APOCOPE OF LATE OLD CHINESE SHORT *ă:
EARLY CENTRAL ASIAN LOANWORD AND OLD JAPANESE EVIDENCE FOR OLD CHINESE DISYLLABIC MORPHEMES

CHRISTOPHER I. BECKWITH – GISABURO N. KIYOSE†

C. I. Beckwith
Indiana University, 355 N. Jordan, Bloomington, IN, 47405, USA
e-mail: beckwith@indiana.edu
G. N. Kiyose †
University of Hawaii, Emeritus

The morphophonology of Old Chinese has usually been reconstructed as an earlier version of the traditional reconstruction of Middle Chinese, with exclusively monosyllabic morphemes. For Old Chinese some scholars have posited syllabic morphemes with long or short vowels, or even polysyllabic morphemes, and other theories have been proposed, but it is still assumed that by Late Old Chinese any longer morphemes that once existed were already monosyllabic. However, Central Asian loans in Chinese suggest some disyllabic morphemes still existed in Late Old Chinese. They seem to be confirmed by a new study of little-noted Old Japanese transcriptions known as nigō-gana. Thus the hitherto problematic Old Chinese and Old Japanese transcriptions of foreign words such as Saka and Buddha, and the monosyllabic theory of Old Chinese morphology, must be reconsidered. This paper’s new reconstructions conform to the data and have great implications for the reconstruction of Old Chinese as well as for the reconstruction of the early Chinese loans into Japanese.

Key words: Old Chinese, Old Japanese, Central Asian loans, Saka, Buddha.

Over half a century ago Barnabás Csongor (1952; 1955; 1960; 1962) published a series of articles in the pages of this journal presenting the phonetic values of some morphemes in Middle Chinese by investigating the segmental transcriptions of the language in Uighur, Tibetan, and Brahmi scripts. Then, thirty years later, working on the same material, Tokio Takata published a book and eventually other studies, which together include more texts and thus more transcribed Middle Chinese forms. These revolutionary works establish beyond doubt the phonetic value of many phonemes in Middle Chinese.

* Acknowledgements: We would like to thank our colleagues and the anonymous reviewers for their helpful comments and corrections. We are of course responsible for any errors.
Although recent work (Kiyose – Beckwith 2008; Beckwith 2014; 2016b; 2017; Shimunek 2017, pp. 370, 389–390, 399) has presented evidence that some morphemes were disyllabic in Late Old Chinese, unfortunately there are still literally only a few Old Chinese words transcribed in foreign segmental scripts in the Old Chinese period itself. However, work on Old Japanese shows that some early man’yōgana characters regularly transcribe disyllabic Late Old Chinese forms.

It is well known and accepted that monosyllabic Middle Chinese words with a coda (final consonant) are transcribed in Sino-Japanese with the regular addition of an epenthetic vowel, typically i or u. Although putative exceptional vowels have received some attention, their implications for the early history of the donor language, Chinese, have mostly been overlooked.

Takashi Kamei (1984) presents the accepted understanding that disyllabic Sino-Japanese forms of the shape CVCV with, e.g., mu and ni in the second syllable (i.e, CVmu, CVni) written with one character, or nigōgana (二合仮名), derive from transcriptions of forms in monosyllabic Middle Chinese (the donor language) having the nasal codas m and n, to which the regular paragoge (epenthetic) vowels u and i were added in Old Japanese (the receptor language). This is not problematic. However, Kamei’s article discusses only the labial nasal coda and apical nasal coda of Middle Chinese character readings and nigōgana derived from them, of which almost all the second syllables of the former are mu/mi and those of the latter are nu/ni (i.e., CVmu/CVmi, CVnu/CVni). Accordingly, nigōgana readings such as 男nama and 信sina are exceptions for Kamei and the traditional approach. Although he does not discuss other nigōgana in his article, examples with exceptional final a exist after other consonants too.

Based on data from Old Japanese (nigōgana) and from Late Old Chinese itself (especially alternate transcriptions of the same foreign word), we hypothesise, first, that the a endings in nigōgana treated as exceptions by Kamei, as well as exceptions of the shape CVCa that he does not discuss, are not examples of paragoge at all, but retentions of forms with the shape CVCa (i.e., with a final short *ă syllable) in the original donor-language, Late Old Chinese; second, that this short final *ă is also attested after other consonants in nigōgana; and third, that forms of the shape CVCa are also attested as loanwords in Old Japanese, such as saka ‘border, frontier’, from a disyllabic Chinese source.

If our only data were exceptional Japanese nigōgana forms, it would be difficult to be certain that they represent disyllabic Chinese pronunciations. But in addition to many examples of Chinese “monocharacter” transcriptions of disyllabic foreign words (i.e., one character transcribing two syllables, like nigōgana), there are also Chinese “dicharacter” transcriptions of the same foreign words (i.e., two characters each transcribing one syllable), which show that the foreign forms were heard by the Chinese as disyllabic forms and transcribed by them in both ways as disyllabic forms, as discussed below.

