OUKA
Osaka University Knowledge Archive

Title	On the frontness opposition in loanwords in Old Uygur
Author(s)	Erdal, Marcel
Citation	内陸アジア言語の研究. 17 P.3-P.24
Issue Date	2002-09
Text Version	publisher
URL	http://hdl.handle.net/11094/18240
DOI	
Rights	

Osaka University Knowledge Archive : OUKA
http://ir.library.osaka-u.ac.jp/dspace/

Osaka University

On the frontness opposition in loanwords in Old Uygur

Marcel Erdal

In this paper we try to come a bit closer to a determination of how foreign words copied into Old Turkic may have been pronounced. After Professor Shōgaito laid solid foundations for this domain of research in 1978, it was subsequently dealt with by a number of scholars.[1] It also is one of the few matters about which scholars working on this language appear to disagree sharply, and transcriptions of texts show the greatest differences and fluctuations in this particular domain. It would therefore seem useful to attempt a clarification of some aspects of the matter at this stage, especially after most of the Uygur sources in Brāhmī have already been edited.

When lexemes or morphemes are copied from one language into another, the copies find themselves immersed in a new phonological system. To understand what happens during such borrowings we have to contrast the system – in our case mainly the vowel system – of the source language with that of the target language. Systematic frontness oppositions were foreign to all languages with which Uygur was in contact (with the exception – marginal for much of the corpus – of Mongolic). The Uygurs appear to have introduced them into copied lexical material to varying degrees, and in very complicated configurations.

Old Uygur borrowed a significant part of its lexicon from contact languages. Since most of our Uygur corpus consists of texts of Buddhist content, the great majority

1 See especially Shōgaito 1978, Moerloose 1980, Laut 1986:89-148, Röhrborn 1988, Johanson 1991:83-85, Erdal 1991:16 f., Röhrborn 1996.

of such elements ultimately came from Sanskrit, the culture language of pre- and non-Islamic South Asia. We have a very good knowledge of the vowel system of Sanskrit, which was also one of the sources for direct lexical copies in Old Turkic.[2] However, most of the Indic terms got into Uygur through Chinese, Tokharian or Iranian languages,[3] in which case it is their sound systems which become relevant for understanding the shapes which these terms acquire. What we know of the vowel systems of Sogdian and contemporary Chinese, which were among the most important direct source languages for Uygur, is quite sketchy, but the vowel systems of the two varieties of Tokharian are reasonably clear. Our primary question concerning Tokharian is the phonetic status of the character transcribed as *ä* and described as a *shwa* vowel; the rephonologisation of this vowel in words taken up by Uygur will have to be systematically dealt with. The Semitic alphabets, in which Uygur is mostly written, are often far from giving a full picture of the vowels. The central concern of the present paper, the distinction between /i/ and /ï/, is not at all addressed by those writing systems. Nor is it generally addressed by the Brāhmī system as used for Uygur (as detailed below), though Tibetan script, the other Indic writing system in which Uygur is written, does appear to make this particular distinction to a certain degree.[4]

I will here assume that phonemic oppositions did indeed exist in Uygur between

---

2 Orkhon Turkic already has Indic loans such as *išvara* or *maxarač*. Direct borrowing from Sanskrit came into fashion in the latest stage of Buddhist Uygur scripture.

3 The actual source for such terms in these languages is often not Sanskrit but one of the Prākrits or Middle Indo-Aryan languages; the 'direct contact languages' of Uygur were, moreover, also in contact with each other.

4 But then the use which Uygurs made of this script is especially inconsistent, beside the fact that its use for this language is so limited. The Brāhmī ms. edited as TT VIII I differs from the others in writing /ï/ (as well as /e/) as E whereas the letter E in other Brāhmī mss. represents the vowel /e/ both in Turkic and borrowed words. In this it shows the same practice as BuddhKat, which is in Tibetan writing. Unfortunately, this is a medical text which has nothing to offer concerning our topic.

/ï/ and /i/ on the one hand, and between /ï/ and /a/ on the other, though both of these oppositions might, under certain circumstances, get neutralised (as phonemic oppositions in all languages sometimes do).[5] Even when the writing does not give direct information on whether we should read *i* or *ï*, we should therefore still be looking for clues on which reading to prefer.

'Signal' letters

When writing Ottoman Turkish, certain Arabic characters functioned as 'signal' letters showing whether the syllables in which they appeared belonged to the front or the back series; I am here repeating a term used by Andreas Tietze in an unpublished lecture as quoted by Röhrborn 1996:176. The consonant pair used most consistently in Ottoman for this purpose was *qāf* as a signal for back syllables, *kāf* serving as a signal for front syllables. This 'signalling' is found already in the Uygur alphabet, which uses the two Semitic characters X and K for distinguishing between back-harmony and front-harmony strings in originally Turkic words. X[6] is, in the Uygur alphabet, used also for representing /g/ in back-harmony strings[7] but no distinction is made between /k/ and /g/ in Turkic front-harmony words.

Dealing with the Uygur correspondences of Sanskrit *k*, *g*, *h*, *kh* and *gh*, Shōgaito 1978: 96 stated that X of the Uygur script corresponds only to Sanskrit *h* while the other four Sanskrit consonants together receive a different representation. Röhrborn 1988 then

---

5 I will deal with the question whether Old Turkic distinguished between *i* and *ï*, and what the nature of this distinction may have been in its different sub corpuses, in a different paper I am preparing at present.

6 Scholars have generally used Q for referring to this character but I will refer to it as X because it goes back to Semitic *cheth* (and *gimel*, also pronounced as fricative in Sogdian, from where the script got into Uygur).

