THE CHINESE NAMES OF THE TIBETANS, TABGATCH, AND TURKS
Christopher I. Beckwith
Indiana University

Perhaps the single most important historical event in East Asia in Late Antiquity was the migration of Mongolic-speaking peoples into the lands of the former Chinese empire in North China, parallel to the Germanic migration into the northern provinces of the Roman Empire. Chinese sources record the names of these and other peoples, but interpretation of the transcriptions remains problematic.

The name ‘Tibet’ has long been known to be a foreign, non-Tibetan name for the country called by the Tibetans Bod, or Bod-yul ‘Bod-land’ (Beckwith 1977, 1987/1993). Many scholars nevertheless have pursued other ideas (e.g., Bazin and Hamilton 1991), especially the long seductive one that the name ‘Tibet’ reflects or includes the native Tibetan name Bod (Mair 1990).

The medieval Chinese name of Tibet and the Tibetans, T’u-fan 吐蕃, from MChi*thɔːspan (Pul. 312, 89, *thɔː-puɑn'), is explicitly said to be 語譯 ‘a linguistic mistake’ (CTS 196a: 5219). The explanation given is essentially that the name T’u-fan 吐蕃 is a mistaken transcription of T’u-fa 秃髮 NMan tūfā, from MChi *thɔːwkpar (Pul. 311, 89 *thɔːwk-puat), which the Tibetans took as the name of their country from a historical leader of that Hsien-pei clan who set-
tled in their territory. Another explanation given is that the Tibetans were descended from the Western Ch’iang, a branch of whom were the Fa Ch’iang 发羌; the pronunciation of fā, from Middle Chinese *par (Tak. 372-373 phar; b Pul. 89 “puat”) 7 and fān was “close,” so the Tibetans were called T’u-fan 吐番 (HTS: 216a: 6071).8 These explanations appear to be due to external information. The Chinese historians must have learned that the Tibetan name of the country was Bod—though that too is not originally a ‘Tibetan’ name (Beckwith 1977)—while the international name was something like *Töpat ~ *Töppat. In view of the other transcriptions, Arabic ـ tubbat, pointed and read in Classical pronunciation as ـ tubbat, probably represents a foreign *tobbat ~ *toppat ~ *többat ~ *töppat (etc.);9 Classical Arabic did not have the vowels o or ō or the consonant p. The idiosyncratic Old Turkic tüpöl [töpät] could perhaps represent an underlying foreign *[töpät].10 The gemination of the labial is represented only by Arabic. It may be supported by or due to contamination with the Arabic name Tubbā’, with which the name of Tibet was usually etymologized (cf. Pelliot 1915: 20).

It is now known that final n in the rhyme to which 輻 belongs—Starostin’s (1989: 579) rhyme class 元D—was pronounced *r in ‘Old Chinese’ (Sta. 579 輻 *par),11 and long continued to be so pronounced in some dialects (Beckwith 2004: 99-102).

In the usual reconstructions of Middle Chinese the non-nasal coronal final is reconstructed as *t. However, there is no question but that

---

6 There is no phonemic distinction between unaspirated and aspirated unvoiced stops in Old Tibetan. In some texts the allophones are written incorrectly nearly as often as correctly (Beckwith forthcoming a).

7 This word always occurs with final -r [t̪] in the A-mi-t'o ching, the original text of which stūtra is unfortunately not reproduced in Takata (1988: 254-261). However, it is the only word that has this irregular pronunciation (the final non-nasal coronal in the rest of the text is -r as expected) and it always occurs in the same construction. The irregularity is clearly deliberate (probably ritualistic) and worth investigating.

8 Tu Yu (FT 190: 5170) more accurately says, “The origin of their nation is not known.” The proposals of the Chinese historians have previously appeared to be irreconcilable with Chinese reconstructions, and have been ignored.

9 Cf. Pelliot (1915: 20). The somewhat late Greek form τουπατ represents [tupat], but it is in no case clearly a loan from the Classical Arabic reading tubbat.

10 According to Old Turkic internal phonology, tüpät or tüpät would be expected, but this is a foreign name and the runic script is ambiguous about the vowels u and ō. Note the Middle and Classical Mongolian forms tübût and tübēd.