In recent work on Chinese reconstruction a number of scholars (e.g., Shafer 1966–1974; Schuessler 1987; Beckwith 2007; Kiyose – Beckwith 2008) have argued that the structure of the semantically heavy morphemes (nouns, adjectives, verbs) of
Old Chinese must have been, at one time, predominantly disyllabic, as they still are today in Mandarin (Beckwith 2015b). Some (Pulleyblank 1962; Starostin 1989) have argued instead for a monosyllabic model with long and short vowel phonemes, resulting in a contrast between monomoraic (short) and bimoraic (long) syllables. That model may be supported by the fact that even in modern Mandarin, syllables in the third tone (a phonetic heir of the Middle Chinese shǎngshēng 上聲 ‘rising tone’) are much longer in duration than syllables in any other tone, so that some consider such syllables to have three moras. Y. R. Chao (1968) and other linguists have noted that a “monosyllabic” third tone “often breaks into two syllables” with a glottal stop in between, e.g. nǐ 你 ‘you’ > [niːʔi] (Duanmu 2000, pp. 232–233), though the break is more likely a creaky tone. However, it suggests that rising tone syllables actually descend from the Late Old Chinese form *CVʁV attested by early Chinese loanwords into Japanese and Tibetan (Kiyose–Beckwith 2008), by many loanwords from various Central Eurasian languages, and by a word borrowed from Aramaic into Late Old Chinese, attested in several transcriptions, and borrowed from Chinese, probably via Koguryo, into Japanese, Jurchen, and Middle Korean (Beckwith 2014; 2017).

According to traditional HSR, by contrast, the Old Chinese transcriptions of foreign names from the early Former Han Dynasty (206 BC–9 CE) written with a single Chinese character were pronounced with a single syllable in Old Chinese times. For example, it has long been established that the early Western Han period Old Chinese transcription 塞 MSC sè ~ sāi ~ sài ‘to block up, stuff; border, frontier’ from MChi *sǝk (P. 273, 271)1 corresponds to Scythian dial. Saka ‘Scythian (an ancient Iranic-speaking people of Central Eurasia); eastern Scythian’. The name is earliest attested in Old Persian, as Saka (late 6th century BC); in Greek, as Saka- (5th century BC to 6th century CE); and in Gandhārī, as Saga (ca. 100 CE, from earlier Saka, cf. Pali Saka and Sanskrit Śaka),2 which explicitly reveals the preservation of the second syllable vowel a because k is regularly voiced intervocally in Gandhārī.3 Yet the Chinese transcription of the name (ca. 135 BC) is reconstructed by HSR as MChi *sǝk (P. 273) from OChi *sǝk (S. 551).

Known loanwords related to Buddhism are solidly attested in Chinese from the mid-2nd century CE on, though some were evidently introduced slightly earlier. All are traditionally reconstructed as monosyllables. For example, the usual putatively monosyllabic Late Old Chinese transcription (2nd century CE) of Gandhārī buddha (Pali buddha, Sanskrit buddha) is now read 佛 MSC fó and reconstructed in HSR from theoretical MChi *but (P. 96) from putative OChi *but ~ *bud4. As for Gandhārī

1 The term “Early Middle Chinese” (e.g., in Pulleyblank 1991) refers to literary character readings collected in the lost Qieyun (601 CE). We use the term to refer to the period before the Qieyun.

2 The name Saka is clearly yet another dialect form of North Iranian *skuða ‘shooter (i.e., “archer”), the source of the English names Scythian, Sogdian, etc. (Szemerényi 1980), as shown in Beckwith (2009, pp. 377–380).

3 Gandhārī forms are cited from Baums–Glass (2012), Pali from Rhys Davids–Stede (2012). We are indebted to Andrea Schlosser (p.c. 2017) for helpful information on the interpretation of the Gandhārī forms.

4 On voiced Old Chinese codas in HSR (Zhengzhang 1991) see note 12.
Samgha (Saṅgha) ~ Sagha (Pali Saṅgha, Sanskrit Samgha) ‘Buddhist monk, the community of monks and nuns’, the Late Old Chinese transcription, now read 僧 sēng, is traditionally reconstructed as from MChi *səŋ (P. 273) from OChi *səŋ. There are many more examples.

It is generally believed that such Late Old Chinese transcriptions represent foreign words which had already lost their final short *ā vowel. That is, it is usually thought that the native forms of these words in the foreign languages themselves were actually *sək, *səŋ, etc. respectively – as some words did come to be written many centuries later in mediaeval works in some of those languages. Alternatively, they are assumed to be abbreviations of the foreign words by the Chinese, who spoke a “monosyllabic language” and liked monosyllabic words better than longer forms.

The first argument is contradicted by the actual attested non-Chinese data in Iranian, Indic, Greek, and Aramaic texts. We find no examples of such shortened forms in these languages until well into the Middle Ages, if then. The language cited by Chi Hsien-lin (1947) as the source for the monosyllabic forms of Buddhist words transcribed in Chinese is Tokharian A, which is not firmly attested before the middle of the first millennium CE (solidly in the Middle Chinese period), but in any case it is now accepted that the crucial earliest Chinese translations were made from Gāndhārī Prakrit, not Sanskrit or the local Central Asian languages. Moreover, the origin or ancestry of all the early translators was from further west, in the Kushan or Parthian empires, and although some apocope did eventually occur, the words in question still have their final a within their Indic and Iranian languages even in Late Antiquity.

