Old Japanese and Proto-Japonic Word Structure

Blaine ERICKSON

Kumamoto Gakuen University

1. Introduction

In the search for external language relationships, certain internal facts must be ascertained first. Here, I examine these facts as they apply to Old Japanese and its predecessor, pre-Old Japanese. Then, I review the word structure of Old Japanese, and, in conjunction with what is known about Old Japanese phonology and morphology, posit probable word shape for proto-Japonic.

2. Procedure

The first step in comparing one language to others is to use the oldest form of the language attested. Fortunately, Japanese is one of the best-attested languages in the world, with a written tradition going back to the eighth century CE. The language attested in the oldest records was used in central Japan, and is known as Old Japanese (OJ). While there are written records in Japan that predate this era, they are far fewer in number than those of the Old Japanese period. The primary texts of this period are the *Kojiki* (*Records of Ancient Matters*, 712); the *Nihon Shoki* (*Chronicles of Japan*, 720); and the *Man'yooshuu* (*Collection of a Myriad Leaves*, circa 759), an anthology of some 4,500 Japanese poems (Okimori 1989; Seeley 1991). *Senmyoo* ('imperial edicts,' 697–789) are another important source of Old Japanese writings. It should be understood that not all writings in these documents necessarily reflect the eighth-century language of central Japan; for example, some sections of the *Man'yooshuu* contain poems from eastern and southern parts of Japan, and other poems are possibly from as early as the fourth century (Habein 1984: 11 et passim). Furthermore, not all sections of these works are written in the most-easily decipherable form of Old Japanese writing, the phonograms called *ongana*. Regardless, these documents comprise a large body of connected text and are the indispensable core of Old Japanese.

The next step is to gain a thorough understanding of the phonology and morphology of the language in question, since comparison of these areas is far more likely to give useful results than comparison of other areas. Starting with the Edo-period Japanese scholars Motoori Norinaga (1730–1801) (Shiraishi 1957) and his student Ishizuka Tatsumaro (1764–1823) (Kasuga 1980), scholarship on Old Japanese has a long history, and recent research has advanced our understanding of this language to the point where we can make many solid claims about it.

Finally, internal reconstruction and comparison with languages demonstrably related

to the one in question can give us pre- and proto-languages, respectively, which can then be meaningfully compared to other languages whose genetic affiliation to the language in question remains to be demonstrated. For Old Japanese, we have both internal reconstruction (Unger 1977) and the comparative reconstruction of a related language group, namely Ryukyuan (Thorpe 1983). Although work in these areas is not as advanced as that on Old Japanese itself, it has progressed enough to allow significant comparisons of proto-Japonic (i.e., the common ancestor of both mainland Japanese and the languages of the Ryukyus) to other languages.

Another important consideration is time. Although languages may borrow words from a dead language, such as the neo-Latinate vocabulary in English, the normal situation is for languages to borrow from contemporaneous varieties of other languages. Dating language contacts and excluding them from the native vocabulary is an important method of clearing potentially confusing data from the investigator's purview. In the case of Japanese, we have a study of early loans (Miyake 1997), as well as certain knowledge of later loans from many varieties of Chinese and several European languages. These loans include what could be called *neo-Sinitic*, which are words not taken from any variety of Chinese but were created in Japan, such as *denwa* 'telephone.' Similarly, in comparing reconstructed languages, it is important to compare languages of similar time depths. For example, the time depth of proto-Austronesian (about 6,000 years) is far too great for it to be usefully compared directly to proto-Japonic (about 2,000 years) (Vovin 2001). If there are any Austronesian elements in Japonic, they are most likely from a stage later than proto-Austronesian.

Finally, proper application of the comparative method is crucial. For example, language relationships are transitive: if language A is related to language B, and if one makes the claim that language X is related to language A, then X must also be related to B. To say otherwise is similar to claiming that you are related to your cousin, but not to her sister. Unfortunately, there is an all-too-large body of naive work by both laypeople and otherwise-respectable scholars that is sadly lacking in the correct application of the comparative method. These works should be ignored by anyone wishing to do serious work on the origins of the Japanese language.

3. Phonological Structures in Old Japanese

The central tool of comparative historical linguistics is the determination of regular correspondences between (potentially) related languages. Of course, it is meaningless to directly compare phoneme inventories. What is compared is the context in which phonemes occur: words. To this end, it is crucial to determine the words of languages in question, as well as to understand the phonology and morphology of these languages to uncover certain facts that might otherwise be obscured.

For Old Japanese, we have a large body of texts, mainly poetry, to inform us of at least part of its lexicon. There is enough variation in this corpus to be able to determine some of the phonological processes that were active in Old Japanese, or possibly its

immediate predecessor, which is sometimes called Archaic Japanese (AJ) (e.g., Unger 1977: 1). Additionally, we have enough examples in the Old Japanese corpus to make informed decisions about Old Japanese morphology. For example, it is clear that Old Japanese had a rich system of verbal inflection, but that the primary form of noun formation was compounding. It is important to note that unlike Austronesian languages, no attested stage of Japanese shows any evidence of infixation, and reduplication, though attested from Old Japanese onwards, is not as common in Japonic as it is in Austronesian.

