of Science and Technology
No.43, Keelung Rd., Sec.4, Da'an Dist.,
Taipei City 106335, Taiwan (R.O.C.)
linchihkai@gmail.com

Abstract

This paper investigates vowel sequences in Old Japanese from a corpus-based approach by using data from the Oxford-NINJAL Corpus of Old Japanese (ONCOJ). Three conditions are taken into consideration to analyze the corpus data, and they are (a) within a phonological word, (b) through morphological process, and (c) by grammatical element. The results have shown that vowel sequences in Old Japanese are not random combinations of vowels, but they are phonologically constrained. In the first and second conditions, the vowel sequences tend to show vowel_[-front] + vowel_[-low]. In the third condition, the first vowel of the vowel sequence is -i-. In addition to frequency, this paper also discusses the types of words and compares types with frequencies.

1 Introduction

This paper investigates vowel sequences in Old Japanese from a corpus-based approach. One of the phonological characteristics of Old Japanese is that the basic syllable is open, CV, and V (Vovin, 2005; Fellesvig, 2010). While CV and V syllables are legitimate in Old Japanese, contact of two vowels generates a sequence like CV.V or V.V, which is uncommon in Old Japanese.

To avoid vowel sequences, a repair like vowel deletion, consonant insertion, or vowel coalescence becomes a possible strategy in Old Japanese. Much previous research has focused on vowel deletion (Oho, 1955, 1977; Yamaguchi, 1985; Unger, 1993; Kishida, 1998; Russell, 2003; Vovin, 2005;

Frellesvig, 2010) and has tried to adopt different mechanisms, such as sonority scales and syllable numbers, to account for vowel deletion in Old Japanese. Although vowel deletion indeed takes place in Old Japanese, not all vowel sequences render vowel deletion, and vowel sequences still exist in Old Japanese. In general, the vowel sequences occur in three conditions, as shown in (1).

- (1) a. within a phonological word;
- b. through morphological process;
- c. by grammatical element

In the first condition, consonant deletion in the intervocalic positions would lead to the contact of two vowels. For example, *makadi* 'paired oars' has a variant *makai* without the consonant *d*. Another example for the first condition is *mi-ato* HON-foot.print 'foot print'.¹ The formation of this word is a prefix of honorifics *mi-* and the word *ato* 'foot print'.

In the second condition, most examples are compounds. When the first word is followed by another word beginning with a vowel, a compound is formed and the two vowels contact, as in *opo-isi* 'big-stone'. In addition to phonological (1a) and morphological (1b) factors, grammatical element also renders vowel sequences in Old Japanese. For

¹ The abbreviations of the glosses in this paper are listed as follows: ATTR= attributive; CAUS = causative; CONC = concessive; COP = copular; DAT = dative; GEN = genitive; HON = honorifics; INF = infinitive; NEG = negation; PART = particle; PASS = passive; PERF = perfective; PFX = prefix; TOP = topic.

example, the infinitive marker in Old Japanese is vowel *-i-*, as in *k-i-ir-i* ‘come-INF-enter-INF’, and there is a vowel sequence *-i-i-* from the grammatical element *-i-* and the first vowel *i* in the root *-ir-* ‘enter’.

Although the three conditions account for vowel sequences in Old Japanese, it remains unknown what the exact distributions of vowel sequences are in the three conditions and what the phonological constraints are, if there are, in Old Japanese. Therefore, before we explore phonological changes like vowel deletion, there is no doubt that we have to probe into the distributions of vowel sequences in Old Japanese. To achieve this goal, this paper adopts a corpus-based approach by collecting data from an online database, the Oxford-NINJAL Corpus of Old Japanese (ONCOJ). To facilitate the discussion, this paper is organized as follows. Section 2 introduces the Oxford-NINJAL Corpus of Old Japanese (ONCOJ) and discusses data selection criteria and analysis procedures. Section 3 reports the results, and section 4 discusses the phonological constraints in the three conditions. Besides, section 4 discusses the types of words in the corpus. Section 5 concludes this paper.

2 Corpus and data selection criteria

This section discusses the corpus used in this paper and data selection criteria. The data are collected from <https://oncoj.ninjal.ac.jp/>, with seven major philological sources, as listed in (2).