7 In back-harmony words the distinction between /k/ and /g/ is shown – though not quite consistently – by double-dotting the X when it represents /k/. The sound χ often gets a single-dot diacritic.

pointed out that the spelling system applied to stems of Indian origin is different from that used for Turkic elements but identical to the system applied to words of Indian origin when writing Sogdian (from where it was no-doubt copied): Whether the matrix language was Sogdian or Uygur, words having the consonants *k*, *kh*, *g* or *gh* in Sanskrit or Prākrit were in the receiving language spelled with K, whether the adjacent (originally) Indic vowels were *a*, *o*, *u* (classified as back in Turkic) or *i*, *e* (assigned to the Turkic front series).[8] If an element in a word goes back to Indic *h* and is followed by *i* or *e*, it is also spelled as K in Uygur writing. The letter X is in borrowed Indic elements used only for Indic *h* followed by the other vowels, *a*, *o* or *u*.[9] Stops and fricatives are distinguished, then, but not voiced velars from unvoiced ones. All this would mean that, as far as elements copied from Indic in Uygur are concerned, X does have 'signal' value (since it turns up only if followed by *a*, *o* or *u*) but K does not (as it can be followed by any vowel).[10] Sanskrit *h* is not all that common, of course, so that – all in all – the 'signal' value of velar letters would in fact be quite limited in such borrowings (where their

8 A small selection of such instances can be found in Röhrborn 1988:236. Counter-examples, X used in stems of Indic origin to represent [k], are quite few. All three instances with X mentioned in Röhrborn 1996:184 are quoted from the St. Petersburg Suv ms.. In *lakšan* 'body mark', one among these, the letter probably represents [x] and not [k]: There is solid evidence that the string /kš/ was pronounced as [xš] in Uygur words and this pronunciation appears to have gotten extended to a borrowing in this case. Another instance appears in Buyan Ävirmäk, a passage added to the text of this ms. in the 17th century, which should by no means be taken as evidence for Old Uygur although it manages to ape that language to a large extent. The real Suv text can only be established by a critical edition taking all the numerous older mss. (mostly in Berlin) into consideration and would exclude Buyan Ävirmäk. The phenomenon is therefore marginal at best.

9 A few examples are mentioned in Röhrborn 1988:240. In Sogdian the letter is also exclusively used for the representation of fricatives; *ḥeth* is a pharyngeal fricative which in Semitic languages attracts vowels to the back of the oral cavity. Cf., on the other hand, *tamŋaq* < Tib. *gdams-ŋag* 'instruction' with X at the end of a foreign word.

10 The presence of X as /h/ is triggered only if the vowel following it is /a o u/; the triggering vowel is the subsequent one, as with the allophones of Modern Greek /x/; not the one preceding the fricative, as with German [x ~ ç].

evidence would have been most needed for establishing the pronunciation). This does not apply to suffixes added to these lexemes: There the signalling system shows the normal Turkic vowel + velar dichotomy, with *+qa* vs. *+kä*, *+taqï* vs. *+täki*, *+lïq* or *+luq* vs. *+lik* or *+lük*. Dealing with the spelling of words of Indo-Aryan origin in Uygur, Röhrborn 1996:177 writes: „Was die Konsonanten betrifft, so kann nur das Vorkommen von sogenannten ‚Signalzeichen' als sicherer Hinweis auf Velarität (Signalzeichen Q^{11}) bzw. Palatalität (Signalzeichen K) einer Silbe gelten." However, contrary to this statement, these letters have no 'signal' status as far as stems of such origin are concerned: When using Uygur script for these words, the use of K is not automatically connected with the harmony class of a syllable within which it is used. There is therefore nothing irregular or non-harmonic about the spelling ŠLWKWX (found e.g. in HTs VIII 1924) for 'stanza (accusative)' with a lexeme ultimately coming from Skt. *śloka*: The K only shows that the velar is a stop and not a fricative. We know that, in Turkish, consonants such as /k/, /l/ or even /t/ have back and front variants and that their pronunciation alone can, if they appear at the end of a stem, determine the harmony class of suffixes added to it; thus in Turkish *saat+e* and *hal+de*.[12] Reading Ottoman Turkish one could, if a syllable contained the phonemes /k/ and partly /s/ and /t/, guess by the use of Arabic letters whether the syllable would be pronounced with front or back vowels. This phenomenon did not *a priori* hold in Uygur. Our information on the frontness or backness of Uygur syllables therefore mostly follows from the way the vowels are spelled in the stems and from the shapes of the suffixes. The Semitic writing systems are rather unhelpful concerning vowels: [ä] looks like [a] except in the absolute onset and [i] and [ï] are spelled identically (except when [ï] is spelled as A); [ö] and [ü] are generally spelled with the

11 What is meant is the character I have referred to as X.

12 We needn't here discuss the question whether [t'] and [l'], these front variants, should be considered to be phonemes although the matter arises only in connection with borrowed words, or whether a supra-segmental model is better suited for the topic.

digraph WY in the first syllable but there are numerous exceptions (such as *tözün* ‚noble' being spelled TWZWN).

Concerning the last syllable of a stem we therefore have to rely on suffixes:[13] When we find the accusative of the word for 'stanza' coming from Skt. *śloka* as ŠLWKWK with K and not X we are entitled to assume that it was, beside *šlok*, also pronounced as *šlök*; *šlok+lug* and *šlok+larïg* (both ending in X) are also quite common, but so is *šlök+lüg* (with twice K). The word was probably borrowed through Tokharian, where one of the two variants of the A dialect is spelled as *ṣlyok*. Whether the vowel of this dialect was itself actually pronounced as [ö] (*yo* may, after all, have been the Tokharian spelling for [ö]), whether it sounded closest to [ö] to Turkic ears, whether it was rephonologised as /ö/ because of the sequence /šl/ (broken apart by the Turks, as shown in spellings such as ŠWLWK) – there is no reason to doubt the /ö/ in this particular variant of the Uygur word. Since Sanskrit only had one low rounded vowel (*o*) whereas Uygur had two (*o* and *ö*), there is always the possibility of a mapping o → ö also if Sanskrit were the direct source. There are many cases other than the fluctuating *š(u)lok* ~ *š(ü)lök* where lexemes originally coming from Sanskrit get front vowels.