An important fact about Old Japanese is that it made a greater number of orthographic distinctions than modern writing does, at least for some vowels in some positions (Hashimoto 1917). This graphic distinction is called *koo* and *otsu*, or type A and type B, and reflects phonemic distinctions in Old Japanese. See Miyake 1999 for full discussion and reconstruction.

3.1 Old Japanese Syllable Structure

In order to understand Old Japanese word structure, it is important to understand structure at lower levels as well. I start the discussion with syllables.

Old Japanese allowed only two kinds of syllables word-internally, V and CV, with the former found almost exclusively in word-initial position. There are also a few words of the form (C)VV[+high], such as *kai* 'oar' and *uu* 'to plant,' but these are rare, and in poetry, the former appears to have counted as two syllables. CV_iV_i sequences appear to have counted in poetry as one syllable, as did most instances of CV_iV_j. To put it another way, Old Japanese was a syllable-timed language, unlike Middle or Modern Japanese, both of which are mora-timed.

What is the evidence for these claims? Old Japanese poetry consists overwhelmingly of lines of five or seven timing units. In poetry written in *ongana*, each unit is usually written with one character. In the modern language, these units are moras, but in Old Japanese, they were syllables. Some poems exhibit what is called *jiamari*, 'extra characters,' and these poems have more than five or seven characters in a line. However, the extras in Old Japanese poetry are almost always V, not CV. When poems with 'extra characters' are read with CVV syllables, these poems then have the proper scansion of five or seven syllables per line. Such poems are found in the *Man'yooshuu*. A few examples are given below, along with their *Man'yooshuu* numbers and the characters used to write them.

(1) 809: 多陀爾阿波須 *tadani apazu* 'not meeting directly'
(2) 850: 由吉能伊呂遠 *yukyi no iro wo* '(taking) the color of the snow'
(3) 3662: 安気婆安気奴等母 *akeyba akeynu tomo* 'whether or not it is opened'
(4) 3665: 伊毛乎於毛比 *imo wo omopyi* 'thinking of my beloved'
(5) 3937: 多妣伊爾之伎美我 *tabyi inisi kyimyi ga* 'you, who left on a journey'
(6) 3948: 比奈爾安流和禮乎 *pyina ni aru ware wo* 'me, in the countryside'
(7) 4129: 等利安宣麻儛爾於吉 *tori agey mapye ni okyi* 'lift up and place in front'

In some of the examples, a CV_iV_i sequence—*ba a-* in example 3, *wo o-* in example 4, and *byi i-* in example 5—counts as one syllable. This strongly suggests that vowel length was not contrastive in Old Japanese, and that there was no distinction between long and short syllables; at the very least, vowel length did not matter in poetry. However, when read in the modern language, these lines have eight, six, and eight moras, respectively.

In the other examples, lines with C*i*V sequences suggest that the medial *-i-* was treated as a glide. The line of poem 3948 (example 6) shown has eight characters, but if *nia-* of *ni aru* is read as one syllable, then this line has seven syllables and does not violate the rule of five or seven timing units per line. The same holds for the other examples shown above. For further discussion, see Erickson 1998: 36–51.

Based on phonogram usage, it is clear that Old Japanese had absolutely no closed syllables of any kind. However, syllabic *m* was permitted word-initially (Erickson 1998: 44–45). Syllabic *m* was, in all likelihood, an innovation in Old (or possibly Archaic) Japanese, and need not be considered at a higher level. Accordingly, I will not consider it further here.

These data suggest that Old Japanese did not have a long-short syllable distinction. Put another way, Old Japanese, unlike later stages, was not a mora-timed language: it was a syllable-timed language. This is also ascertainable from what happened when two vowels came together as a result of compounding or consonant loss, as well as from what happened when two consonants came together as a result of vowel loss. This is discussed further below (§ 3.3), as well as in Erickson 1998 (38–42).

3.2 Old Japanese Phonemes

It goes without saying that the phoneme inventory of Old Japanese must be considered in the attempt to relate it to other languages. The state of the art on the values of the Old Japanese phonemes is Miyake 1999; I present his results below.

Labial	*Labial*	*Dental*	*Palatal*	*Velar*
-voi obstruents	*p	*t *s		*k
prenasalized +voi obstruents	*mb	*nd *nz		*ŋg
nasals	*m	*n		
liquids & glides	*w	*r	*y[1]	

Table 1: OJ Consonants (from Miyake 1999: 465)

1 Miyake's y is IPA [j].

	Front	*Central*	*Back*	*Diphthong*
High	yi = *iy	iy = *ɨ[2]	u = *u	
Mid	ye =*e	o̲ = *ə	wo = *o	ey = *əy
Low		a = *a		

Table 2: OJ Vowels (from Miyake 1999: 592)

It must be understood that certain of these phonemes are considered to be secondary, i.e., ultimately derived from other sources. For the consonants, the prenasalized voiced obstruents are taken as having come from earlier sequences of a nasal followed by a voiceless obstruent; these in turn may be from earlier NVC sequences. As for the vowels, *iy*, *wo*, and *ey* are from earlier sequences of two vowels, which in turn are often from earlier VCV sequences. Furthermore, *ye* is also suspect, as the overall frequency of both kinds of *e* in Old Japanese was exceedingly low: just 9% of all vowels (Martin 1987: 47–48). In fact, some instances of *ye* can be shown to come from VV sequences (see example 22 below). These and other facts suggest that of the eight Old Japanese vowels, only *four—yi, u, o̲,* and *a*—are underived and can therefore be posited for an earlier stage of the language. (Evidence from Ryukyuan not considered here suggests that proto-Japonic had six vowels (Serafim 1999); perhaps the other two vowels were preserved in proto-Ryukyuan but lost in pre-Old Japanese.)