- (2) a. *Kojiki kayō* (KK), 112 poems; 2,527 words. Compiled 712 CE
- b. *Nihon shoki kayō* (NSK), 133 poems; 2444 words. Compiled 720 CE
- c. *Fudoki kayō* (FK), 20 poems; 271 words. Compiled 730s CE
- d. *Bussokuseki-ka* (BS), 21 poems; 337 words. Compiled after 753 CE
- e. *Man'yōshū* (MYS), 4,685 poems; 83,706 words. Compiled after 759 CE
- f. *Shoku nihongi kayō* (SNK), 8 poems; 134 words. Compiled 797 CE
- g. *Jōgū shōtoku hōō teisetsu* (JSHT), 4 poems; 60 words. Date unknown

In ONCOJ, there are 4850 poems in total, and 97% of the poems are from MYS.

The steps of data collection are as follows. First, Old Japanese is transcribed in Chinese characters in two formats: phonographic and logographic. To unveil the phonetic values of the transcriptions, only the phonographic forms are analyzed in this paper. Second, vowels in ONCOJ are in the following notation: *a*, *e*, *o*, *i*, *u*, *wo*, *wi*, and *ye*. The eight vowels are conventionally divided into general vowels and paired vowels. The former includes vowels *a* and *u*; the latter includes pairs *i/wi*, *ye/e*, and *wo/o* (the vowels *i*, *ye*, and *wo* in the pair are known as *korui* vowels and the second vowels *wi*, *e*, and *o* in the pairs are known as *otsurui* vowels). The eight vowels lead to 64 possible combinations for vowel sequences. Third, the data are classified into three conditions depending on how the vowel sequence is recognized:

- (a) in a phonological word, as in *kai* ‘rudder’ and *mi-ato* ‘foot; foot print’
- (b) compound, as in *opo-isi* ‘big-stone’;
- (c) grammatical element, as in *k-i-ir-i* ‘come-INF-enter-INF’.

Finally, since most of the data in Old Japanese are poems, this paper focuses on the situations where the vowel sequences occur in a phonological word. In other words, vowel sequences across a phonological word or a phrase, as in *wa-ga opo-kimi* 1st.pronoun-GEN big-lord ‘my great lord’ (MYS.18.4059), are not analyzed.

3 Results

In this paper, there are 627 examples of vowel sequences from ONCOJ, and Table 1 below shows the general distribution of the combinations of the eight vowels. In Table 1, the tendencies in the two vowels are different. In the first vowel, vowel *a* outnumbers other vowels. Vowels *i* and *o* are ranked second and third in Table 1. Other vowels, such as *u* and *wo*, take less than 10% of the total examples. The other three vowels *ye*, *e*, and *wi*, are rare in the corpus (< 5% in total). On the other hand, vowel *wo* is the majority in the second vowel, and vowel *ye* is in the second ranking. Fewer than 100 examples in the corpus, vowels *i*, *wi*, *a*, *o*, *u*, and *e*, less frequently appear in the second position of vowel sequence.

Table 1: The distribution of the combinations of the eight vowels

1 st vowel \ 2 nd vowel	a	i	u	ye	wo	e	o	wi	Total
a	7	38	0	0	0	2	1	1	49
i	10	28	12	1	0	6	7	6	70
u	11	10	2	1	2	0	0	0	26
ye	68	25	19	0	25	0	21	0	158
wo	115	23	14	4	14	0	56	0	226
e	0	1	0	0	0	3	0	0	4
o	3	28	1	0	0	0	0	0	32
wi	30	13	0	0	1	4	14	0	62
Total	244	166	48	6	42	15	99	7	627

In the 64 possible combinations of the vowel sequences, the cell that exceeds 100 examples is that of the combination of vowels *a* and *wo*, with 115 examples (18%, 115/627), as in *sa-wo-wo-ni pa* PFX-small-peak-DAT TOP ‘to the small peak’ (KK 89), and *awo-ni yo-si* red-clay good-COP ‘good red clay’ (MYS.5.806).

There are two cells with more than 50 examples. The first cell is that of the combination of vowels *a* and *ye*, as in *kasuka-ye-no* KASUKA-bay-GEN ‘of Kasuka Bay’ (KK 95); the second cell is that of the combination of vowels *o* and *wo*, as in *opo-wo-ni pa* big-peak-DAT TOP ‘to the big peak’ (KK 89). The following subsections discuss the distributions of vowel sequences in three categories: in a phonological word (3.1), compounds (3.2), and grammatical element (3.3).