The phonic influence of foreign *k*

When looking at all the cases where we find the mappings a → ä, u → ü or o → ö, it turns out that the great majority take place in the vicinity of K (as stated by Röhrborn 1988). In a number of instances, some of them mentioned in Röhrborn 1988:237, words in Sanskrit having *ku°*, *kū°*, *ko°*, *go°* etc. in the onset are spelled as KWY° in Uygur, no doubt to be pronounced as *kü°* or *kö°*; e.g. *Göd(a)vari* (BT VII 42) < *Godāvarī* or *kümut* < Skt. *kumuda* 'lotus'. Some such lexemes were subsequently borrowed by Mongolian with a front vowel, e.g. *kümüd* just mentioned, or *Gödäm* (in Uygur attested e.g. in BT

13 With loans, this statement holds only if the suffixes have front harmony, according to the rule (restated below) that back harmony is the default class for suffixes added to such lexemes.

VIII) < *Gautama*. A versified *avadāna* text edited by Professor Shōgaito has KWŠ'L'MWL (from Skt. *kuśalamūla*) alliterating with *küčlüg* and therefore to be pronounced as *küšalamul*, if not with additional fronting elsewhere in the word. The Brāhmī spelling *Gäŋ* shows that the Uygurs pronounced (or *sometimes* pronounced) the vowel in the name of the river Ganges as /ä/. Röhrborn 1996:182-3 tries to explain this as follows: „Wenn ... ein Schreiber nicht mit der Norm des Subsystems vertraut ist, dann wird er ein K-Zeichen des Subsystems als palatales Signalzeichen interpretieren. ... Verläuft die Anpassung an die türkische Norm in Richtung Palatalisierung, dann scheint mehr der optische Eindruck des K-Zeichens auf den Schreiber verantwortlich zu sein." Röhrborn appears to have been thinking solely of copyists of holy texts intent on (but not quite capable of) retaining traditional spelling. Instead of assuming unprofessionality one might rather speak of the gradual and sporadical influence of spelling pronunciation. Had the influence been a merely visual one, *g(ä)rx+kä* 'at (the ascendance of) the planet' in ZiemeVorr[14] dated to the year 1022 would not have had a front suffix: The fronting in this word must have been caused by the *g*°.

Moreover, words not transmitted over the copying of religious texts but through general diffusion show the same fronting, e.g. *käpäz* 'cotton seed' coming from Indo-Aryan via Tokharian (*karpāsa* in B but *kappās* in A) with the original /a/ vowels realised as /ä/ in Turkic or *kürküm* 'spice used as saffron surrogate', a term found everywhere between India and the Mediterranean with two /u/s. Tokharian *kuñcit* 'sesame', a third term from the same non-religious semantic domain, is also in Uygur rephonologised as /ü/, as shown by the Brāhmī spelling as *künčid* in TT VIII M 28. If *$k_u$rekār* of Tokharian A (*kwrakār* in B) gives *kürekar* 'ceremonial tent' (spelled with KWY° several times) in

14 T II Y 37b r4 as read by A.v.Le Coq; the actual fragment is now lost. See the UW for this and other abbreviated references to texts. *g(ä)r(ä)x* would also be a possible reading. TokhAB has *grak ~ grah* coming from Skt. *graha* but the defective spelling makes Sogdian a likely direct source.

Uygur, this is a matter between these two languages which has nothing to do with the fact that the ultimate source is Skt. *kūṭāgara* 'chapel'.

Loans starting with /k/ which arrived from or by way of Sogdian appear not to have participated in the fronting process: Such terms are *kolti* 'a very great number' << Skt. *koṭi* (cf. Laut 1986:91, 105), *kalp* 'an age in divine reckoning', *kšan* 'a very short moment', *karmaput* 'precept' or karte << Skt. *grhastha* 'householder'. If this is correct, it would link fronting by onset /k/ to Tokharian mediation.

There appears to be a general correlation between /k/ in the coda of borrowed nouns and harmony fluctuation in suffixes added onto them: The word spelled as Č'TYK or Č'DYK ultimately comes from Skt. *jātaka* 'story about the life of Buddha in a previous existence'. It has, in Suv, three examples with the suffixes *+ig* and *+lig* (one of the latter in Suv 235,10) as against two examples with *+lïg* (one of the latter in Suv 235,2); *čadïklïg* appears also e.g. in TT X 4. The Tokharian AB cognate *jātak* is not very likely to be its direct source because of the stem's second vowel. As with Skt. *śrāvaka* appearing in Uygur as ŠR'VYK, Sogdian mediation may here be likelier (cf. Laut 1986:117): The heightening of vowels does not appear to be typical of the Tokharian route. *užik / üžik* 'letter, akṣara', a Sogdian term (spelled with both 'WY and 'W in the first syllable) perhaps ultimately coming from Chinese, shows both +YK and +YX as accusative suffix.[15] A common variant is, in Semitic script, spelled as *užak*, which also must have had *užïk* as underlying form. The word spelled as *ābhišik* in TT VIII D 17 (from Skt. *abhiṣeka* through TokhAB *abhiṣek*) shows a similar suffix fluctuation between the suffixes *+lïg* and *+ig* (accusative). With other borrowed °*k* lexemes I have come only across instances with back suffixation: The form *saŋgik+qa* is attested in SA

15 *čäk* 'dot, diacritic' coming from Tibetan *cheg* is attested with the suffix *+kä*; it is therefore another instance of a foreign word not showing back suffixation because of final *k*. Like *užik* (with which it conceivably shares its etymology) it is not of Indic origin and was therefore not considered by Röhrborn.