3.3 Phonological Processes in Old Japanese

Here, I discuss the phonological processes that can be inferred for Old Japanese (or possibly Archaic Japanese; see Erickson 1998: 41–42 et passim). I start with processes affecting vowels, and then move on to those affecting consonants.

3.3.1 Vocalic Processes

Some of the most important phonological processes for this discussion are those that affected vowel-vowel sequences. Old Japanese had virtually no underived vowel-vowel sequences, and VV sequences were not tolerated across morpheme boundaries. (The verb mentioned above, OJ *uu* 'to plant,' shows medial -/w/- in other forms, e.g., *uwete* 'plant CONTINUATIVE'; no known variety of Japanese has allowed homorganic high glide-high vowel sequences.) Normally, when vowel-vowel sequences arose, one of three changes occurred: deletion of the first vowel,

(8) *waga-* 'my' + *imo* 'lover' > OJ *wagyimo* 'my lover'
(9) *waga-* 'my' + *ipye* 'home' > OJ *wagyipye* 'my home'

2 I believe that [ɨ̟ɨ̞] is a better value for *ï* than [ɨ], and will use it henceforth.

(10) *ara-* 'rough' + *umyi* 'sea' > OJ *arumyi* 'rough seas'

deletion of the second vowel,

(11) *waga-* 'my' + *ipye* 'home' > OJ *wagapye* 'my home'
(12) *yama* 'mountain' + *no̲* 'GENITIVE' + *upey* 'top' > OJ *yamano̲pey* 'mountain-top'
(13) *panare* 'separated' + *isô* 'rocky beach' > OJ *panaresô* 'shoal'

or crasis and reduction to one syllable (Unger 1977; Yamaguchi 1971).

(14) *taka-* 'high' + *iti* 'market' > OJ *takeyti* '(proper noun)'
(15) *to̲no̲* 'palace' + *iri* 'to enter' > OJ *to̲neri* '(a kind of imperial servant)'
(16) *sakyi* 'to bloom' + *ari* 'to be' > OJ *sakyeri* 'be blooming'

Words such as these are attested throughout the Old Japanese corpus. Some scholars, such as Kishida (1942) and Yamaguchi (1971), have given phonetically-based conditioning factors for these changes, whereas Unger presented a theory of contraction 'not based on phonetic considerations' (1977: 41). A summary of the these accounts is found in Erickson 1998 (38–40). Although Unger's analysis matches the data better than either Kishida's or Yamaguchi's, the existence of doublets like OJ *panaresö~ panarisö* 'shoal' (see example 13) remains problematic for any account. Given the large number of doublets found in Old Japanese, perhaps any explanation will need to rely on the notion of dialect mixture (Serafim 1978).

The phonological processes exemplified by examples 14–16 are part of a group of processes that can be lumped together under what Martin called *vowel crasis* (1987: 45 et passim). As noted above, Old Japanese did not tolerate sequences of vowels, and when two vowels came together due to consonant loss or compounding, the result was a single vowel, as discussed above. The cases of vowel loss (examples 8-13), though important for demonstrating that Old Japanese was syllable-timed, are of significantly less interest for this discussion than those in which the two vowels resulted in a third (examples 14–16). These 'third vowels' are the secondary vowels mentioned above, namely *iy*, *wo*, *ye*, and *ey*.The sources of these vowels are presented below (after Frellesvig 1995: 61; Martin 1987: 58; and Miyake 1999: ch. 8).

(17) *iy* : *u* + *yi* = [ui] > [i̯i̯] ,e.g., OJ *tuku-* 'moon' + C*i* '(suffix of unknown meaning)' > *tukiy* 'moon'
(18) *iy* : *o̲* + *yi* = [əi] > [i̯i̯] ,e.g., pJ **pəri* 'fire' > **pəi* > OJ *piy* id. (Unger 2000: 659)
(19) *wo* : *u* + *o̲* = [uə] > [o] ,e.g., OJ *situori* '(a kind of woven cloth)' ~ OJ *sitwori* id.
(20) *wo* : *u* + *a* = [ua] > [o] ,e.g., OJ *kazu* 'number' + OJ *apu* 'to combine VT' > OJ *kazwopu*

'to count'

(21) *ye* : *i* + a = [i a] > [e], e.g., OJ *sakyi* 'to bloom' + OJ *ari* 'to be' > OJ *sakyeri* 'is in bloom'

(22) *ye* : *i*+*o̲* = [i ə] > [e], e.g., AJ **utusi* 'real' + AJ **o̲myi* 'person' > AJ * *utusyemyi* 'livingperson' (as opposed to a dead person or one encountered in a dream) > OJ *utusemyi* id. ~ OJ *utuso̲myi* id. (*Utusi* and *omyi* are both attested in Old Japanese; see the discussion in Omodaka et al. 1967: 122, 123.)