3.1 Vowel sequences in a phonological word

The first category of vowel sequences is that in a phonological word. Words with affix, prefix in particular, are also included in this category. In total, there are 305 examples in this category, as shown in Table 2 below.

In Table 2, if we focus on the first vowel, more than half of the examples are in vowel *a* (157 examples). On the other hand, vowel *wo* is the majority in the second vowel (140 examples). In Table 2, the vowel *a* tends to cooccur with vowel *wo* or vowel *ye*, as in *awo-ni yo-si* red-clay good-COP ‘good red clay’ (MYS.5.806) and *aye-n-u gani* fall.off-PERF-ATTR PART ‘fallen off’ (MYS.8.1507). Other frequent combinations like vowels *a* and *wi* are also found in the corpus, as in *kurenawi n-i* safflower COP-INF ‘the safflower’ (MYS.18.4111). Other first vowels with sufficient examples are vowel *o* (52 examples) and vowel *wo* (40 examples), as in the conjunctive particle *monowo*. There are some sporadic examples such

Table 2: The distribution of the combinations of the eight vowels (in a phonological word)

1 st vowel \ 2 nd vowel	a	i	u	ye	wo	e	o	wi	Total
a	2	9	0	0	0	0	0	0	11
i	8	0	6	0	0	0	0	4	18
u	7	2	0	0	1	0	0	0	10
ye	51	3	11	0	26	0	8	0	99
wo	68	8	7	0	13	0	44	0	140
e	0	0	0	0	0	1	0	0	1
o	2	4	0	0	0	0	0	0	6
wi	19	1	0	0	0	0	0	0	20
Total	157	27	24	0	40	1	52	4	305

as *kwoye-te k-ite* pass.over-GEN come-GEN ‘pass over and come’ (MYS.15.3762).

3.2 Vowel sequences in compound

In the second category of vowel sequences, there are 131 examples in the corpus. Table 3 shows the distribution of the combinations of the eight vowels in compounds.

Some other sporadic examples are also found in the corpus, such as vowels *o* and *wi* (13 examples) in *kumo-wi nasu* cloud-be.at COP ‘clouds’ (MYS.17.4003), and vowels *u* and *ye* (8 examples) in *sidu-ye pa* lower-branch TOP ‘lower branch’ (KK 43).

Table 3: The distribution of the combinations of the eight vowels (in compounds)

1 st vowel \ 2 nd vowel	a	i	u	ye	wo	e	o	wi	Total
a	5	0	0	0	0	2	1	0	8
i	2	0	3	1	0	5	3	0	14
u	4	0	1	1	1	0	0	0	7
ye	4	0	8	0	0	0	0	0	12
wo	47	5	7	4	1	0	4	0	68
e	0	0	0	0	0	2	0	0	2
o	1	0	1	0	0	0	0	0	2
wi	3	1	0	0	1	0	13	0	18
Total	66	6	20	6	3	9	21	0	131

If we focus on the first vowel in Table 3, approximately half of the examples are in vowel *a* (66 examples). Likewise, the majority in the second vowel is vowel *wo* (68 examples).

The commonest vowel sequence in Table 3 is that of vowels *a* and *wo* (47 examples), as in *masura-wo no* strong-male COP ‘strong man’ (MYS.5.804).

3.3 Vowel sequence in grammatical element

In the third category of vowel sequences, there are 191 examples in the corpus, as shown in Table 4 below.

Different from the distributions in Tables 2 and 3, the one in Table 4 shows that the majority of the first vowel goes to vowel *i* (70%, 133/191), as in *wor-i-ak-as-i mo* sit-INF-bright-CAUS-INF PART

Table 4: The distribution of the combinations of the eight vowels (in grammatical element)

1 st vowel \ 2 nd vowel	a	i	u	ye	wo	e	o	wi	Total
a	0	29	0	0	0	0	0	1	30
i	0	28	3	0	0	1	4	0	36
u	0	8	1	0	0	0	0	0	9
ye	13	22	0	0	0	0	12	0	47
wo	0	9	0	0	0	0	8	0	17
e	0	1	0	0	0	0	0	0	1
o	0	24	0	0	0	0	0	0	24
wi	8	12	0	0	0	4	1	2	27
Total	21	133	4	0	0	5	25	3	191

‘sit up all night’ (MYS.18.4068) or in *yeda kir-i-oros-i* branch cut-INF-let.down-INF ‘cut down the branches’ (MYS.15.3603). Other vowel sequences are mostly attested when the second vowel is *ye*, as in *i-no ne-raye-n-u ni* sleep-GEN sleep-PASS-NEG-ATTR COP ‘(I) do not sleep’ (MYS.15.3665), and *kik-o-ye sikado mo* hear-INF-PASS CONC PART ‘Although I am heard,’ (KK 45).