8; it comes from Skt. *saṃghika*, an adjective derivate from *saṃgha*, and according to Laut 1986:145 has Sogdian *snk'yk* as direct source: Its second syllable could be /ï/ as well as /i/. Another relevant term of ultimate Iranian origin is *psak* 'crown', always spelled with K and receiving the suffix *+lïg* in U II 59,2_2; its direct source may be Tokharian *pässak*, since the Sogdian counterpart has an initial *alef*. The question whether the spelling of the velar has any automatic relevance for the pronunciation of the suffix is, I think, resolved by runiform $n^1g^1ws^1k^2l^1r^1$ = *nagošaklar* 'lay believers' in ms. TM 332 (KöktüTurf p.1047) : The word is spelled with a front k^2 but the plural suffix is *+lar* and not *+lär*.[16] This shows that the presence of a front velar did not, by itself, have any influence on the quality of the suffix vowel. We have a similar phenomenon also in Manichæan script, *t(a)razuk* 'scales for weighing' spelled with undotted *caph* in M II 12,8; in U II 77,25 we find it spelled with K in Uygur script as well. This word of Iranian origin has back vowels in all Iranian and Turkic sources.[17]

Loans with final /k/ spelled as K have fluctuating suffixation which is not indicative of the vowel class of the last syllable in the stem. No certain choice between readings such as *abišik*, *abišïk* or *abïšïk* can be proposed for any particular instance if there is no additional evidence. We have already shown that the presence of K is no reason for discarding the latter two readings for this word. The K appears to have saved the front pronunciation for some speakers, against the more general rule that the back harmony is the default class for suffixes added to loans. The vowels of the last stem

16 Zieme 2001:217 took Mz 386 (TM 333) v1-2 to contain another runiform instance of this term, reading the passage as $š^1{}'k^2l^2r^2{:}t^1w[g^1{}']r^1{:}r^2t^2I$. This he interpreted as *[nigo]šaklar tu[g]lar ärti* and translated it as '[audi]tors were born'. This would mean that *nigošak* here gets the front variant of the plural suffix. However, the first character after the lacuna does not, I think, look like s^1 but rather like k^1. Furthermore, the verb phrase *tugar ärti* would imply durative aspect or a continuous or iterative event, which seems unlikely; besides, babies are not born as auditors. Another possible reading is *]ka k(ä)l(i)r tu[ša]r (ä)rti* 'They were coming to meet (+ dative)'.

17 Clauson is wrong in writing it with °*g*; the *caph* of Manichæan script can only be /k/.

syllables can however, in some cases, be determined by other means: Variation with *a* (as in *užak*) would, e.g., speak against the reading '*užik*' and for *užïk*. I would also plead for *čadïk* beside *čadik* (not excluding *čädik*) because vowel raising (apparently linked to Sogdian mediation) does not necessarily imply fronting and because we know from cases like *psak* that K does not front vowels before it.

Shōgaito 1978 had limited his scope to words of Indic origin because his aim was to answer the question of their route into Uygur. When dealing with the spelling of loan words of Indic origin as against the spelling of original Turkic stems and suffixes, Röhrborn 1988 and 1996 disregarded the spelling of borrowings coming from other languages into Uygur, as the ones just quoted. Röhrborn 1988:232-233 speaks of the whole Sanskrit lexicon being open to the learned Uygur usage, which is used even when the source language of a text being translated is Chinese. The fact is, however, that such Sanskrit terms do not generally turn up in their original shape but in the shape they have in Sogdian or, more commonly, Tokharian. While learned translators like Šiŋko Šali tutuŋ and later Uygur scholars would have a direct knowledge of Sanskrit, the normal source of terminology would be indirect, not to speak of non-terminological loans. We could therefore speak of two lexical permeabilities as far as Buddhist terminology of the post-Sogdian stage is concerned: The Indic - Tokharian permeability and the Tokharian - Uygur permeability. Dealing with the way Uygur represents loan words in writing, we should be looking at the phonic and graphic mechanisms of borrowing as a whole, as it is unlikely that a language should, at any particular stage, have dealt with a specific ultimate (i.e. indirect) source in any way not related to the rest of the sources. We here see that Manichæan texts in two writing systems not dealt with by the scholars mentioned show the same phenomenon as the ones they did deal with. Suffix fronting by borrowed K is clearly not limited to Tokharian mediation.

A third possibility with terms borrowed with K is, of course, that the K is not in the beginning or the end but somewhere in the middle of the stem. The base of a form

like *šaki+lïg* (ShōAgon 203 u. 22; < Skt. *śākya* by way of TokhAB *śākki*) could also, of course, have ended in /ï/, influenced by its first vowel; this in spite of the spelling with front K, as, in fact, *šaki* itself could (e.g. in TT X 66, 136, 334) have been pronounced as *šakï* in spite of the K. *lakšan* < TokhAB *lakṣaṃ* is also generally spelled with K.

säŋräm < Skt. *saṃghārāma* 'monastery' also has /a/s in Sanskrit but front vowels in Uygur.[18] Röhrborn 1996:183 proposes to account for this by assuming that SNKR'M (the spelling without *alef* after the S, which was in use in Sogdian and is attested once in an Uygur source in Sogdian script) was misunderstood as S'KR'M, *alef* and *nūn* being similar in Uygur writing; the K is then supposed to have been interpreted as a 'signal' character for fronting, while retaining [ŋ] phonically. All this does not seem very likely as it presupposes a simultaneous interpretation of a character string as both NK and 'K and the absence of oral transmission. The term appears in Tokharian AB as *saṅkrām* or *saṅgrām*, with a velar nasal followed by a velar stop. If Sogdian was the immediate source, the process remains obscure but can be an inner-Sogdian problem; if it was Tokharian, the velar stop with which the word appears in the Tokharian dialects could have been responsible for the fronting.