11 There is partial external confirmation of the Old Chinese reconstruction, in that 安 NMan ㄠ—which belongs to Starostin’s (1989: 576-577) 元A rhyme class XXXV—was used to transcribe foreign ar, as in the first syllable of the name of the Arsacids. Though Starostin (1989: 578) reconstructs this as *tān, as he himself shows through his listing of many ‘inexact’ rhymes among the different supposed subclasses, the latter were undoubtedly all the same basic rhyme if a unitary ‘Old Chinese’ is assumed. However, it is well known and accepted that dialect and period differences exist in the Shih ching poetry collection; a unitary reconstruction is by definition not rigorous. It is also true that this final became -n very early in some cases, because the name Alexandria was transcribed in the second century B.C. as Ἄρακα Δαυίδ Νίκα (EH 96a: 3888-3889), where (l) NMan shān ‘mountain’, MChī sān (Tak. 366; Pul. 274 ‘šān/šān…’), corresponds to -san in ‘Aleksandria’. The conflict between these two correspondences has not been explained. There is solid foreign transcriptional evidence for OChi *r in the name Ch’in—source of the name ‘China’—which is earliest known in Greek transcription as σρη- σήρ-.
it was \( r \) phonetically in Middle Chinese, since there are many long Chinese texts from Tun-huang transcribed in Old Tibetan script, most dating from the late 8th to the early 10th centuries, in which \( -r \) regularly appears instead of \( *-t \). There are also numerous contemporaneous transcriptions in other scripts, including Khotanese Brahmi, Uighur, and Arabic scripts, and loanwords into neighboring languages, which attest to Middle Chinese final \( r \). In short, there is no question but that there was no final \( *-t \) in ‘standard’ Middle Chinese during the early medieval (Sui-T’ang) period, i.e., at least from the last quarter of the sixth century up to the early tenth century. However, on the basis of the early medieval Chinese material alone (i.e., excluding all non-Chinese transcriptions) it is impossible to say what a given foreign non-nasal coronal final was phonetically in the foreign language transcribed by the Chinese. It is also impossible to determine if the final \( r \) of the Chinese transcriptions of foreign words continues, in part, Old Chinese \( *r \), if it is an innovation, or if it represents a merger of Old Chinese final \( *r \) and Old Chinese final \( *t \), both of which phones are firmly reconstructible for ‘Old Chinese’ and have the regular reflexes \( n \) and \( r \) respectively in standard Middle Chinese.

Reconsidering the Chinese words for Tibet in light of what is known about Middle Chinese, the ‘correct’ transcription T’u-fu 秀髮 (Pul. 311, 89 ӹቻtֽąًwkuat) proposed in the Chiu T’ang shu would have been pronounced in the T’ang ӹቻtֽąًwkp\text{ar}. This is not ‘close’ to the Middle Chinese pronunciation of T’u-fan 吐蕃 as ӹቻtֽąًspan, nor to the international pronunciation of the name, ӹቻtֽąًpət\text{a}. Even if the post-T’ang scholars knew about the archaic reading of 蕃 as ӹቻtֽąn, the first syllable (T’u-) is quite different and would seem to be what was meant by a ‘linguistic mistake’. The ‘correct’ transcription is further supported by the ‘unofficial’ transcription of the name of Tibet in the bilingual glossary from Tun-huang (Pelliot 1961: 143-144, ms. Pelliot tibétain 2762) in which the final velar in the first syllable is attested, namely T’e-fan 特蕃, in standard Middle Chinese pronunciation *dakpan \( \sim *tֽąًkpan \) (Pul. 304, 89, 90 ӹቻtֽąn k\text{pua\text{u}}; LMC ӹቻtֽąn k\text{p\text{ua\text{u}}n}), representing a transcriptional *tֽąnpar or *tֽąkpar. Moreover, while the reconstruction of the Old Chinese final of 蕃 N\text{Man} t\text{u}, which belongs to Starostin’s (1989: 561-564) 魚 rhyme class XIII, is debated, it is agreed that it was, or contained at one point, a velar or laryngeal. In view of the attested alternate transcriptions of the name of Tibet, the first syllable final must have been a velar and in the archaic transcript-

---

12 Virtually the only exceptions in attested material are the loanwords into Old Japanese, which reflect final \( -t \), indicating that the loans either entered Japanese from a southern or eastern dialect that retained final \( -t \) or that the Japanese heard Middle Chinese final \( -r \) as \( [t] \). The latter is a good possibility considering the unusual phonetic nature of the native Japanese phoneme usually transcribed as \( r \). Foreign names transcribed into Chinese for the first time in the Sui and T’ang periods generally use final \( [r] \) to transcribe all foreign non-nasal coronal syllable finals, including \( t, d, r, \) and \( l \), though often they are transcribed as a separate syllable instead. Cf. Note 7.

13 By ‘standard Middle Chinese’ is meant the prestige dialect, based on the dialect of Ch’ang-an (the capital) and the area of northwestern China in general—which was in direct contact with the non-Chinese peoples to the north and west.
tion dialect 吐 (OChi *tʰāŋ) must have been pronounced *tʰǎŋ ~ *tʰawŋ ~ *tʰǎŋ,14 so that 吐蕃 was actually read *tʰawŋpar ~ *tʰawŋpar ~ *tʰǎŋpar in that dialect15 in the early seventh century. In any case it is clear that the dialect which recorded the transcription T’u-fan 吐蕃 retained Old Chinese readings of the rhymes of both characters. There are accordingly three different basic transcriptions of the name in Chinese, an Old Chinese-style one, 吐蕃 T’u-fan (*tʰawŋpar ~ *tʰawŋpar ~ *tʰǎŋpar); a Middle Chinese-style one, T’u-fa 禿发 (*tʰawŋpar); and a mixed one, T’eh-fan 裔蕃 (*tokpar ~ *takpar). All of them transcribe the same spoken Chinese pronunciation of the same foreign name, but they do so using different values for the transcriptional characters.