If the monosyllabic forms of disyllabic foreign words – which are attested as disyllabic in the foreign languages themselves – are abbreviations made by the Chinese, surely one must ask whether the Chinese did not sometimes transcribe them with two monosyllabic characters, as they could if Old Chinese morphemes were largely monosyllabic. In fact, there actually are very many examples, in Late Old Chinese, of just this kind of “dicharacter” transcription, i.e., using two characters to write disyllabic loanwords or to transcribe foreign names and other words. For example, the Gāndhārī word for ‘Buddha’ could be and was transcribed just as early with two characters as with one character, necessarily indicating that the Chinese transcriptions represent an underlying disyllabic foreign word *[bu.da]. For example, Buddha is also transcribed 浮屠 MSC fútú from MChi *buw.-dɔ (P. 98, 311) from OChi *būdā (dated 65 CE) and 浮圖 MSC fútú from MChi *buw.-dɔ (P. 98, 311) from OChi *būdā (dated 193 CE). These early transcriptions do not record a form such as *[bud.da] (which would reflect Sanskrit Buddha) but rather *[bu.da], reflecting Gāndhārī budha [bu.dɦa], as expected. Why then did they not always use two characters to transcribe the word in the second and third century translations, but sometimes preferred the apparently monosyllabic transcription, 佛 MSC fŏ, supposedly reflecting a theoretical Late Old Chinese *bud ~ *but, of the solidly disyllabic word attested

---

5 For the sources see Zürcher (2007, pp. 26, 329).
6 Sanskrit did not become the dominant language of Buddhist texts in northern Buddhism for perhaps two more centuries after this time, when the Chinese increasingly transcribed Sanskrit forms, including, eventually, [bud.da]. See Boucher (1998, p. 477, note 38).
APOCOPE OF LATE OLD CHINESE SHORT *ā

in exactly contemporaneous Gāndhārī (as Budha), Bactrian (as Boddo), Greek (as Boutta), and in Chinese itself (as 浮屠 *būdâ, etc.)?7

The reason seems to be that using two characters gave two equally stressed syllables, whereas disyllabic words in Iranian, Indic, and many other languages – evidently including Late Old Chinese – normally had only one stressed syllable. Accordingly, the monophthongic transcription Fo 佛 ‘Buddha’ represents *būda /ˈbu.da/ – i.e., *būdâ [ˈbu.dǝ]. This reveals the motivation for the later apocope of short *ā in Chinese, as well as in some mediaeval Central Asian Iranian and Indic languages, namely stress. In the donor languages and in the borrowing language (Chinese), the first syllable of these words was stressed, but if no appropriate monophthongic transcription was available, the Chinese had to use two or more characters anyway. Of course, although some characters were used as monophthongic transcriptions of disyllabic foreign language words, the same characters wrote Chinese words, too. That is, there must also have been monophthongic writings of native Chinese disyllabic words in Old Chinese.

The dicharacter transcriptions thus tell us that the foreign pronunciation of foreign disyllabic words was still disyllabic and was known to the Chinese in disyllabic form, and that the monophthongic transcriptions which the Chinese used for these particular words actually had disyllabic readings in Old Chinese, and such readings continued to be disyllabic in Late Old Chinese, surviving in some dialects into the beginning of the Middle Chinese period. The monophthongic transcriptions are often written with distinctive transcribing characters, too. For example, Gāndhārī Budha ‘Buddha’ is written both 佛 MSC ㄈ and 浮屠 MSC ㄈ, as pointed out long ago (Chi 1947), the former being read in MChi  but (P. 96) from LOC *būda (i.e., *būdâ), and the latter read in MChi buw-*dɔ (P. 98, 311) from LOC *būdâ [ˈbu.ˈdâ].

Further support for this analysis comes from traditional scholarship, in which one may find discussion of examples such as 不 + 之 → 弗, in which each character is read monosyllabically. The usual explanation is that 弗 is a contraction of two monosyllables, and the result must have been read as a monosyllable in Old Chinese. However, the example itself actually suggests that the monophthongic 弗 originally represented two (or more) syllables, even if one of them was unstressed. This is solidly supported by the graphically derived, phonetically related character 佛 LOC *būdâ [ˈbu.dǝ], which became the usual Chinese representation of Gāndhārī budha ‘the Buddha’.

7 The name Buddha is earliest attested in Western sources in the Stromata of Clement of Alexandria (fl. 180–211 CE), where it is written Boutta, i.e., [but.ta]; see Beckwith (2015a, p. 100) for the text and commentary. For a study of the coin of Kanishka I (ca. 127–152 CE) showing a Buddha image and the name Boddo ‘Buddha’ in Bactrian, see Cribb (1999–2000).

8 See Beckwith (2014) for examples that strongly support this proposal. It is also likely that the final short *ā was (or became) phonetically *[a], which might not have been available for dicharacter transcriptions, as suggested by the reduction of unstressed final a in foreign ‘Buddhist’ words borrowed in monophthongic transcriptions (cf. Beckwith 2017).

Examples such as *būdâ and *sākâ suggest that earlier, both voiced and unvoiced intervocalic stops existed, but they were neutralised along with the apocope of final short vowels at the end of the Late Old Chinese period.