(23) *ey* : *a* + *i* = [ai] > [ə i̯], e.g., OJ *ama-* '*rain*' + C*i* '(suffix of unknown meaning)' > *amey* 'rain'

(24) *ey* : *o̲* + *i* = [əi] > [ə i̯], e.g., AJ **to̲no̲* 'palace' + AJ **iri* 'entering' > AJ **to̲neyri* '(an imperial servant)' > OJ *to̲neri* id. (both *to̲no̲* and *iri* are attested in Old Japanese)

As seen above, the secondary vowels are from earlier sequences of two vowels, which in turn are from earlier VV or VCV sequences.

The observant reader will have noticed that [ə i̯] is a source of both *iy* and *ey*. Although many instances of *iy* are from /ui/ sequences, some, such as *ko̲* ~ *kiy* 'tree,' are not. Clearly, there are two different phonological processes at work. One, responsible for some instances of *iy* , partially assimilates [ə] to [i] by raising [ə] to [ɨ] The other, which accounts for *ey* , partially assimilates [a] to [i] by raising [a] to [ə]. These processes must have been active in counter-feeding order, because if they were not, /ai/ would also be a source for *iy* –but it is not. Since synchronically-active phonological processes apply simultaneously and are unordered with respect to each other (Donegan and Stampe 1979 §§ 3.1–3.2), we know that the first process ([əi] > [ɨi̯]) was no longer active by the time the second one ([ai] > [ə i̯]) started to apply. Accordingly, *o̲* ~ *iy* alternation must be older than *a* ~ *ey* alternation, and *ey* was a later entrant into the historical Japanese phonological system than *iy*.

3.3.2 Consonantal Processes

In addition to the changes discussed above, another significant change was the loss of a vowel, which caused two consonants to come together. When this happened, the word also lost a syllable, as shown below. Significantly, words resulting from such interconsonantal vowel deletions did not have consonant clusters; this is in marked contrast with Middle Japanese, where such deletions always resulted in consonant clusters.

(25) AJ **amyi* 'net' + AJ **pyikyi* 'pulling N' > OJ *abyikyi* 'net-'shing N' (both *amyi* 'net' and *pyikyi* 'pulling N' are attested in Old Japanese)

(26) OJ *hayapyito̲* ~ *hayato̲* '(proper noun)'

(27) OJ *kyigyisi* 'pheasant' ~ OJ *kyizi* id.

(28) AJ **mura* 'village' + AJ **nusi* 'owner' > OJ *murazi* '(a title)' (*mura* 'village' and *nusi* 'owner' are both attested in Old Japanese)

(29) OJ *nurite* ~ *nute* 'kind of bell'

(30) AJ **nusumyipyitoo* > OJ *nusubyitoo* 'thief'

(31) AJ **osisaka*? **osasaka*? '(proper noun)' > OJ *osaka* id. (see the discussion in Hamada 1946)

The voicing caused by nasal and voiced consonants preceding the elided vowel (examples 25, 27, 28, and 30) shows that this was not simply the loss of an entire syllable. Rather, vowel deletion occurred first, and then the remaining voiced or nasal consonant induced voicing on the following consonant. (I believe it is better to analyze this as the complete assimilation of the "deleted" vowel to the preceding consonant, but I will not make those arguments here.) As for why such a series of changes is proposed, recall that Old Japanese voiced obstruents were prenasalized (§ 3.2). When the vowel after a voiced or nasal consonant was lost, the nasality of the consonant remained and was transferred to the consonant following the now-deleted vowel. See Martin 1987: 20-29; also Hamada 1952; Miyake 1995; Serafim nd; Unger 1977; Vance 1982; and Whitman 1996 for discussion.

A word like OJ *abyikyi* 'net-fishing N' developed as follows: AJ *[**amĩpiki**] underwent vowel loss to give *[**ãmpiki**]. In Old Japanese, nasals preceding consonants always caused voicing of the following consonant, so this became *[**ã^{m}biki**], which was reanalyzed as OJ /abyikyi/. The other examples shown above can also be analyzed as vowel deletion followed by consonant cluster simplification.

3.3.3 Summary of Phonological Processes

From the discussion above, it is seen that the phonological processes active in Archaic and Old Japanese served to shorten words. From this, I conclude that earlier stages of the language had, on the whole, longer words than are found in Old Japanese; this will be demonstrated below.

4. Old Japanese Word Structure

Now that we have examined the syllable structure, phoneme inventory, and phonological processes of Old Japanese, we may look at the structure of Old Japanese words, as well as what this implies for proto-Japonic word structure.

The starting point for this discussion is Samuel Martin's 1987 work *The Japanese Language through Time.* This monumental tome is indispensable for any serious work on Japanese historical linguistics, and is a treasure trove of information.

On pages 600-664, Martin presented data on Japanese nouns. These data are divided into sections based on length and accent class. Martin also presented statistical analyses of these data. However, Martin's database was not limited to Old Japanese. Since I want to focus on Old Japanese, I have taken inspiration from Martin and compiled a database

of Old Japanese nouns.

4.1 Structure of the Database

I used *Jidaibetsu Kokugo Daijiten: Joodai hen* (*Dictionary of Japanese by Historical Periods: Old Japanese*) (= JDB) as the source for Old Japanese words. At this point, the database consists entirely of nouns, and contains only a small percentage of the nouns listed. Although JDB marks some words as pronouns, no attested stage of Japanese actually has pronouns: the words translated as such into languages like English function grammatically as nouns, and are therefore included in my database. Excluded from the database, however, are *makurakotoba*, which are conventionalized words and phrases found in poetry. These words are often of uncertain meaning, and as poetic conventions, were not necessarily part of regular speech.