In view of the unanimous non-Chinese transcriptions of the word-final consonant as r, the reconstruction of the underlying name—the foreign word the Chinese were trying to represent in their transcriptions—should have been *tάŋpat ~ *tʰawŋpat ~ *tʰɔŋpat,16 though the syllables 點 NMan fā (MChi *par) and 髮 NMan ṭā (MChi *par) of the ‘correct’ transcriptions proposed by the post-T’ang historians belong to the same Old Chinese rhyme class, *-uət, which was actually the same as the rhyme class *uɑ̌ (where *ç represents an Old Chinese affricate, [ts] or [ʈʂ]), as shown below.

According to the above analysis, the name T’u-fan ‘Tibet’ probably really is connected to the Hsien-pei name T’u-fa,17 but because both are ultimately transcriptions of a dialect form, *tʰāŋpat, of the name T’o-pa. The name itself thus long predated its application to Tibet, with which place it originally had no connection.18 It cannot be forgotten in this connection that T’o-pa was also the name of the most powerful clan of the Tang-hsiang (TT 190: 5169), the ancestors of the Tangut, who came from northeastern Tibet and moved into the area of Kan-su and Inner Mongolia during and after the Tibetan imperial period (Beckwith 1987/1993: 169-170, n. 174; Beckwith 1987b).

**T’O-PA 拓跋 ‘TABGHATCH’**

According to the attested Middle Chinese pronunciation of inherited alveodental final *t as r, the name of the Mongolic-speaking Hsien-pei people who founded the Northern Wei dynasty, T’o-pa 拓跋 NMan tuóbâ, or *tʰakbat in theoretical Middle Chinese reconstruction

15 In the standard dialect 吐蕃 was undoubtedly read *tʰawŋ (*tʰawŋân) or the like, eventually becoming *tʰawan, and continued to be so pronounced thereafter.
16 Someone from the same dialect may have been responsible for the transcription of the name of the Jou-jan (which has several variants, including Juan-juan and Jui-juan) with what appears to be the same final r. The center of their kingdom was located directly to the north of Tun-huang and Chang-yeh (TT 196: 5378).
17 Similarly, “Peut-être après tout *Thuk-p’anh [Tu-fa], [*Thur-p’anh [T’u-fan], Tüüpit et Tibet ne sont-ils qu’autant de formes d’un même nom” (Pelliot 1915: 20).
18 T’u-fan first occurs in an entry for the year 634 (Beckwith 1987/1993: 21).
were in fact the same name. Since both are Hsien
by Boodberg. The ‘etymologies’ are actually folk
dialectal continuation of Old Chinese final *
the existence of other Chinese transcriptions of this name or variants
of it, including T’u-fa 和鞑 and the names of other Hsien-pei peo-
people in which the first syllable must reflect an archaic dialect
pronunciation *takbar or the like, as shown above. But the attested
Old Turkic form Taβyač has final -č ([tʃ], i.e., [ʃ]), not -t, or -r, both
of which would be perfectly possible phonologically in Old Turkic.
Neither of the early medieval non-Chinese transcriptions (Old Turkic
and Greek) has a final liquid or simple dental stop in this name.22
The final -č is thus significant.

If the final -r in the actual Middle Chinese reading of the name
represented a continuation of Old Chinese final *r rather than Old Chi-
inese final *t, the native pronunciation of the name might have been
something like *takbar as well. But the name was transcribed into
Chinese early enough that it should not be reconstructed according to
Middle Chinese values. Moreover, all sources confirm that the name
of Tibet, which is in origin a form of the same Hsien-pei name, had
final *-t, not *-r. The character 鄂 NMan bá ‘to trample’ in the name
T’o-pa 拓跋 NMan tuòbá is not attested in the Shih ching (Book of
Odes) rhymes, but its character belongs either to Starostin’s (1989:

19 Ligeti (1970: 290) notes, “il y a lieu d’admettre que l’étymologie sans doute
populaire recueillie par un texte chinois d’après laquelle t‘o-pa, ach. t‘ak-bar, c’est-
dire tay-bar (ou tayo-bar) signifie ‘maître de la terre, du sol’ repose sur le fait que
dans le sien-pi de cette époque il existait réellement un terme tay signifiant ‘terre, sol’.
Note also the Old Chinese pronunciation of 土 ‘earth, soil’, *t‘aγ, and its archaic
dialectal continuation attested by the transcriptions *t‘awγ ~ *t‘ak, etc.
There seem to be no examples of *-γβ in Old Turkic; see Clauson (1962: 169-
170), who does not discuss *-βγ in Taβyač and Yaβyač, both of which words he does
mention. Pelliot (1959, I: 217) inexplicably says that the Chinese form is meta-
thesized and “Taβyač, Taβyač […] most probably renders the original form of the
name of the Altaic tribe which founded […] the dynasty of the Northern Wei (386-556).”
21 Ligeti (1970: 290 n. 45), citing Chinese etymologies of the name T’o-pa quoted
by Boodberg. The ‘etymologies’ are actually folk-etymologizing variant transcrip-
tions. The archaic Chinese dialect pronunciation of T’u-fa was so close to the pronun-
ciation of T’o-pa that one was considered a form of the other by the early Chinese.
Since both are Hsien-pei clan names, it is difficult to avoid the conclusion that they
were in fact the same name. Cf. the discussion of T’u-fan.
574-575) Old Chinese rhyme class XXIX 祭 A\textsuperscript{23} *-ač, or to XXXII 月 A *-at. He reconstructs the related characters 輯 and 拾 as *bh(á)t and *b(h)áč respectively (Starostin 1989: 575).\textsuperscript{24} According to Starostin, this is a неточный (‘inaccurate, inexact’) rhyme, but these putatively distinctive Shih ching rhymes actually freely rhyme with each other, and in fact, both rhyme classes include more attested ‘inaccurate’ rhymes than ‘accurate’ ones, and each occurs equally often in rhymes with the other class; similarly, other syllables with the same final *-č interrhyme with syllables reconstructed with final *-t (Starostin 1989: 570-576). A clear example of this is Shih ching Ode 16, in which there could only have been one rhyme, though Starostin and Baxter have reconstructed three rhymes, one with final *-t and two with final *-č (or *-ts),\textsuperscript{25} evidently based on the traditional rhyme categories (Legge 1966: 26-27).\textsuperscript{26} This is clearly a very old error. Its significance for the Ode in question—and for the Shih ching rhymes as a whole, or ‘late’ Middle Old Chinese—is that there is no reconstructible phonemic difference between the codas *-t and *-č (or *-ts). In short, 拾 NMan ǝ́ in the name T’o-pa 拾 could have been pronounced either *bat or *bač in an archaic Chinese dialect; it is impossible to decide on the basis of the Chinese alone. Similarly, the character 髪 NMan ǝ̄ ‘hair’, representing MCh. *par (Tak. 372-373 phar) in the Hsienpei ethnonym T’u-fa 尸 in the Hsienpei ethnonym T’u-fa 尸 (Pul. 311, 89 ǝ̄₃₂₃wk-puat), also belongs to this rhyme, and could be reconstructed for Old Chinese as *par ~ *pač. The Old Turkic affricate (along with its Greek version) suggests that the final consonant was the affricate.

The Chinese sources relate that the T’o-pa themselves say the name T’o-pa means 土土 ‘ruler of the earth’. The account in the Wei shu (1: 1),\textsuperscript{27} after recounting the legend that the Hsienpei were descendants of the Yellow Emperor (黄帝 Huang tí), says:

The Yellow Emperor ruled by the power of Earth;\textsuperscript{28} in the North they call 土 (NMan ǝ̄) ‘earth’ 拾 *Ɂa and they call 后 (NMan ǝ̄) ‘ruler’

\textsuperscript{23} Rhyme classes are often referred to rather than specific rhymes of specific attested characters because in many cases a particular character is not attested in a Shih ching rhyme and it is then necessary to use a phonetically related character. Naturally, the resulting uncertainty causes major differences among reconstructions.

\textsuperscript{24} Cf. Karlgren (1957: 87); Baxter (1992: 746) reconstructs 拾 *bat and 拾 *bots.

\textsuperscript{25} The two rhymes with final *-č are reconstructed with different vowels, *a and *o. But this is problematic. It is certain that the vowel *o split and became *wa in most cases very early in Old Chinese (Beckwith forthcoming b). Since the first set of rhymes has the vowel *a as well, the supposed distinction among them in Ode 16 disappears.

\textsuperscript{26} In Middle Chinese there is no distinction between the two categories, despite some reconstructions. The Tun-huang texts regularly have -ar, although the rhymes involved are spread over several categories in the Ch’ieh-yün (Lung 1968), each including examples from the series built on the phonetic 龍 (Karlgren 1957: 87, phonetic 276a), including 羽 (rhyme 月9; Tak. 372: phar), 拾 (rhyme 月11; Tak. 364 *bar), and 拾 (rhyme 齿12; Tak. 366 par, phar). For 拾 *bar, the attested spellings are bad and ba. They may reflect archaic pronunciation; cf. note 7.

\textsuperscript{27} It is unclear why Tu Yu (TT 196: 5373) has altered this to read: ‘Some say that they are the distant descendants of the Yellow Emperor. Because the earth of the Yellow Emperor is virtuous, and they call 土 ‘earth’ 拾 *Ɂa, and 后 ‘ruler, lord’ 拾 *bač, accordingly they took [*Ɂakbač] as their clan-name.’