Thus far, we have discussed the distributions of the vowel sequences in the corpus, and the results have shown that the vowel sequences in the first and second conditions in (1) differ from those in the third condition in (1).

4 Discussion

This section discusses phonological constraints in the vowel sequences in Old Japanese (4.1) and the words with the vowel sequences (4.2).

4.1 Phonological constraints

The three categories of vowel sequences in the corpus show two phonological constraints in Old Japanese. In section 3, we have discussed the three categories of vowel sequences in Old Japanese. Of the three categories, the third category is different from the other categories, as the first vowel in high frequency is an infinitive marker, a high front vowel *i* (vowel *o* as a variant of vowel *i*).

If we temporarily remove the distribution of grammatical element from Table 2, the combined distribution of a phonological word and compound, as shown in Table 5 below, reflects the first phonological constraint of the vowel sequences in Old Japanese.

The five vowel sequences that exceed 5 % of the corpus examples are *a + wo* (115 examples, 26%), *a + ye* (55 examples, 13%), *o + wo* (48 examples, 11%), *wo + ye* (26 examples, 6%), *a + wi* (22 examples, 5%).

The vowel sequences in Table 5 reflect the ranking of the unmarkedness of the vowels in the two positions: $a > o > u > wo > i > e > wi$ for the first vowel and $wo > ye > wi > i > a > u > o > e$ for the second vowel. It is apparent that vowel sequences in Old Japanese are not just random combinations of any two vowels. The vowel sequences are preferred to be vowel_[-front] plus vowel_[-low].

The second phonological constraint based on the corpus data is that the general vowels, *a* and *u*, and the paired vowels *i/wi*, *ye/e*, and *wo/o* (*korui* vowels including *i*, *ye* and *wo*, and *otsurui* vowels including *wi*, *e*, and *o*) are positionally different in the vowel sequences. Table 6 below shows the nine combinations of general and paired vowels.

The combination of general and *korui* vowels outnumbers other combinations in the corpus. The data in Table 6 suggest that the first vowel is preferred to be the general vowels, and the second vowel is preferred to be the *korui* vowels.

Table 5: The distribution of the combinations of the eight vowels (without grammatical element)

1 st vowel \ 2 nd vowel	a	i	u	ye	wo	e	o	wi	Total
a	7	9	0	0	0	2	1	0	19
i	10	0	9	1	0	5	3	4	32
u	11	2	1	1	2	0	0	0	17
ye	55	3	19	0	26	0	8	0	111
wo	115	13	14	4	14	0	48	0	208
e	0	0	0	0	0	3	0	0	3
o	3	4	1	0	0	0	0	0	8
wi	22	2	0	0	1	0	13	0	38
Total	223	33	44	6	43	10	73	4	436

Table 6: Combinations of general and paired vowels

1 st vowel \ 2 nd vowel	G	K	O	Total
G	19	14	3	36
K	222	61	68	351
O	26	7	16	49
Total	267	82	87	436

*G = general vowels; K = *korui* vowels; o = *otsurui* vowels

To sum up, the two phonological constraints in vowel sequences in Old Japanese are vowel_[-front] plus vowel_[-low] in terms of vowel types and general vowel plus *korui* vowel in the terms of vocalic categories.

4.2 The words with the vowel sequences

In this section, we discuss the vowel sequences in the word differences in the corpus. In Table 5, we have shown the frequencies of vowel sequences in the corpus, while some words are frequently used in the corpus. For example, the word *monowo* ‘focus particle’ is attested 38 times in the corpus, and the word *awo* ‘green or blue’ is attested 32 times. The fact that two words are highly frequently used in the corpus makes vowel *wo* prominent in the second vowel. To fully understand the vowel sequences in Old Japanese, it is necessary to calculate the vowel sequences in terms of types of words, not the frequency of words.