Acta Orient. Hung. 71, 2018
Literary Chinese later completed the monosyllabicisation of its morphemes so that by the Sui-Tang period no monochacter (i.e., disyllabic) readings of single characters appear to have survived. For instance, the Late Old Chinese disyllabic reading of 佛 as *būdā had become the theoretical Middle Chinese monosyllabic reading 佛 but – actually attested as [bur], later [fur], etc. in foreign transcriptions of Late Middle Chinese (Takata 1988, pp. 382–383), a hitherto unexplained phenomenon (v. infra.). However, the dicharacter transcriptions of these disyllabic words remained in use too, written as before with two characters, but in Middle Chinese times each character came to be viewed as representing one morpheme, so to the Chinese of the period Old Chinese was marked by an exotic “monosyllabicism”, which, as lovers of the ancient and arcane, they strove to emulate. Belief that Chinese has always been “monosyllabic” continues to drive virtually all reconstructions of the language.

A fair number of examples of monochacter transcriptions of disyllabic words can already be identified for Old Chinese, from the Warring States (Classical) period to the point in Late Antiquity when spoken Late Old Chinese became spoken Proto-Mandarin and the literary language became Middle Chinese, i.e., some two to three centuries before the Qiéyun.

As mentioned, the Scythian ethnonym Ṣaka is transcribed in Old Chinese with the single character 塞 MSC sè ~ sāi ~ sài ‘to block up, stuff; border, frontier’ from MChi *sǝk (P. 273, 271), reconstructed as from traditional OChi *sǝk (B. 785: *sik [i.e., *sǝk in Baxter’s later system that follows Starostin’s]; S. 551: *sǝk). It is one of many Central Eurasian names recorded in Chinese and Central Asian or Near Eastern languages or Greek in Antiquity. In Western languages, these words are disyllabic. Only the Chinese transcriptions are interpreted as representing exclusively monosyl-

The early Japanese linguistic tradition implicitly recognises the above phenomena, though its evidence has largely been overlooked, Chinese 博 MSC bó ‘extensive’ is reconstructed in the traditional method from MChi *pak (P. 41; B. 748) from OChi *pak (B. 748) ~ *pāk (S. 565). In Japanese, however, from Old Japanese times to the present, many native disyllabic morphemes and some disyllabic foreign loan morphemes are written with a single character that has a disyllabic nigōgana reading. A well-known modern example is haka- 博 OJpn *paka- in hakase 博士 ‘learned master (now usually in the sense “Ph.D.”)’ (JDB p. 572), which is also used to “phonetically” transcribe haka- in the native Japanese place name Hakata (from OJpn *Pakata). The same phenomenon is attested with loanwords. For example, that Chinese 博 MSC sè ~ sāi ~ sài ‘to block up, stuff; border, frontier’ – all from MChi *sǝk (P. 273, 271) – which was used in Old Chinese times to transcribe the name Ṣaka ‘Eastern Scythian’, as discussed above, must actually have been pronounced *sǝkā ~ [’sa.kǝ], not *sǝk, in Old Chinese is shown by OJpn *saka ‘border, frontier’, a loan from OChi *saka ‘id.’ Though written “semanti-

10 In Old Japanese the apico-alveolar fricative onset s and affricate onset ts appear to have been in free variation (Kiyoše 1989a, p. 143; 2013, p. 186).
cally” with the character 境 ‘border, frontier’, OJpn *saka (JDB p. 318; cf. Martin 1987, p. 516)11 is an exact synonym of 塞 OChi *sáka ‘border, frontier’.

The phenomenon of nigōgana (二合仮名) has been practically unstudied in Japanese linguistics, despite its recognition and naming (as “nigō no kana”) already by Motoöri Norinaga (1730–1801) in his paragraph “Kana no koto” [‘On kana’] in Volume 1 of his famous Kojiki-den (Motoöri 1790).

Because the man’yōgana writing system in Japan came into use between the 5th and 7th centuries and was highly developed in the 8th century, the readings of some characters used as kana reflect the phonology of an earlier period of Chinese and others a later period. In addition, since almost all the scribes in Japan at the time are thought to have come from various regions of the vast land of China, the sound of man’yōgana characters transcribed by a given scribe must have phonetically reflected the Chinese dialect of his native region. Thus, for example, the man’yōgana character 翼 ‘cloud’ is given as uru (in the *VrV section) and una (in the *VCa section) below (Ōno 1977; Osterkamp 2011). This variety is reflected in nigōgana, too.