\textsuperscript{28} For an explanation of the ideology referred to see Wang (2000: 138 et seq.).
While there is no way to know if ‘ruler(s) of the earth’ is indeed the actual etymology of the name *Taγβαc, it is unquestionable that the T’o-pa (or at least an who knew the T’o-pa language) thought it was and so explained it. There is thus some likelihood that the pieces of the word did have the meanings the Chinese say the T’o-pa said they had. Based on the form *taγ in the premetathesized form *taγbač of Old Turkic taβyač, and on the reconstruction of other Chinese forms of the name, including T’u-fan ‘Tibet’ (q.v. above), T’o- ṭo MChi *ṭak- should represent a T’o-pa form *taγ-. Ligeti notes that the Chinese did indeed identify the first syllable correctly with a word in the T’o-pa language since there is a word *taγ meaning ‘earth’ in Mongolic (Ligeti 1970: 289-290), to which language family T’o-pa, a dialect of Hsien-pei, belongs, as demonstrated conclusively by him in the same magisterial article.29 Accordingly, the second syllable should be right too. That syllable, based on the Chinese transcription 跡, i.e., NMan ṭa from theoretical *bat ~ *bač, should, as shown conclusively by him in the same magisterial article, and on the premetathesized form of the Old Turkic transcription, should be *bat ~ *pač ~ bač (i.e., T’o-pa language) as their clan-name.

In the century and a half preceding the T’o-pa conquest, North China was intellectually conquered by Buddhism, a religion originating in India. Hundreds of Indic loan words and calques (literal translations of foreign expressions) were introduced into Chinese. The expression ‘lord of the earth’, meaning ‘sovereign’, well known in Buddhist texts (Wogihara 1986: 728, Skt viṣayapati), contains the Indic word pati30 ‘lord, ruler’, usually translated as 主 NMan ṭa ‘ruler’. The Chinese transcriptions reconstructed as *bat ~ *pati are fairly accurate representations of a Prakrit form of the Sanskrit word pati.31 Assuming this is the origin of the second syllable of T’o-pa would give the name the theoretical underlying pronunciations *taypati and *taγpati, the latter being in fact the reconstruction of its form in the name T’u-fan ‘Tibet’ (q.v. above).32 In the T’o-pa name borrowed into Old Turkic it was *bač (or vač ~ wač) and in the Chinese transcription T’o-pa it is *bač ~ *bat.

---

29 The doubts that have been expressed by some since are based on the quotation of earlier works, notably Pelliot’s, on whose actual views see Ligeti (1970: 277).


31 Words with initial ṭ in Sanskrit very often have voiced initials in the Central Asian Prakrits and their Chinese loanforms. “Il se peut très bien que les emprunts aient été faits non pas au sanscrit, mais à un de ces prâcrits si copieusement sonorisés dont les inscriptions de l’Inde, les manuscrits d’Asie centrale et les transcriptions chinoises archaïques nous attestent l’ancienne expansion” (Pelliot 1915: 25 n. 2).

32 This suggests a local Kansu-Kokonor area pronunciation of pāti as *pati. The later Chinese transcriptions (such as 跡提 MChi *patey) have initial *p-.
The attested final ē of the Old Turkic form is not problematic, as it is the regular Mongolic development of *t when followed by *i. This change would have been normal for T’o-pa, a Mongolic language, and probably for Jou-jan, which seems to have been a dialect of it. The continuant initial *β is not directly reflected in any of the Chinese transcriptions of the name T’o-pa or its variants T’u-fa, T’u-fan, etc. While the T’o-pa form of the word is ultimately from a Central Asian Prakrit form of Indic pati ‘ruler’, Indo-Buddhist terms were largely first introduced to China by Central Asians who used not Sanskrit but Prakrits (cf. Coblin 1983), in which initial p- is often voiced (Pelliot 1915). In Gândhārī Prakrit the reflex of pati sometimes appears as vati, and indeed the word is most commonly transcribed in the early (third century CE) Chinese Buddhist text translations as 越 NMan yuē33 (cf. Soothill and Hodous 1937/2003, s.v. 椰越 ‘donor’, corresponding to Skt dānapati)—theoretical Middle Chinese *wat (CY 5: rhyme 9 月; Pul. 388 *wuat) from Late Old Chinese *wat ~ *wač. In view of the Turkic transcription (*β’ač), the Chinese transcription in the name T’o-pa (*bat ~ *bač), and the Chinese loanform *wat ~ *wač of this word, it appears that the Hsien-pei Mongolic speakers borrowed the word from foreign Buddhists who used the Prakrit pronunciation vati, giving the early T’o-pa Mongolic form *β’ač, which was transmitted eventually to the Turks as *β’ač. Because the early Chinese Buddhists had already borrowed a form of the word pati, Prakrit vati, as *wat ~ *wač before the rise of the T’o-pa, the *bat ~ *bač in the name T’o-pa should represent the T’o-pa pronunciation of the loanword *β’ač ~ *vač ~ *bač ‘ruler’.