Fronting without velars

The only solidly attested example for the fronting of a stem which has no K (and is therefore not mentioned in Röhrborn's papers) which I am aware of is *ärdini* 'jewel', whose ultimate source is Sanskrit *ratna*. Its first vowel is shown to be /ä/ by Brāhmī evidence; the front suffix variants *+lig*, *+kä* and *+g*, with which it is frequently attested, prove its last syllable to have been /i/ and not /ï/. I have come across a single instance of +X' in BT VIII A122. The UW entry for this lexeme allows us to trace its history within Uygur: Manichæan texts (including the relatively late Pothi Book) never have the initial vowel, which is clearly a Turkic addition; they show the spellings RTNY, RTNYY,

18 '*saŋram*' in UW 277b is an error.

RDNY and RDDNY. Both Sogdian and Parthian are relevant for the shape of this lexeme: Sogdian retained the Sanskrit *t* while Parthian shows *d* instead and may possibly have influenced the form. RDNY is further attested in ms. L in TT VI 243 (as against six other mss. extant for the passage) as only non-Manichæan evidence for such spelling.[19] *ärdni* is the second stage the form went through, the later stages being *ärdini*, occasional *ärtini* (found in the DLT as a variant) and (rarely) *ärdäni* (perhaps borrowed back from Mongolian). The Manichæan variant was presumably pronounced with a vowel after the /r/ but we do not know what that vowel was; if *ärdni* came about by metathesis, it was /ä/ (and not /a/, as assumed in Laut 1986:144). Here we clearly have fronting with Iranian intermediary. We do not know with what vowels this intermediary was pronounced; the vowels of borrowings which reached Uygur by that way need not concern us as the fronting could have occurred within those languages. We have to state, at any rate, that not all words of ultimate Indic origin have back suffixes in Uygur, even if they do not end in /k/.

The fate of Sanskrit *°a* in Uygur

Sanskrit *°a* stems referring to animate beings generally appear in Tokharian with *°e* while those referring to inanimate entities drop the vowel. As discovered by Shōgaito 1978 (and independently from this restated by Moerloose 1980), this distribution is generally mirrored by Uygur, showing that Uygur borrowed much of its Buddhist lexical stock from Tokharian. In Semitic writing systems, with which the vast majority of Uygur sources are written, the coda vowel is in these borrowings spelled as Y. Y is in Uygur used for transcribing three sounds, [i], [ï] and also [e]. From sources in Indic writing we know that /e/ was in Uygur admissible also in non-first syllables, both in

19 Manichæan influence on the Säkiz Yükmäk Yarok has been variously discussed in the recent years; this evidence of one ms. against all others *might* speak for a specific Manichæan background of this particular scribe.

borrowings and in original Turkic words. It might therefore be assumed that the sound appearing as Y was, in the coda of such words, pronounced as [e], following the source language. This is no doubt the reason why the UW transcribes such words with °*e*, both when the author can adduce evidence that such a Tokharian word exists and when no such evidence is adduced: The words attested as ''ČYTY, as ''RČWNY or 'RČWNY and as ''YWKŠ'NY are respectively transcribed as *ačite* (with Tokharian cognate), as *arčune* and as *ayukšane* (no Tokharian cognate mentioned). Such etymologically motivated transcriptions are not, as far as I could determine, supported by Brāhmī sources. There is much evidence against reading Uygur Y as *e* every time Tokharian has *e*, a part of it mentioned in the next section: The lexeme written as *asure* in Tokharian and coming from Skt. *asura* is in Neujahr 73 (a Brāhmī text edited in Zieme 1984) spelled as *asuri*. The UW quotes this Uygur instance as *asure* and writes: „Trotz des Brāhmī-Belegs transkribieren wir *asure*". Brāhmī evidence deserving to be taken seriously, there seems to be no reason for not reading the word as *asurï*. In similar disregard for the data, Laut 1986:140 writes *tirte* (<< Skt. *tīrtika* 'teacher of false doctrine') because of TokhB *tīrthe*, in spite of *tirthi* in TT VIII O 3; another Brāhmī instance, *tirti* in MaueKat Nr. 44a v8, appeared subsequently. Other examples with *i* / *ï* << Skt. *a* in the coda are *čit]t[a]čaytasiki* < Skt. °*caitasika* in MaueKat Nr. 44a v3, *nigranti* < Skt. *nirgrantha* ibid. v8 and further terms in that line. In Neujahr 75-78 we find the proper names *Māndhāta*, *Maṇibhadra*, *Vallabha* and *Vaiśravaṇa* all in Uygur versions with *i* and not *e* in the end. It is true that we find *asanke* 'a very large number; a very long period of time' in MaueKat 19 Nr. 11 from TokhA *asaṃkhe* and *šarmire* 'novice' in MaueKat 3 Nrs.78 and 84 from TokhB *ṣarmire*. These two words are not, however, instances of Skt. °*a* > Tokh. °*e* at all, as they come from Skt. *asaṃkhyeya* (and not '*asaṃkha*') and *śrāmaṇera* respectively.