Use of the nigōgana reading haka 博 OJpn *paka in hakase 博士 to write native Japanese haka- (OJpn *paka-) in the place-name Hakata indicates unambiguously that this particular Chinese literary borrowing was heard as two syllables, even though according to the usual rule for mediaeval Chinese borrowings, which are in many cases disyllabic today, all are thought to have been borrowed from Chinese monosyllables. Thus, because Old Japanese was phonotactically an open-syllable language that allowed no codas, a vowel was added by paragoge – either i (OJpn i/i) or u, both of which are high vowels, the least sonorous – to monosyllabic Chinese loan-words with codas, so as to avoid closed syllables (i.e., with codas). Also, o (OJpn o/ö) was in a few cases added by paragoge to avoid the same kind of syllables as above, because OJpn o is internally reconstructed as a high back vowel [u] phonetically (Kiyose 1982; 1991, pp. 198–210; 1995, pp. 145–153). OJpn ö was a mid-vowel, not a high vowel, unlike o, but the vowel ö could be added by paragoge in the same way as above (while o could not) if the vowel ö preceded in the first syllable, due to a kind of “vowel harmony”, the Arisaka-Ikegami law.12

Why then does the word haka 博 have a low vowel, a, which is in fact the most sonorous vowel, as its second vowel? There is no regular rule for a paragoge in Japanese, and the final vowel a occurs unpredictably after any first syllable vowel (v. infra). The word must have been borrowed as *paka – perhaps along with other early Buddhist borrowings – before the great mass of literary loans from Middle Chinese, when exactly the same word was reborrowed, but treated as a monosyllable with a stop coda *k, to which Japanese added the vowel *u according to the rule for

11 This is considered a “native” or kun ‘semantic’ Japanese character reading, not a “borrowed” or on ‘phonetic’ Sino-Japanese reading. In modern Japanese it is usually sakai ‘border, boundary’ from 佐加比 OJpn *sakapi (JDB p. 320; cf. Martin 1987, p. 516), morphologically consisting of the noun saka ‘border’, plus the iterative/durative verbal suffix -ap (Kiyose 1989, p. 149; 2001, p. 249; 2013, p. 191), forming the stem sakap- ‘to border’, to which the nominaliser -i (Kiyose 1989a, pp. 13–14; 1995, pp. 11–12, 53; 2013, p. 79) is suffixed, making *sakapi.

12 See Kiyose (1989b) for the inferred sound values of all eight OJpn vowels.
paragoge, producing the alternate reading also preserved in modern Japanese, *haku*, alongside the more conservative form *haka*. The review of Sven Osterkamp’s important book (2011) by Heiko Narrog and John Whitman (2013, p. 72) states that many Middle Chinese syllables, unlike Modern Standard Chinese, ended with a consonant or semivowel, while Old Japanese syllables are open, concluding: “Thus, for example, the Chinese character 博; MC[hi] /pak/ could be used to render the OJ[pn] mora sequences /paka/ and /paku/. As seen in this example, the second vowel in the sequence was freely added.” However, in Old Japanese – the language in and before the Nara period (A.D. 710–784) – the reading of the character 博 is *paka only, never *paku (Ōno 1977, pp. 453–454); the reading *faku for 博 appears later in the Heian period (A.D. 794–1192) in a place name (Osterkamp 2011, p. 84).

However, there were also more distinctions in syllable types in 8th-century Japanese than in later periods. As an orthographic convention in Old Japanese, sets of man’yōgana – Chinese characters used for their phonetic value – functioned as a kind of syllabic phonogram system marking two values, type-A and type-B, for what became MSJ *i, e*, and *o* in certain syllables.

For example, the Old Japanese predecessor of the modern syllable *ki* in MSJ aki ‘autumn,’ yuki ‘snow,’ kinu ‘cloth,’ kimi ‘chief,’ etc. is transcribed in Old Japanese with one of the characters classified as the “type-A *ki,*” such as 企, 伎, 岐, 吉, 企, etc. However, the Old Japanese predecessor of the modern syllable *ki* in MSJ kiri ‘fog,’ tuki ‘moon,’ ki ‘tree,’ kuki ‘stem,’ etc. is transcribed in Old Japanese with one of those classified as the “type-B *ki,*” such as 紀, 奇, 忌, 歸, 納, etc. The sound value of the type-A vowel *i* must have been approximately the same as that of the later *ı*, i.e., a high front-vowel, and the type-B *ı* must have been high, too, but a little further back, toward the central position or the like. In Latin transcription, it is customary to place a diacritical mark – umlaut – on the Type-B *ı* i.e., ı in tukı ‘moon,’ but no mark on the Type-A *ı*, as in aki ‘autumn’. Similarly, the Old Japanese front vowel which became MSJ *e* had the A–B distinctions in the 8th-century after velars (*k, g*) and labials (*p, b, m*), and the Old Japanese mid-back-vowel which became MSJ *o* had the A–B distinction after velars (*k, g*) and dentals (*t, d, s, z, n, r, y*, and in the *Kojiki* (A.D. 712) after *m*. The umlaut is added to each of these vowel letters in the B-series in Latin transcriptions of Old Japanese. After the Nara period, all Type-B vowels merged with the corresponding Type-A vowels. Until then, there had been two mutually exclusive groups of man’yōgana characters, type-A and type-B, in specific sets of syllables. These distinctions were quite consistent, and characters were not freely interchanged. In fact, almost no scribal errors with respect to the A–B distinction are seen in any documents produced during the Nara and preceding periods.

In short, Narrog and Whitman’s (2013, p. 72) statement that the second vowels of putatively monosyllabic Middle Chinese words written with nigōgana characters were “freely added” cannot be accepted.