The database has several fields for each word. These are as follows: Part of speech; Tergum; Compound; Voicing; Secondary; Deverbal, and Definition. Part of speech is self-explanatory; Tergum is the entry spelled backwards; Compound is for entries that are clearly compound words; Voicing represents those words with voiced consonants; Secondary is for words with secondary vowels; Deverbal is for nouns derived from verbs; and Definition gives the meaning. Additional fields include combinations (e.g., for words that are both compounds and have voiced consonants) and an information field for notes.

Initial word selection was as follows. Starting with page 2, I entered the nouns on facing pages of every ten pages (i.e., pp. 2–3, 12–13, 22–23, etc.). Once these pages were entered, I started with the same procedure on subsequent pages, i.e., 4–5, 14–15, etc. I did this to ensure that a variety of words was included, and so that the preliminary form of the database would not be overly weighted towards the beginning of the dictionary. The five nouns on page one, excluding the ghost word **a* 'net,' were also included.

I continue to add words to the database. When entry for all nouns is complete, I intend to perform the analyses shown below again and publish the results.

One shortcoming of using JDB as the source for Old Japanese is that its lexicon is not strictly limited to the Old Japanese period. Some of the words included therein are from slightly later works, such as *Shinsenjikyoo* (*New Selected Word Mirror*), which is from 900 (Okimori 1989: 178). Such words have been excluded as well.

4.2 Analysis of the Data

The data that I have gathered so far can be analyzed in many ways. Here, I present a number of these analyses, starting with straight interpretation of Old Japanese word length. (The total does not add up to 100% due to rounding.)

monosyllables	1%
bisyllables	16%
trisyllable	29%

quadrisyllables	37%
pentasyllables	9%
six or greater	7%

Table 3: Percentages of Old Japanese Nouns by Length

Right away, it is clear that the monosyllable is the least common word shape in Old Japanese. Nearly two-thirds of the words analyzed—66%—are three or four syllables in length, and four out of five words—82%—are two, three, or four syllables long. As for longer words, almost all of the words that are five syllables or longer are clearly compounds, with just a few that are the nominal form of verbs (e.g., OJ *utukusibï* 'kindness, mercy' derived from *utukusibu* ~ *utukusimu* 'to be kind, merciful').

These figures are not the end of the story—they are only the beginning. Analysis of the words based on their composition yields information about earlier stages of Japonic.

4.3 Word Structure at Earlier Stages

The words in the database present a starting point for analysis of earlier, unattested yet inferable, stages of the language. The following chart shows the levels of Japanese language history relevant here. In this chart, *proto-* refers to a reconstruction based on the comparative method, and *pre-* refers to a language inferred from internal reconstruction. Both Archaic Japanese and pre-Old Japanese are internal reconstructions, but the time depth of the latter is much greater than that of the former. By using the name *Archaic Japanese*, I can also avoid the ungainly term *pre-pre-Old Japanese* and the confusion it may engender, since pre-Old Japanese is based on Old and not Archaic Japanese. In any case, the difference between Old and Archaic Japanese is much smaller than that between either of them and pre-Old Japanese, as will be shown.

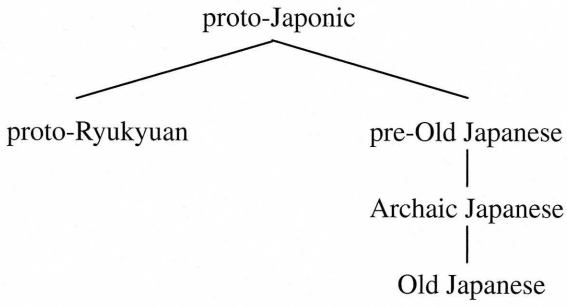

Chart 1: Stages of Japanese Language History

Although it may be possible that proto-Japonic came to the Japanese archipelago when the Yayoi people and culture entered northern Kyushu around 300 BCE, it is not at all clear that the language spoken by the Yayoi people was genetically ancestral to

modern Japanese. Since the Ryukyuan archipelago was not peopled until the third century CE, which is the same time that the Yayoi period ended and the Kofun period began, the breakup of proto-Japonic did not occur until around 300 CE . Did proto-Japonic develop *in situ* in Kyushu and western Honshu before spreading into the Ryukyus 600 years after its arrival, or did proto-Japonic come to the archipelago from the Korean peninsula with the Kofun people during the fourth century CE ? This is a question for future research, one that requires input from the disciplines of history and anthropology.

Returning to the different stages of historical Japonic shown above, the main feature of Archaic Japanese that distinguishes it from Old Japanese is that the A/B (i.e., *koo/otsu*) vowel distinction was found in all (relevant) positions, as opposed to the limited distribution in Old Japanese. In Old Japanese, the A/B distinction was found in full only after the velars. The distinction between /yi/ and /iy/, and between /ye/ and /ey/, was not found after coronal consonants, nor was there a distinction after labial consonants for /wo/ and /o̲/. Old Japanese itself can be dated to 700-850 CE (Erickson 1998: 3). Unfortunately, due to the absence of adequate records before the Old Japanese period, it is impossible to date Archaic Japanese precisely.