The name T’o-pa thus refers at once to both the Chinese Yellow Emperor and an Indo-Buddhist world ruler, and in both traditions, not only to the element Earth and the color Yellow but to the Center, the direction of rule.34 It is not surprising that a people attempting to establish their rule over a vast realm would have carefully chosen a royal clan name that literally means ‘Rulers of the Earth’.

T’U-Chūeh 突厥 ‘TURK’

The Chinese transcription of the name of the Turks, T’u-chūeh 突厥 NMan tūjuē, in Middle Chinese *turkwar ~ *durkwar (Pul. 168, 311 التوکوات ~ الوکوات; 35 Tak. 372 kwar)36 has long been problematic. The now traditional explanation of Pelliot, going back ultimately

33 Jan Nattier, p.c., 2005. I am of course fully responsible for any errors.
34 The Indo-Buddhist system of five elements varies somewhat from the Chinese one, and many of the correlations are different, but they agree on Earth being yellow and square and in the center (Soothill and Hodous 1937/2003, s.v. 土, 地, and 方).
35 Pulleyblank (1991: 311) says, “The modern reading may be for E. *t’wat, written in the Guangyin with radical 40 instead of 118 but given in the Jyoun as a reading for this character.” In fact, both the *t and *d- readings of 突 are given in the Ch’ieh-yün (CY 5: rhyme 10月: 他骨反 and 豕忽反; cf. Kar. 134 *t’uat ~ *d’uat-
36 The Ch’ieh-yün gives both unaspirated and aspirated initial pronunciations for 突 (CY 5: rhyme 9月).
to Marquart (Harmatta 1972: 263-264), is that it represents *Türküt, though it has since been argued that it derives not from Turkic directly via a Mongol plural suffix but from a hypothetical Sogdian *twrkyt, i.e., *turkît (Harmatta 1972), a Sogdian -t plural form of *twrk ‘Turk’. Yet despite the existence of many hundreds of contemporaneous early medieval transcriptions of the name Türk in Greek, Arabic, Tibetan, and many other languages, there are no transcriptions like *Türküt or *Turkît in any language, including Sogdian. The name *Tu'chüeh 窪, meaning ‘Türk’, is first recorded in an account dated to 545 A.D., less than two decades before the appearance of the Greek transcription Τουρκ [turk] (Moravcsik 1958, II: 320-321), followed about a century later by the Arabic and Old Tibetan transcriptions. The first syllable of the underlying name Türk is thus clear. The remainder of the syllable written with 窪 is the problem. Several facts must be taken into consideration.

Firstly, because theoretical Middle Chinese syllable-final *t was actually pronounced r, the transcription T’u-chüeh in that period in the standard dialect was read *turkwar. Transcriptions in non-Chinese writing systems represent only the root τürk. The second syllable vowel in the Old Tibetan and Old Khotanese transcriptions is clearly an internal Tibetan or Khotanese development. Except for Chinese, then, there is no second syllable in any early transcription of the name Turk from anywhere, including the earliest clear examples—Greek Τουρκ [turk], Arabic *twrk- [turk], and Old Tibetan drug-[drük] ~ [drük].

Secondly, the one putative textual attestation of the theoretical underlying Sogdian plural form *Turkît (Harmatta 1972: 273) cannot be a form of the word τürk at all, not to speak of a plural form of it. The exception is said to occur in the first line of the Sogdian inscription from Bugut, dated to the late sixth century A.D., which is a hypothetical transcription of the word *türk.

Bailey (1982: 85) comments on Harmatta’s theory, “the Chinese may have learnt the name of the Turks from the Sogdians and hence the name may have been *turkît (this would be rather -êt) as a Sogdian plural. But this -êt (rather -êt) can only be the later Sogdian plural of a stem in -aka-, which is in older Sogdian -k, but in Manichean Sogdian ʒʒ, -y (that is, -ê, later -i, from which -aka-tâ- gave -êtâ-.” The one putative example in Sogdian script does not exist; see below. Pulleyblank (1965) argues that the two characters of T’u-chüeh are a unitary transcription of the word Türk.

Clauson (1962: 84) argues that the original form of the name was türkö be-
cause a second syllable stem vowel *-u- could theoretically be attested in one of the two Old Turkic spellings, which is ambiguous on this point. However, the unambiguous spelling τürk also occurs in the very same texts, as he himself shows. Pulleyblank (1965) has disproven the rest of Clauson’s theory.