---

13 Some Old Japanese reconstructions write the vowel ı as wi or the like. Martin (1987) writes it as iy (contrasting with the vowel *i*, which he writes as *yi*).
Although the traditional understanding of this historical phenomenon is well known, it may be helpful here to first give some examples of Chinese characters for syllables that have stop codas in Middle Chinese, but were made pronounceable for Old Japanese speakers by paragoge, producing disyllabic Old Japanese forms. For example: OJpn *sapi 杂 (Ōno 469; Ost. 46) from MChi *dzǝp/ǝdzǝp (P. 391); OJpn *tapu 塔 (Ōno 470; Ost. 30) from MChi *tǝp (P. 299) ‘stupa’, a known transcription of a form of the Indic word stūpa (thus perhaps a nigōgana transcription of a LOC *tapu that actually retains both metathesised vowels); OJpn *swot 越 (Ōno 432, 465; Ost. 53) from MChi *swuat 薯 (Ōno 462, 510; Ost. 58) from MChi *thap (P. 299) ‘stupa’, a known transcription of a form of the Indic word stūpa ‘a being (who aims to achieve) enlightenment (bodhi)’ (thus perhaps actually a nigōgana transcription of a LOC *satu or the like); OJpn *woti 越 (Ōno 432, 465; Ost. 53) from MChi *wuat 薯 (Ōno 462, 510; Ost. 58) from MChi *thap (P. 299) ‘stupa’, a known transcription of a form of the Indic word stūpa ‘a being (who aims to achieve) enlightenment (bodhi)’ (thus perhaps actually a nigōgana transcription of a LOC *satu or the like); OJpn *sikï 式 (Ōno 483–484; Ost. 78) from MChi *śik (P. 285: ɕik); OJpn *raku 楽 (Ōno 488–489; Ost. 77–78) from MChi *lak (P. 185); OJpn *tökö 德 (Ōno 427, 453) from MChi *tǝk (P. 74); etc.

Note also that the reconstruction of all instances of the Middle Chinese final coronal stop as *t has long ago been shown to be incorrect. Depending on dialect and period, it was usually *r in Chinese, and some examples are attested in nigōgana, e.g. OJpn *tari 達 (Ōno 437; Ost. 177), also read *tata (Ōno 463; Ost. 60), from MChi *dǝr (T. 362–363) ~ *dat (P. 69); and OJpn *pari 播 as in *parima 播磨 (Ōno 838–839; Ost. 189), which is not from the Middle Chinese reading *pa3 (Ost. 189; P. 42) or the analogous misreading *pan ~ *ban (Ost. 189) but from MChi *par (not attested in T.) ~ *pat (P. 40), i.e. actual *par. Osterkamp (2011, p. 50) discusses this issue, mentioning the fact that transcriptions in Tibetan, Uighur, and Khotanese segmental scripts, as well as in Sino-Korean, all regularly represent HSR MChi *t as r. Also striking is the existence of a good number of nigōgana with Vr V, as in OJpn *tsuru 駿 (Ost. 389–389; Ost. 189), which is not from the Middle Chinese reading *pa3 (Ost. 189; P. 42) or the analogous misreading *pan ~ *ban (Ost. 189) but from MChi *par (not attested in T.) ~ *pat (P. 40), i.e. actual *par. Osterkamp (2011, p. 31, note 7) discusses this, but concludes “für den Gebrauch des Zeichens nach der Entlehnung ins Japanische ist die allerdings irrelevant”.

14 Osterkamp (2011, p. 31, note 7) discusses this, but concludes “für den Gebrauch des Zeichens nach der Entlehnung ins Japanische ist die allerdings irrelevant”.

15 The misreading *pan 播 is not found, but only ban is seen in MSJ, used as an abbreviation for Harima (see above). Otherwise the character 播 ‘to seed’ is read regularly as ha in MSJ.

16 Tōdō (1965, p. 657; 1977, p. 553) gives OChi *puar > MChi *pua.
OJpn *pari from 半 MChi *pan3 (P. 29) < LOC *par(i); OJpn *maro from 滿 MChi *man3 (P. 207) < LOC *mar(o); etc.

The period before Early Middle Chinese was Late Old Chinese, when the language clearly still had some disyllabic morphemes. The loanword *saka 境 ‘frontier, border’ is striking, as it shows the word had already been completely nativised long before, accounting for the Japanese choice to write it “semantically” with 境 MSC jìng from MChi *kiaŋ2 (P. 159) – rather than with 塞, though the date of the borrowing is unknown. It is further known that some Old Chinese words for things that are not native to Japan were imported from the Asian mainland along with the things themselves in the pre-Old Japanese period, so it is not surprising that they are often earlier Old Chinese loans and thus preserve anomalies which, as Meillet (1984) stresses, are crucial to historical linguistic reconstruction. Some words, which are written in Old Japanese with the characters associated with them in Old Chinese at the time of borrowing, are solidly disyllabic as well, and are thus effectively a kind of nigōgana, not kun ‘semantic’ readings.17

The following preliminary list gives examples of forms shortened by apocope of *ā after Late Old Chinese, as confirmed by external transcriptions or loanwords. An acute accent [́] marks the stressed syllable preceding the final unstressed *ā that was later deleted.18 Syllables with a long vowel, which is assumed to have automatically carried or attracted the stress accent, are not further marked. The doomed short allophone of /a/ is written *ā.