What, then, is the distinction between Archaic Japanese and pre-Old Japanese? Both Old and Archaic Japanese had eight vowels, four of which are secondary. In pre-Old Japanese, the secondary vowels had not yet developed, which gives it a four-vowel system consisting of /a i u ə/, with pOJ /ə/ corresponding to AJ/OJ /o̲/. What this means for the reconstruction is that every word with a secondary vowel in Old Japanese is (at least) one syllable longer in pre-Old Japanese.

Not directly mentioned above is the issue of *hei-rui*, or type C, vowels. These are the vowels for which the A/B distinction potentially existed, but either was not indicated in the orthography or was not found in Old Japanese. Since both types of OJ *e* are secondary, I have included type C *e* as a secondary vowel in my calculations for pre-Old Japanese.

4.4 Further Analysis of the Data

Here, I analyze the data beyond the superficial level shown above and start to use the techniques of internal reconstruction. Due to the difficulty of cleanly analyzing words with multiple factors (e.g., a compound word with a secondary vowel may not necessarily have the secondary vowel as a result of compounding), I take each step separately.

4.4.1 Expansion of Secondary Vowels

As shown above (§ 3.3), all the secondary vowels of Old Japanese are derived from earlier VV sequences, and many of these are, in turn, from yet earlier VCV sequences. If we allow that an earlier form of Japonic permitted VV sequences as two separate syllables, then all words with secondary vowels must be counted as one syllable longer than their Old Japanese forms. Table 4 shows percentages based on lengthening all such words.

monosyllables	1%
bisyllables	10%
trisyllables	22%
quadrisyllables	37%
pentasyllables	20%
six or greater	10%

Table 4: Percentages of pre-OJ Nouns after Expanding Secondary Vowels

Comparing these figures with those in Table 3, it is seen that bisyllables drop from 16% to 10%, and trisyllables decrease from 29% to 22%. Quadrisyllables remain unchanged, but longer words increase their shares, with pentasyllables jumping from 9% to 20%, and words of six or more syllables increasing slightly from 7% to 10%.

It must be noted that not all secondary vowels arose due to crasis of a VCV sequence; some arose due to the compounding of a vowel-initial word as the second element of that compound. As noted above (§ 3.1), Old Japanese was a syllable-timed language, and so VV sequences counted as one syllable. The timing system of pre-Old Japanese is unknown, but whatever it may have been, words with secondary vowels in Old and Archaic Japanese were longer in pre-Old Japanese.

4.4.2 Expansion of Voiced Obstruents

Voiced obstruents in Old Japanese were, like half of the vowels, secondary in nature, as all of the original voiced obstruents were lost in an earlier stage of the language (Martin 1987: 20; Ramsey and Unger 1972). As first hypothesized by Asayama (1943) and as shown conclusively by Miyake (1999), Old Japanese voiced obstruents were prenasalized. Ultimately, they derived from voiceless obstruents preceded by nasals. The Old Japanese corpus has doublets showing variation between C_1VNVC_2V and C_1VC_2[+voi]V forms , as shown in examples 25 and 30. It is therefore reasonable to think that voiced obstruents were reintroduced in this way, and so it is possible to continue to reach back to an earlier stage by expanding words with voiced obstruents by an additional syllable. (Again, percentages do not add up to 100 due to rounding.)

monosyllables	1%
bisyllables	6%
trisyllables	18%
quadrisyllables	34%
pentasyllables	26%
six or greater	14%

Table 5: Percentages of pre-OJ Nouns after Expanding Voiced Obstruents

There is yet further reduction of the number of bisyllabic forms, from 10% to just 6%, and trisyllables also go down from 22% to 18%. Quadrisyllables also drop slightly,

from 37% to 34%, while pentasyllables are up from 20% to 26%, and words with six or more syllables go from 10% to 14%. At this point, 74%—nearly three-quarters—of all pre-Old Japanese words are four syllables or longer.

4.4.3 Breaking Apart Compounds

The results above will now be countered with another analysis: the removal of compounds from the calculation, which will serve to skew the percentages back down. This will have a tempering effect upon the previous modifications, though it is certainly too strong. As noted by Serafim (1977), some techniques available to internal reconstruction are very powerful, and they may project back to a time before proto-Japonic. In fact, the breaking apart of compounds is, in a sense, quite unrealistic, because compounding is perhaps a universal method for creating new words, and there have probably been no languages that have not made use of it. Therefore, these results should be taken with a large grain of salt.

For these calculations, I simply removed compound words from the database, then performed the expansions listed above anew.

monosyllables	4%
bisyllables	23%
trisyllables	43%
quadrisyllables	25%
pentasyllables	3%
six or greater	2%

Table 6: Percentages of pre-OJ Nouns after Removing Compounds

This recalculation has the effect of making most pre-Old Japanese words—fully 91% —two, three, or four syllables long, with close to half of the words in the middle at three syllables. As mentioned above, these figures are artificial in a way that the others are not, and some of the pre-forms created through this overly-powerful technique are certainly older than proto-Japonic. Even so, the picture that all these calculations draw is clear: both Old Japanese and its antecedents had words shapes that were fundamentally polysyllabic.