Back-vocalic suffixation as default in loans

It was Peter Zieme who first (1969: 57-8) pointed out that foreign lexemes appearing in Uygur get back vowel suffixes even if the stems have front vowels; Gabain 1950:58 had stated that there are irregularities in this domain but had not dealt with them, the only one she mentions being that *nïrwan* (as she spells the term) gets *+qa* whereas *frnibran* (<< *parinirvāṇa*) is attested with *+kä* as dative suffix. Zieme discussed a number of examples in Manichæan sources, mentioning in a note that the phenomenon is present also in Buddhist texts. The matter is then elaborated upon by Röhrborn 1988 and 1996, who limits himself to Indic terms in Buddhist sources.[20] The fact is best observed in texts written with Brāhmī writing when the stem ends in [e], which is spelled as *e*; e.g. in *asanke+larta* (MaueKat Nr. 19,11), *šarmire+larï* and *šarmire+larka* (Nrs. 3,78 and 84). The stem of *tetse+larïnïŋ* ‚of his pupils' (ĀṭSū 1 v4, also Brāhmī) shows the same phenomenon but comes from Chinese. In BuddhKat 23 (Tibetan writing) *didim+lïg* 'wearing a diadem' seems to be the preferable reading against the editors' '*dïdïmlïg*', but *didïmlïg* cannot be excluded. The form is attested with *+lïg* also in Maitr 2v23, BT VII and elsewhere. There is no known reason for *frišti* 'angel', of Iranian origin, to have had back vowels; even this lexeme with two /i/s shows fluctuations in suffix harmony: *frišti+lar+ka* (M II 10,4, TT IX 94 and elsewhere) vs. *frišti+lär+kä* (M III Nr.1 I v3 and elsewhere). Since we know from Brāhmī evidence that the last syllable of Uygur *asanke*

20 Röhrborn 1996:178 refers to a single exception, *hetu+nüŋ* in MaueKat Nr. 44a r2, which he considers to be an error. Another exception, in a Manichæan text and with an Iranian word, is *možag+kä* with K in the suffix, referred to by Zieme 1969:57. *rajagṛ+dä* in MaueKat Nr. 3,86 is considered an error as well, as the editor transcribes the base as *rajagïr*; Röhrborn 1996:178 does not agree, thinking the frontness of the suffix is caused by the presence of *g*. Since we must assume the presence of an auxiliary vowel after the *g* (which Röhrborn does not do), rules for such vowels (discussed below) should apply. We should not, on the other hand, limit our view to borrowings: Since we find *ugur+dä* in Nr. 3,90 as well as 96 (in the following line in the ms. and 4 lines further on), the scribe appears to have been going though a singular stretch of confusion.

is /e/, *asanke+lïg* (documented in the UW; consistently spelled with X in the suffix) and *asanke+daqï* (DKPAMPeters 263, 403) can also serve as examples for the phenomenon. Similarly with *den+ka* 'to the religion' in TT II A 46, in view of the first syllable of what should be read as *dendar* in Tibetan script in BuddhKat 26, 29 and 30. Without the information on the base vowel from an instance in an Indic script, we could also have read it as '*dïn*' (as Zieme 1969:37 in fact does).

The rule that borrowings should generally be followed by suffixes of back harmony has a quite wide diffusion; in Qarakhanid it holds for Islamic loans: The QB and the Atabatu 'l-Ḥaqā'iq spell stems with consonants indicating front vowels but have their suffixes with back-harmony indicating consonants; cf. Mansuroğlu 1959:90. The rule still stands – strangely enough – in modern Tatar with Russian loans (though not with earlier loans from Persian and Arabic), as described in Öner 1998:22-23: e.g. *appetitlïlïk* 'appetite' *dokumentlaštïrïlgan* 'documented', *protsesïgïzda* 'to your process', *printsiplarï* 'their principles' etc.; there are no irregularities among back-vowel stems. Öner points out that the matter cannot be motivated by recent spelling habits, as it is found also among the Tatar community in Finland, which emigrated during the revolution and uses Tatar mainly as a spoken medium. The sources of the Tatar phenomenon are to be found in Classical Chagatay.

We transcribed the *i* of the possessive and genitive suffixes of *tetselarïnïŋ* with *ï* because Brāhmī does not show the difference between these two sounds. If the last syllable of a borrowed stem in a Brāhmī text has *i*, that can also represent [ï] as well as [i] (but not [e]): The word spelled as *pravarikda* in ZweiFrag 5 can be read as *pravarïk+da* as well, since the same fragment also writes /ï/ as *i* in *yavïz*, *monï* (accusative), *atlïg* and many other words. We might thus possibly read *rašï+ta* instead of *raši+ta* in TT VIII L32, from Skt. *rāśi* 'zodiac sign'; *jñātiputri+lïglar* in TT VIII G13 might also have *ï*. For the accusative form spelled *utpatti+sin* 'its emergence' in Brāhmī TT VIII H4 the reading *utpatïsïn* is possible as well. The word in the UW and in Laut 1986:144 spelled as

arži, coming from Skt. *ṛṣi* presumably via some Prākrit and then via Sogdian *rž'y*, was probably pronounced as *aržï* in Uygur: *ṛṣi+larda* in TT VIII D6 (with historical spelling of the stem, if read correctly) and *+larïg*, *+ka*[21] etc. in Uygur script may be no proof, but *arže* in BuddhKat 26 and 29 makes this likely; see Maue & Röhrborn 1984:294 and footn. 42 (also referring to the Tibetan itinerary) for the use of *e* to represent /ï/ in that source. The coda vowel of this word is unlikely to have been *e* as it has *i* in the source languages. *nizvani+lïg* 'characterised by moral blemish' in Brāhmī (TT VIII E47) and instances of the addition of *+larïg*, *+larka*, *+lïg* and *+ka* to this base in Uygur writing are also given as proof that foreign words with front codas are followed by back-class suffixes. However, the spelling *nizvane* in BuddhKat 9, 27, 37 and 43 points in the direction of *nizvanï*; the reading proposed by Maue & Röhrborn 1984 is in fact *nïzvanï*.[22]

Then we have a group of Uygur °*i* stems corresponding to Tokharian °*e* stems: *kumbandi+lar* in ĀṭSū 1 v6, coming from Skt. *kumbhāṇḍa* through TokhB *kumbhāṇḍe*, can safely be spelled with *ï* in the last syllable of the stem; note that *tetse+larïnïŋ* appears in the same ms., so that the scribe did have °*e* in his inventory. *gandarvi+lar* from Skt. *gandharva* over TokhB *gandharve* is quite unlikely to be an error, as it appears twice (ĀṭSū 3 r2 and v1). Nor is *nigranti+lar* (TT VIII G 13) < TokhB *nigranthe* < Skt. *nirgrantha* 'anchorite', supported by *nigranti* in MaueKat Nr. 44a v8: Both probably have *ï*. *ništani+larï* (MaueKat Nr. 44b B2) presumably belongs to the same group if it ultimately comes from Skt. *niṣṭhāna* 'food', as the editor thinks.

abišik / *abišïk* mentioned above is another word of Indic → Tokharian origin with heightened foreign *e*, attested with the back suffix *+lïg* on the one hand, with the accusative suffix *+ig* (BT VIII 355 and 361) on the other. Why was the /e/ in *kumbande*,

---

21 Röhrborn 1988:240 gives *arži+ka* and *šaki+lïg* as proof that Indic loan words get back-class suffixes even when they end in front vowels; erroneously, to my mind.