Examples

Scythian *saka : 境 MSC sе ~ sāître < MChi *sak ‘border, frontier; Saka, Scythian’; cf. OJpn *saka ‘border, frontier (境 · 境)’ (JDB pp. 318, 320) ← OChi *saka 境 ‘border, frontier’.

Scythian hinduka : 天竺 MSC tiānzhú < MChi *t’en1 (P. 306), read like 正 MSC xiān < MChi *xɛn (P. 334) + MChi *tuwk (P. 414) < OChi dial. *hendukă ‘Hinduka; India, Indian’.19

Hsiung-nu bagatur : 蒙古語 MSC mòdùn < MChi *mǝk (P. 217) + *twǝn3 (P. 84) < OChi *mǝ́gǝur ‘p.n. or epithet (Old Iranian baga ‘god, lord’ + unidentified morpheme *tur ~ *twir)’.

---

17 See Kiyose – Beckwith (2008) for examples and chronology.

18 This form of notation is used here (rather than the IPA system), because it does not require explicit marking of syllable boundaries. For longer words (however transcribed in Chinese), the stress was likely more complicated.

19 Identified by Pulleyblank (1991, p. 414); he gives 竹 MChi *tuwk [tuwk] (P. 414), an exact homonym of 蒙. However, Old Iranian had a phonemic distinction between voiced and unvoiced stops, so the transcription attests an additional voiced-onset reading of 蒙 in a Late Old Chinese dialect.

20 This has long been thought to be an Old Chinese transcription of the Central Eurasian culture-word bagatur ~ bagyat ‘hero’ (Beckwith 2009, p. 387).

Gāndhārī *budha : < MChi *but (P. 96) < LOC *būtṣa ~ *būtsa ‘Buddha (the awakened one)’.23

Gāndhārī *dharma : < MChi *dama ~ *dam (P. 300) < LOC *dāmā ‘Dhamma, Dharma’.24

Gāndhārī Śaka- (Baums – Glass 2012) in Śākamuni- ‘the Śaka (Scythian) sage’ ~ Pali Sakkamuni ~ Sanskrit शाक्य (in Śākyamuni):25 釋 MChi *si < MChi *siājk (P. 285) ‘Śākya; Buddhist’ (Sanskrit abbreviation for the Buddha’s traditional putative clan, Śākya); 仏 in OJpn *Wesaka 仏 (Ost. 102) < LOC *śākā.

Gāndhārī *samgha ~ saṅgha (Baums – Glass 2012): 僧 MChi sēng < MChi *sāŋ (P. 273) < LOC *sāŋga ~ *sā̆ŋga ‘the saṅgha (saṅgha, = Gāndhārī samgha /saṅga/)’, the community of Buddhist monks and nuns; cf. 僧伽 MChi sēngjī < MChi sī/saṅ (T. 396–397) ~ *sāŋ (P. 273) + 伽 MChi ga (T. 306) ~ *gā (T. 306; not in P.) < LOC *sāṅga ‘the saṅgha’.


Gāndhārī godama ~ ghudama (Baums – Glass 2012): 精舍 MChi *gū (P. 260) + MChi *dam ~ *dām (P. 300) < LOC *gudāmā ‘the personal name of the Buddha (Sanskrit Gautama)’ (Beckwith 2015a, pp. 118–121); cf. ‘dharma’ above.

---

21 Beckwith (2014), q.v. for the reconstruction, dicharacter transcriptions, and the alternate early monocharacter transcription of the 相 sī (incorrectly sī in the article); for a detailed study of this transcription see Beckwith (2017). Both transcriptional words have the same onset and belong to the same Old Chinese rhyme (K. 255–256, No. 967d and No. 972h). The pre-Middle Chinese dicharacter transcriptions of the Aramaic donor form *dērā include 仏舍 MChi jīnglu < pre-EMC *dzērā; 寺舍 MChi *sī < pre-EMC *dzērā; and 仏舍 MChi jīngshè < pre-EMC *dzērā, all meaning ‘monastery’. For word-internal coda deletion and the reconstruction of 相, see Beckwith (2014); cf. Beckwith (2017).

22 Starostin’s (1989, p. 550) reconstruction of 相, which rhymes only once in the Shih ching, is erroneously printed as “smāhi”, which is actually his reconstruction of 相 OChi *smāh (S. 550), the word with which 相 rhymes. Both belong to his Middle Chinese rising tone category.

23 See above for two of the dicharacter transcriptions; for the r of the attested MChi bur (etc.) v. infra.

24 This word seems not to be attested in the earliest dated translations (Nattier 2008), which mostly translate dharma into Chinese as 法 ‘(law). However, dharma is evidently transcribed in quite a few of the early translations as 譚道 MChi dāo (lit., ‘the way’), in LOC *dāwā < OChi *dāwā ~ *dāwā ~ *dāwā, though the transcription was clearly chosen for its semantics as well. This transcription precedes the “formal” historical introduction of Normative Buddhism to China by several centuries (Beckwith 2015a, pp. 122–123). Cf. Ostertamp (2011, pp. 50–51) on OJpn *dārowa 真諦 ‘dharma’ and *kārowa 真諦 ‘karma’.