Bailey (1982: 84) notes that türkə- ‘Turk’ occurs only once in Old Khotanese texts, and “elsewhere always trûka-.” I am grateful to Peter Golden for this reference. Clauson (1962: 86) adds the Khotanese spellings türki, türki. The variants show that the base form was türk- or trûk-, and that there are no forms with final -u. The second Turkish vowel thus did not come from Khotanese. The Tibetan base form is probably drug, which usually occurs only in compounds. It has been extended within Tibetan by the addition of the diminutive noun suffix form -a (after g; as in myugu ‘grain sprout’, smyuγu ‘pen’, lugu ‘lamb’), perhaps to help distinguish it from drug ‘six’. The name is used in the Old Tibetan Annals exclusively to refer to the Western Turks (Beckwith 1987/1993: 63-64, n. 56).
from Bugut, dated to the late sixth century A.D., which is a funerary inscription for Tatpar (formerly read ‘Taspar’) Qaghan erected by his son and successor (Yoshida and Moriyasu 1999).42

In the first line of the inscription the clearly written letters tr’wkt occur. They have been read as the Sogdian -t plural of Türk by Kljaštornjy and Livšic (1972: 85, 87-88; 1978: 54) and Yoshida (p.c., 2005). These editions interpret tr’wkt as a metathesized form [truk] of the regular Sogdian spelling tərk [türk] ‘Turk’.

Yoshida (p.c., 2005; Yoshida and Moriyasu 1999: 123) reads and translates Line 1 of the Bugut Inscription as follows:

\[rt\ y\ mw\ 'n\ k\ n\ w\ m\ n\ k\ 'wst\ 'nt\ tr'wkt\ 'šyn\ k\ w\ tr'wkt\ 'xšywn'k\]

‘Kings of the Turkish Ashinas clan have established this stone of the (Buddhist) law.’

He notes (Yoshida, p.c., 2005) that there is a similar phrase in the Sogdian face of the Karabalgasun Inscription; he reads and translates the sentence in which it occurs as:43

\[šn\ k\ nty\ t\ r\ k\ 'xš\ w\ n\ d\ r\ 'st'nt\]

‘They took the ruler of the Tujue [= T’u-chüeh] of the Ashinas clan.’44

The string tr’wkt in the Bugut Inscription is clearly written, and could be either one word or two words, tr and ‘wkt, as there is a slight space between the resh (r) and the aleph (‘)—as noted in the alternate transcription tr’-wkt (Yoshida and Moriyasu 1999)45—if not three words or morphemes, such as tr, ‘w, and kt. The putative word trwkc in the second line according to Kljaštornjy and Livšic (1972: 85; 1978) is

42 I am deeply indebted to Yutaka Yoshida for generously answering my many questions about this inscription and other Sogdian issues. I would especially like to thank him for sending me photocopies of his and Moriyasu’s article and of his photographs of the rubbings, as well as a copy of the handout from his 2003 lecture at the Collège de France giving parallel quotations of his and Kljaštornjy and Livšic’s editions and translations. Without his help I would not have been able to discuss this problem in any detail. I should note that he does not agree with me on several important points, and I am of course responsible for any errors that might remain.

43 Y. Yoshida, p.c., 2005. He notes further that kwtr’t ‘is a plural form of kwtr, loanword from Sanskrit gotra, and knty is an original Sogdian word. Although there is a difference in the word order of Turk and Ashinas-clan, I do not think it significant. Incidentally, the passage in KB [the Karabalgasun Inscription] describes the event in which Uighurs, conspiring with Basmils and Qarluqs, conquered the second Tujue empire and there is practically no doubt as to what the expression refers to.” Yoshida’s edition of the Bugut Inscription also eliminates the putative occurrence of ββββμον or ββυ ββμον proposed by Kljaštornjy and Livšic (1978: 54), who interpret it as “Бымн [= Bumîn]” Qaghan. There is no -ββμον- in the text, as far as can be seen on the photocopy of the rubbing. Yoshida reads it as wmn’; ‘Umna’ (Qaghan). Also, the Chinese texts transcribe the name ‘Bumîn’ with an initial *t, ±[3] T’u-men, i.e., *Tumîn (CS 50: 908). The Chinese transcription, dated to 545, is two centuries earlier than the Old Turkic inscriptive form of the name, Bumîn, but only a few decades earlier than the Bugut Inscription. There are also other considerations, but unfortunately there is no space here to discuss this problem further.

44 Translating this completely into English, it would mean, according to his interpretation, ‘They took the ruler of the Turks of the Ashinas clan.’ His and Moriyasu’s (1999: 123) published translation of the Bugut Inscription reads, “Kings of the Turkish Ashinas tribe have established [this] stone of law […].”