22 Sogdian, the source language, has *nyzβ'n'k* / *nyzβ'ny(y)*, probably to be transcribed as nizβānē.

asure or *abišek* not retained[23] although *asanke* (< Tokh. *asaṃkhe*) does have it? Non-first-syllable /e/s may have been gradually eliminated from Old Turkic: *yete* '7', *yegirme* '20' and *süvre* 'pointed' may have originally have had /e/ and lost it in the course of time. So this was an inner-Uygur process.

I am not aware of a single instance of Tokharian °*e* < Skt. °*a* appearing as °*e* in Uygur. If this is really so, then transcriptions such as *arčune* should be corrected.

A possible retracting influence of the accusative suffix

One factor which may have changed coda /i/ (or even /e/ as well) into /ï/ could have been the accusative form added to it in the back-harmony variant. Take *nizvanï+g* (attested in Suv and in BT VII H 20) and *ačarï+g* (attested in Suv), which reached Uygur from Sogdian (the ultimate source of the latter being Sanskrit), *išvarï+g* (TA 4272), *kušatrï+g* (Suv), *p(ï)raptï+g* (six times in TA) or *darnï+g* (Suv 484,17), all spelled with X in the suffix and here – tentatively – transcribed with /ï/ at the end of the stem. The fact that the coda vowel gets into the *same syllable* as a back-consonant is highly likely to have retracted it to back position. This *may* then have influenced the whole paradigm. I am aware of no good argument against reading *darni* as *darnï* and taking *darnï+lïg*, *darnï+larïg* etc. to be instances of standard vowel harmony.

Auxiliary vowels

Consonant clusters of foreign words are in Uygur, as in other Turkic languages, often broken up by auxiliary vowels; they usually involve an *r*, whether a syllabic one between consonants or a consonantal one beside a vowel. There is no reason why such auxiliary vowels should be taken to be /i/, as written by some scholars, if adjacent vowels belong to the back series. The UW is inconsistent in this matter: The author correctly

23 The UW gives *abišek* as lemma; this is unacceptable as it is against Brāhmī documentation.

writes *amïrta* << Skt. *amṛta*, *anasïravadatu* << °*āsrava*° and *anapan(a)sïmartilïg* << °*smṛti*°, but *abidarim* of UW (< TokhAB *abhidharm*) should be *abidarïm*, especially since a Maitr variant is spelled *abidaram*. *abipiray* (UW; < Tokh AB *abhiprāy*; Uygur Brāhmī *ābhiprāy*) might better be *abipïray*: The preceding vowel is *i* but *pray* > *pïray* would seem to concern a development within the domain of the original second syllable. *an(a)ntiriš* in Laut 1986:124 footn. 4 should be *anantïriš*: Other Uygur variants of this word (< TokhA *ānändārś*, B *anantārś* < Skt. *ānantarya*) mentioned in the UW are *anantraš*, *anantïraš*, *anantaraš* and *anantrïš*. The UW's *apiramane* << Skt. *apramāṇa* 'incommensurate' should be *apïramanï* while *ardir* << Skt. *ārdrā* (name of a moon station) should be *ardïr*. The UW entry for *arži* (thus!), ultimately < Skt. *ṛṣi*, quotes instances where double or single *alef* of the first syllable are replaced by 'Y from Shō, ShōKenkyū, BT VIII and Suv; the author thinks they should be read as '*erži*' (because two of them alternate with single *alef*) or '*irži*'. But *ä* is by no means the same as *e* and Y cannot be used to represent it, a spelling with a single *alef* cannot have been used for representing [e], nor is there any historical justification for [e]. /ï/ does, on the other hand, alternate with /a/ as auxiliary vowel in many of the instances quoted here. Laut 1986:97 writes *čakir* for *čakïr* (< Tokharian AB *cākkär* and Sogd. *ckkr* < Skt. *cakra*); the /ï/ is supported by the variants *čak(a)r* and *č(a)kar* and by what we find or do not find in the second syllable of the source languages.[24] Similar considerations should make us write *patïr* and not *patir* (< TokhA pātr), *šastïr* instead of *šastir* (<< *śāstra*), *vačïr* and not *vačir* (< TokhAB *vājär*)[25] and so forth. Occasionally auxiliary vowels appear beside consonants other than /r/; here we just mention one of the examples, *ašïvači* < (TokhA *aśvaji*), which is in Laut 1986:126 quoted as *ašivač[i]*.

---

24 Laut is against Moerloose's (1980:69) derivation of the word from Tokharian, presumably because of the Maitrisimit's defective spelling as *čkr*.