25 This epithet is highly irregular even within Indic.
Old Japanese *mimana: 任那 MSC rènnà < MChi *nim1 / *nim3 (P. 266–267) + *na3 (P. 221) < LOC *mī́mänà [ˈmiː.mɔ,na] ‘ancient realm on Korea’s south coast; Kara’.26

The phonotactics of the words Buddha, Dharma, and Sangha, which seem to have helped them maintain a disyllabic pronunciation in Prakrits as bud.dha ~ bu.dha, dham.ma, saŋ.gha, make it unlikely that the second vowel (perhaps phonetically schwa [ə] even in early times) would have been dropped in Antiquity. In fact, their final vowel has rarely been dropped in any language. It is only since completion of the post-Late Old Chinese monosyllabicisation that monocharacter transcriptions came to be read as monosyllables in Chinese.

It is unlikely that all Late Old Chinese disyllabic morphemes became monosyllabic at the same time. As is well known, sound shifts can take many generations to work their way through a language, and sometimes they stop before they are completed (Kiyose – Beckwith 2006).

It remains to say something about the hitherto mysterious coda r of Middle Chinese transcriptions in Old Tibetan, Khotanese, Uighur, and other segmental scripts. It is partly attested even in Japanese (v. supra), instead of the expected *t, which is only indirectly attested (and only in Japanese). The coda r seems clearly to be the result of intervocalic lenition of *d (i.e., words like *CVdV became CVrV) in northern dialects of Late Old Chinese. When apocope occurred, all morphemes of the shape CVrV became attested MChi CVr. Most Old Japanese forms, which have *CVtV, are evidently from a southern dialect that preserved *d and *t as *t.27 This detail further confirms the disyllabic nature of many Late Old Chinese morphemes.

Although this paper focuses only on apocope of final short *ã, it seems clear on the basis of the Japanese data that other vowels have been lost in Chinese too. However, because Middle Chinese loanwords with codas were made pronounceable in Old Japanese via regular paragoge addition of either i or u (or sometimes o/ō), disyllabic examples with such vowels in the second syllable are ambiguous historically, making it difficult to distinguish between genuine borrowings from Chinese with a disyllabic monocharacter reading and examples of a Middle Chinese monosyllabic borrowing plus a paragoge vowel. The present study therefore limits nígōgana readings to those which have a in the second syllable. There are a good number of them, as seen in Ōno (1977, pp. 420–423), including apa 鴨, kapa 甲, opa 邑; wata 26 For this name (and Kara, the usual name for the same region) and what is known about the language spoken by its inhabitants in Antiquity and the Early Middle Ages, see Beckwith (2007). Osterkamp (2011, p. 180) discusses Starostin’s reconstruction of the Old Chinese form of Han 韓 MChi *γan (P. 118) – as OChi *γbär, i.e., *yar, noting that OJpn kara ‘China, Korea’ seems to transcribe this word, q.v. Beckwith (2016b). The issue is complex, but this is clearly a case of OChi *Cara > MChi *Can, where OJpn preserves the OCh disyllabic *Cara. It must therefore be reconstructed *ara. Toponym examples of this kind are Sarashina 佐良志奈, 譲信 (Ost. 190) OJpn *tsaratsina (~ *saratisna) and Sarara 佐良々, 譲良 (Ōno 412; Ost. 190) OJpn *tsarara (~ *sarara) ← OChi *tsara > MChi *tsan3 (P. 391) > tsar 譲 MSC zăn.

27 It seems that OChi *t also became *d intervocalically, i.e., *t and *d merged in that environment before becoming *r. The alternate scenario of a traditional “monosyllabic” LOC *CVt becoming attested MChi CVr is phonologically unmotivated and typologically unlikely.

Acta Orient. Hung. 71, 2018
In conclusion, this study shows that at least some Old Chinese characters had disyllabic readings in Chinese itself – the extent of such disyllabic readings in Old Chinese being as yet undetermined – and that some of these readings were transmitted directly or indirectly to Japan by the Old Japanese period as etymological nigōgana, disyllabic “Chinese” readings of single characters in Japanese that preserve Late Old Chinese *ā.

Abbreviations and Sigla

An asterisk (*) marks reconstructions. Forms based on Chinese split-character ‘spellings’, rhyme-tables, or character transcriptions, and non-Chinese forms written in Chinese characters, are marked with a star (†). Notated tones are given as subscript numerals.

B.: Baxter (1992)
dial.: dialect
HSR: Historic Sinological Reconstruction
JDB: Omodaka et al. (1967)
K.: Karlgren (1972)
LMC: Late Middle Chinese
LOC: Late Old Chinese
MCh: Middle Chinese
MSC: Modern Standard Chinese (Mandarin)
MSJ: Modern Standard Japanese
OCh: Old Chinese
OJpn: Old Japanese
Ost.: Osterkamp (2011)
Ôno: Ôno (1977)
P.: Pulleyblank (1991)

References

Beckwith, Christopher I. (2007): Koguryo, the Language of Japan’s Continental Relatives: An Introduction to the Historical-comparative Study of the Japanese-Koguryoic Languages, with