The jump in monosyllables, from 1% to 4%, is partially an artefact of the still-small size of the database. Whether 4% is too high or too low remains to be seen. Regardless, the ultimate percentage of monosyllables will be reduced by the elimination of forms which appear to be truncations of bisyllabic words, such as OJ *ka* 'deer' (cf. OJ *sika* id.), OJ *ka* 'oar' (cf. OJ *kadi* id. < **kaNVti*), and OJ *pa* 'feather, wing' (cf. OJ *pane* id.). The loss of an edge syllable of a longer word, thereby creating a doublet, can be found for other, longer words as well, such as OJ *katati* ~ OJ *kata* 'shape, form.' Forms such as these will be accounted for in future calculations made on the database. Anyone who

would like to see these as compounds must account for the 'other' part of the word, e.g., **si* '?' in OJ *sika* 'deer,' preferably with words attested in Old Japanese.

4.5 Further Analysis and Discussion

Another interesting set of forms are those associated with the words for *water*. Both OJ *myidu* 'water' and OJ *myina*[3] 'id.' are attested as independent words, though the latter most commonly appears in compounds. The truncated form *myi-*, which does not appear as an independent word, is also found in compounds, as in OJ *idumyi* 'wellspring,' which is a transparent compound consisting of OJ *idu* 'to come out' and *myi-* 'water.' Although some see the *na-* of *myina* as some sort of concatenative form in words like OJ *myinatwo* 'port' (OJ *two* 'entry') and OJ *myinamoto* 'source (of a river)' (OJ *moto* 'source, origin'), this analysis fails to account for the absence of putative *na-* in parallel forms like OJ *yamato* 'the Nara plain (which is surrounded by mountains); Japan' (OJ *yama* 'mountain') and OJ *sakamoto* 'base of a hill' (OJ *saka* 'hill'). If the loss of an edge syllable is an irregular change that accounts for certain doublets in Old Japanese, it is then possible to posit the form **myinatu* for 'water.' This word failed to survive, but two of its daughter forms did. OJ *myina* can be accounted for by irregular loss of the final syllable, and OJ *myidu* is due to the loss of the medial *-a-* and subsequent voicing and prenasalization of the following *-t-* (see § 3.3.2 and examples 25, 28, and 30). Etymologies such as this serve only to strengthen the claim that pre-Japanese and proto-Japonic words were longer, not shorter, than those attested in Old Japanese.

It is also possible that some Old Japanese monosyllables, such as *u* 'cormorant' and *wi* 'well,' could be from earlier bisyllables. To the best of my knowledge there is no evidence to suggest that these words have bisyllabic origins. However, it is possible that, for example, *u* 'cormorant' is from earlier **uCu*, and that *wi* 'well,' rather than being from simple pJ **byi* is instead from **bï* < **buCi* or **bəCi*[4] (the case for a bisyllabic origin for *wi* is stronger, because the A/B distinction was not found after *w*, so it is not possible to say if its earlier form was **wyi* or **wiy*). Of course, it would be foolhardy, unwarranted, and a violation of Occam's Razor to propose that all monosyllables must be from earlier bisyllabic forms. However, if there were compelling evidence in the form of possibly cognate words from other languages, words that would argue in favor of forms such as **uCu* and **buCi* or **bəCi*, then it must be acknowledged that the phonology of Old Japanese fully allows for such derivations.

5. Conclusion

Based on both the raw data and internal-reconstructive interpretation of those data, it is seen that the basic form of Old Japanese and proto-Japonic words was overwhelmingly

3 This appears as an independent word in the *Kojiki* ; my thanks to Alexander Vovin for pointing it out to me.

4 My thanks again to Alexander Vovin for suggesting this analysis of *wi*.

polysyllabic, with the majority of those being three or four syllables long. Both monosyllables and words of five or greater syllables make up a very small percentage of the corpus. Therefore, when looking for external relations to proto-Japonic, it would be most prudent to start looking at languages which also have predominantly polysyllabic forms, either synchronically or diachronically. Bearing this in mind, efforts to relate Japonic to Altaic in general and Korean in particular show the greatest potential. Although Japonic may have connections to Austronesian languages, the paucity of plausible Austronesian etymologies in Japonic (Vovin 1994, 2001) suggests a stratal or borrowing relationship is more reasonable than a genetic one.

Bibliography

Asayama, Nobuya (1943) Kokugo no kashira-onsetsu ni okeru dakuon ni tsuite (On Voicing of Initials in Japanese). *Kokugo to Kokubungaku* 20 (5): 16—26 (422—432)

Donegan, Patricia J. & David Stampe (1979) The study of natural phonology. In: D. A. Dinnsen, (ed.) *Current approaches to phonological theory*, Bloomington: Indiana University Press.

Erickson, Blaine (1998) *The origins and development of Japanese mora nasals.* Ph.D. Dissertation, Linguistics, University of Hawai'i, Honolulu.

Frellesvig, Bjarke (1995) *A case study in diachronic phonology: The Japanese onbin sound changes.* Aarhus: Aarhus University Press.

Hamada Atsushi (1946 [1986]) Sokuon enkaku koo (A study of the development of mora obstruents). In: A. Hamada, (ed.) *Kokugoshi no Shomondai* (*Problems in Japanese language history*), Osaka: Izumi Shoin.