25 The Uygur variant *vašir* may, on the other hand, reflect the TokhA variant *waśir*.

The quality of raised vowels

Laut 1986:96 takes the word for 'sandalwood', originally Skt. *candana*, to have been borrowed into Uygur by way of Sogdian (*cntn*) and not Tokharian (A *candaṃ*, B *candāṃ / cantāṃ*) because of its raised first vowel.[26] What I here object to is his spelling of the Uygur lexeme as '*čintan*' and not *čïntan*: Raised vowels should not automatically be assumed to be fronted. In this case we find the word spelled with n^{l} in the (runiform) Irq Bitig, which (since this source does not belong to Orkhon Turkic but to Uygur) should be sufficient evidence. Another case in point is the spelling '*šazin*' in Laut 1986:138. It alternates with *šazan* and comes from TochAB *śāsaṃ*, both reasons for reading this variant as *šazïn*. The editor of the TA has correctly (as in so many other cases) transcribed this word as *šazïn*.

Conclusion

I have, in this paper, tried to point at clues - some of them obvious but many too often disregarded - concerning the phonic shape of words which Uygur adapted from other languages; I hope to have pointed at some of the problems involved. At present we understand only a part of the forces which were at play in this matter. What is needed in order to get a better grasp of what was really going on is a systematic survey of all the available material. The foremost principle to be applied should, I think, be the trust to be placed in explicit evidence: in this case mostly from Brāhmī and Tibetan script sources. The second principle I would propound is that Uygur is a 'real' language with speakers, possessing an internal logic, and not merely a medium for the transmission of scriptures whose character is determined by scribes. There were, of course, linguistic differences between the dialects of various communities (not all of them sufficiently determined as

26 The term 'Sproßvokal', i.e. auxiliary ('anaptyctic') vowel, used by the author in this and similar cases is inexact as the word in the source languages can be expected to have had the same number of syllables as in Uygur (i.e. in this case two).

yet), but it should be permissible to draw conclusions from one text group towards the other unless there are good reasons to the contrary. The final point is that there obviously was individual variation between users of the language, linked, among other things, to the degree of erudition. In that sense – the writing systems not being quite adequate from the phonetic point of view – there will always be a degree of indeterminacy. It is our task to bring that degree down to a minimum. This task I have, in this paper dedicated to a foremost contributor in this very domain, been able to accomplish only to a very limited extent.

Bibliography

Erdal, M. 1991: *Old Turkic Word Formation.* Wiesbaden: Harrassowitz (Turcologica 9). 2 vols.

Gabain, A.v. 1950: *Alttürkische Grammatik.* 2. verbesserte Auflage. Leipzig: Harrassowitz.

Johanson, L. 1991: 'On syllabic frontness oppositions in Turkic.' *Varia Eurasiatica. Festschrift für Professor András Róna-Tas.* Szeged, 77-94.

Laut, J. P. 1986: *Der frühe türkische Buddhismus und seine literarischen Denkmäler.* Wiesbaden: Harrassowitz (VdSUA 21).

Mansuroğlu, M. 1959: 'Das Karakhanidische'. *Philologiae Turcicae Fundamenta* I. Wiesbaden: Steiner, 87-112.

Maue, D. 1996: *Alttürkische Handschriften. Teil 1. Dokumente in Brāhmī und tibetischer Schrift. Beschrieben und herausgegeben von D.M.* Stuttgart: Steiner. (Verzeichnis der Orientalischen Handschriften in Deutschland XIII 9)

Maue D. & Röhrborn K. 1984-85: 'Ein "buddhistischer Katechismus" in alttürkischer Sprache und tibetischer Schrift'. *ZDMG* 134: 286-313 (Teil I) and 135:68-91 (Teil II). Text here referred to as BuddhKat.

Moerloose, E. 1980: 'Sanskrit loan words in Uighur'. *JTS* 4: 61-78.

Moriyasu, T. 1985: 'Étude sur un catéchisme bouddhique ouigour en écriture tibétaine (P.t. 1292)'. *Memoirs of the Faculty of Letters, Osaka University* 25:1-85.

Oda, J. 1983: 'Remarks on the Indic "Lehngut" of the Säkiz yükmäk yaruq sūtra'. *Sprachen des Buddhismus in Zentralasien*, ed. K. Röhrborn and W. Veenker, Wiesbaden: Harrassowitz, 65-72.

Öner, M. 1998: *Bugünkü Kıpçak Türkçesi.* Ankara: Türk Dil Kurumu.

Röhrborn, K. 1977f: *Uigurisches Wörterbuch. Sprachmaterial der vorislamischen türkischen Texte aus Zentralasien* (Wiesbaden: Steiner). 6 fascs with continuous page numbering, referred to as UW.

— 1981: 'Zum Wanderweg des altindischen Lehngutes im Alttürkischen'. Hans R. Roemer & Albrecht Noth, eds., *Studien zur Geschichte und Kultur des Vorderen Orients. Festschrift für Bertold Spuler zum siebzigsten Geburtstag.* Leiden: Brill, 337-343.

— 1988: 'Zur Darstellung der Gutturale in den indischen Fremdwörtern des Uigurischen'. *CAJ* 32/3-4: 232-243.

— 1996: 'Zur Suffixklassifikation im Alttürkischen'. *UAJb N.F.* 14: 176-186.

Shōgaito, M. 1978: 'Kodai Uigurugo ni okeru Indo raigen shakuyō-goi no dōnyū keiro ni tsuite'. *Ajia Afurika gengo bunka kenkyū* 15: 79-110.

— 1991-3: *Studies in the Uighur Version of the Abhidharmakośabhāṣya-ṭīka Tattvārthā. Vols. I-II. Text, Translation and Commentary.* Shokado. Here referred to as TA.

Zieme, P. 1969: *Untersuchungen zur Schrift und Sprache der manichäisch-türkischen Turfantexte.* Unpubl. Ph.D. dissertation (Berlin).

— 1984: 'Zur Verwendung der Brāhmī-Schrift bei den Uiguren'. *AoF* 11: 331-346.

— 2001: 'A Manichaean-Turkic dispute in Runic script'. P. Mirecki & J. BeDuhn, *The Light and the Darkness. Studies in Manichaeism and its World.* Leiden: Brill, 209-219.