Table 7 shows distributions in a phonological word without affixation. Examples of compounds are also not included in the discussion. In Table 7, there are 50 words, and half of the words are those with vowel *a* as the first vowel, as in *taye* ‘break’ and *sakaye* ‘prosper’ for *a + ye* (9 words), *awo* ‘green or blue’ and *mawos-* ‘speak [hum]’ for *a + wo* (6 words), and *sawi* ‘rushing’ for *a + wi* (4 words).

If we consider the number of words together with the tokens in the vowel sequences in Table 7, the top three vowel sequences are *a + wo*, *a + ye*, and *o + wo*.

5 Conclusion

This paper has discussed the vowel sequences in Old Japanese from a corpus-based approach. This

paper has discussed the vowel sequences in terms of three conditions: in a phonological word, in compound, and in grammatical element. The results have shown that the distribution of grammatical element differs from those of the other two categories in the vowel types since most

Table 7: The number of words in vowel sequences

Vowel 1	Vowel 2	Words	Tokens
a	ye	9	50
a	wo	6	55
a	wi	4	19
i	wo	4	7
u	wo	4	7
o	wo	4	44
a	u	3	7
u	ye	3	11
o	ye	3	8
wo	ye	2	26
a	a	1	2
a	i	1	8
i	ye	1	1
i	wi	1	1
u	i	1	6
wi	i	1	4
wo	wo	1	12
e	e	1	1
Total		50	269

examples in the category of grammatical element are reflected by the infinitive marker, vowel *i*.

On the basis of the corpus data, this paper has also discussed two phonological constraints and the types of vowel sequences in words. On the one hand, the first vowel tends to be [-front] and the second vowel to be [-low]; on the other hand, the first vowel tends to be general vowels and the second to be *korui* vowels. In addition to the frequencies of the corpus data, this paper has explored the words in different vowel sequences. The comparison of words and tokens has indicated that the most frequently used vowel sequences are *a + ye*, *a + wo*, and *o + wo*.

Much effort has been made to discuss the vowel sequences within the domain of a phonological word in Old Japanese. It is also possible that vowel sequences are attested beyond the phonological word. For example, in *wa-ga opo-kimi* 1st.pronoun-GEN big-lord ‘my great lord’ (MYS.18.4059), there is a vowel sequence between the genitive

case marker *ga* and the noun *opo-kimi*. In future research, one topic that needs in-depth investigation could be the vowel sequences across phonological words. After that, a complete picture of the vowel sequences in Old Japanese can be obtained.

Acknowledgments

A research grant from Taiwan's National Science and Technology Council 110-2410-H-011-033- is gratefully acknowledged. The usual disclaimer applies.

References

- Alexander Vovin. 2005. *A Descriptive and Comparative Grammar of Western Old Japanese*. Global Oriental Folkestone, Kent, UK.
- Bjarke Frellesvig. 2010. *A History of the Japanese Language*. Cambridge University Press, Cambridge, UK.
- Kelly Russell L. 2003. 'Contraction and Monophthongization in Old Japanese' in Toshiki O, Vovin A (ed.), *Nihongo keitōron no genzai/Perspectives on the Origins of the Japanese Language*, Nichibunken, Tokyo, Japan.
- Marshall Unger, J. 1993 [1977]. *Studies in early Japanese morphophonemics* (2nd revised edition). Indiana University Linguistics Club Publications Bloomington, USA.
- Susumu Ono. 1953. Manyoujidai no Onyin [Phonology in Manyou period], in *Manyoushu Daisei 6: gengohen* [Collection of Manyoushu 6: linguistics]. Heiponsha, Tokyo, Japan.
- Susumu Ono. 1977. Onyin no Hensen 1 [Phonological changes 1], in *Iwanami Kouza: Nihongo 5: Onyin* [Seminar of Iwanami: Japanese phonology]. Iwanami, Tokyo, Japan.
- Takeo Kishida. 1998. *Kokugo Onyin Henkaron* [A Study of Japanese Phonology]. Musashinoshoin, Tokyo, Japan.
- Yoshinori Yamaguchi. 1985. *Kodai Nihongo Bunbo no Seiritsu no Kenkyu* [A Study of Old Japanese Grammar]. Yuseitou, Tokyo, Japan.