———— (1952 [1984]) Hatsuon to dakuon to no sookankei no mondai: kodaigo ni okeru dakushiin no onka (The problem of the interrelationship of mora nasals and voiced stops: The phonetic value of voiced consonants in Ancient Japanese). In *Nihongo no Shiteki Kenkyuu* (*Historical studies in the Japanese language*), A. Hamada, ed. Kyoto: Rinsen Shoten.

Hashimoto Shinkichi (1917 [1949]) Kokugo kanazukai kenkyuushijoo no ichi hakken: Ishizuka Tatsumaro no kanazukai oku no yamamichi (A discovery in the history of research on *kana* spelling in Japanese: Ishizuka Tatsumaro's *innermost secrets of* kana *spelling*). In: Hashimoto Shinkichi Hakase Chosakushuu Kankoo Iinkai, (ed) *Moji oyobi Kanazukai no Kenkyuu* (*Studies in writing and Kana spelling*), Tokyo: Iwanami shoten.

Kasuga Kazuo (1980) Ishizuka Tatsumaro. In: Kokugo Gakkai, (ed) *Kokugogaku daijiten* (*Comprehensive dictionary of Japanese language studies*), Tokyo: Tokyodo.

Kishida Takeo (1942 [1984]) Joodai kokugo ni okeru boin onsetsu no datsuraku (The loss of vowel-only syllables in Ancient Japanese). In: T. Kishida, (ed) *Kokugo On'in Henka no Kenkyuu* (*Studies in phonological changes in Japanese*), Tokyo: Musashino Shoin.

Martin, Samuel E (1987) *The Japanese language through time.* New Haven: Yale University Press.

Miyake, Marc H (1997) Pre-Sino-Korean and Pre-Sino-Japanese: Reexamining an Old problem from a modern perspective. In: H.-M. Sohn & J. Haig, (ed) *Japanese/Korean Linguistics 6*, Stanford: Center for the Study of Language and Information.

———— (1999) *The phonology of eight century Japanese revisited: Another reconstruction based upon written records.* Ph.D. dissertation, Linguistics, University of Hawai'i, Honolulu.

Okimori Takuya, ed. (1989) *Nihongoshi (History of the Japanese language)*. Tokyo: Oofuusha.

Ramsey, S. Robert, & J. Marshall Unger (1972) Evidence of a consonant shift in 7th century Japanese. *Papers in Japanese Linguistics* 1 (2): 278—295.

Seeley, Christopher (1991) *A history of writing in Japan*. Leiden: E. J. Brill.

Serafim, Leon A (1977) *The relationship of Pre-Japanese and Proto-Japanese*, ms. New Haven: Yale University.

———— (1978) *Doublets in Japanese native vocabulary: A contribution to the consonant phonology of Prehistoric Japanese*, ms. New Haven: Yale University.

———— (1999) *Why Proto-Japonic had at least six, not four, vowels.* Department of Linguistics Tuesday Seminar Presentation, Honolulu: University of Hawai'i at Manoa.

———— (nd) *Japanese language history in a nutshell*, ms. Honolulu: University of Hawai'i.

Shiraishi Taiji (1957). Motoori Norinaga. In: Kokugo Gakkai, (ed) *Kokugogaku Jiten (Dictionary of Japanese Language Studies)*, Tokyo: Tokyodo.

Thorpe, Maner Lawton (1983) *Ryukyuan language history*. Ph.D. dissertation, Linguistics, University of Southern California, Los Angeles.

Unger, J. Marshall (1977) *Studies in Early Japanese morphophonemics*. Bloomington: Indiana University Linguistics Club.

———— (2000) Reconciling comparative and internal reconstruction: The case of Old Japanese /ti, ri, ni/. *Language* 76 (3): 655—681.

Vance, Timothy J (1982) On the origin of voicing alteration in Japanese consonants. *Journal of the American Oriental Society* 102 (2): 333—341.

Vovin, Alexander (1994) Long-distance relationships, reconstruction methodology, and the origins of Japanese. *Diachronica* 11 (1): 95—114.

———— (2001) North East Asian historical-comparative linguistics on the threshold of the third millennium. *Diachronica* 13 (1): 93—137.

Whitman, John (1996) *Comments on the origin of dakuon*, ms. Presentation at Koreo-Japonic Circle Meeting, Honolulu.

Yamaguchi Yoshinori (1971) Kodai Nihongo ni okeru boin datsuraku no genshoo ni tsuite (On the elision of vowels in Ancient Japanese). *Kokugogaku* 85: 1—17.

上代日本語および日琉祖語の語構造

ブレイン・エリクソン
熊本学園大学

上代日本語、およびそれ以前の日本語における単語の形と長さを明らかにするために、上代日本語の音韻と形態を検討する。再構された語形は、音韻過程と形態規則をまったく逆に適用することによって、提示されている。その再構された語形の原型は、これよりも長い音節や短い音節があるにもかかわらず、３音節または４音節、つまりCVCVCV(CV)の形をもつ。

The phonology and morphology of Old Japanese is examined to reveal information about word shape and word length in Old Japanese and its predecessors. Reconstructed forms are adduced through the reverse application of phonological processes and morphological rules. Archetypal forms are three or four syllables long, of the form (C)VCVCV(CV), though both longer and shorter forms exist.