45 He considers that tr and ‘wkt could not be two independent words (Y. Yoshida, p.c., 2005).
translated as “Turkish” in their rendering. However, the first letter is not a  튵 according to the palaeographic forms of the Bugut Inscription, so there is no such word in the text. It may be thought that the reading of 튵kt as *Turkit in line one would seem to be assured by the immediately following string in the same line, which Yoshida reads as ‘ ’. However, Kljaštornyj and Livšić read it as c(y)nst’n ‘China’. In fact, the rubbing is unclear right at that point and it is uncertain what this particular string really represents; other readings could be proposed as well. The putative *Ashinas reading in both inscriptions is itself partly—if not wholly—dependent on the interpretation of 튵kt as *Turkit in the Bugut Inscription. Yet this interpretation is extremely problematic at best.

The word 튵 ‘Turk’ is unambiguously attested in Sogdian in the following texts:

1. a contract for the sale of a female slave dated to 639 A.D., where 튵strn ‘Turkistan’ is described as her place of origin;
2. the Karabalgasun Inscription, where the words 튵c’ny ‘Turkish’ and 튵 ‘Turks’ occur;
3. the Mt. Mugh documents (early 8th century), where the personal name 튵 ‘Turk’ occurs;
4. the Mahrnāmag (early 9th century), where the personal name 튵 ‘Turk’ occurs;

In all of these Sogdian texts the word ‘Turk’ is spelled 튵. The Bugut Inscription, with the putative example 튵k-, would be the only occurrence of such a spelling of ‘Turk’ in Sogdian. This very same problematic example also provides the lone attestation of a putative plural form 튵kt ‘Turks’. Yet 튵 in the Karabalgasun Inscription actually has plural meaning, ‘the Turks’, although it is singular in form. The unquestionably attested form 튵 ‘Turk’ is thus not formally pluralized in Sogdian even in the one undoubted instance when it is plural. In fact, ethnonyms are rarely, if ever, pluralized in any of the contemporaneous non-Indo-European Asian languages in which the name ‘Turk’ occurs, even when pluralization is an option, as in

---

46 Yoshida and Moriyasu (1999) read it as y’rwk’ Yaruka (which occurs in a phrase rendered as “Muqan Qaghan’s Yaruka-brother”). In the second line there is also a string wurkप ’r, read as “Urkupar” (in a phrase rendered as “for the sake of Urkupar Cracu Magha [...]”). What is interesting about this word is the sequence wurk [urk], which agrees with the exceptionless Sogdian spelling of the name of the Turks, 튵, not with the sequence 튵k- in the preceding line.

47 I owe this information to Y. Yoshida (p.c., 2005). I am of course responsible for any errors.

48 As in other Indo-European languages, ethnonyms are sometimes formally pluralized in Sogdian (Y. Yoshida, p.c., 2005). There are to my knowledge no examples of *Türklär ‘Turks’ in the Orkhon inscriptions or in other Old Turkic texts, and needless to say, the hypothetical form *Türküt does not occur there either. Peter Golden (p.c., 2005), notes similarly, “I cannot recall ever seeing Türkler in an Old or Middle Turkic text. The same is true for other ethnonyms (e.g. Qarluq, Türgesh et al.).” The word Türk ‘Turk’ is thus never pluralized in Old Turkic. The situation is similar in other non-Indo-European Asian languages.
Arabic. In addition, the Chinese transcription T'uve-chueh, which is ultimately the motivation for this reading, could never have been read *turkit, since a foreign syllable *kit could not have been transcribed by MChi *kwat, theoretical MChi *kwat (Pul. 168). The only possible conclusion is that whatever the Bugut Inscription string *turkit represents it is not *turkit ~ *turkit ‘Turks’. The meaning of Sogdian tr’wkt in the Bugut Inscription is unknown, and no form *Turkut ~ *Turkit exists in that language.

Thirdly, the second syllable in T’u-chueh belongs to the same rhyme as the second syllable of the name of the T’o-pa, which is transcribed in Old Turkic form as Tañjçu, representing metathesized *Tañýçu. The Chinese name T’u-chueh represents a close transcription of two syllables: *turk ‘Türk’ plus *waç. The second syllable *waç is identical to the second part of the compound name T’o-pa, *Tañýçu [tanýatš], namely *baç, which also occurs in the Chinese transcription of the name, *Tañýkuç, in which the second syllable *baç is explicitly glossed in Chinese as ‘ruler’. The syllable *waç in the name T’u-chueh is thus identifiable with the Old Turkic transcription *baç (waç ~ waç) of the T’o-pa word *baç ‘ruler’, a loan from Indic vati ~ patti ‘ruler’, then a well-known international culture word with powerful resonance that had been transmitted by Buddhist mis-

49 In Arabic the word al-Turk ‘the Turk; Turk; Turks; the Turks’ is generally used as a collective or non-specific plural, e.g., in the frequent expression bilād al-Turk ‘land of the Turks’. When early Arab writers wished to refer to a plurality of Turkic nations or tribes, they would sometimes use al-Arāk ‘the Turks; Turks’, though even in such cases al-Turk is more frequently used, as in the title of Maḥmūd al-Kāshgarī’s famous book, Dīwān lughāt al-Turk ‘Account of the Languages of the Turks’. In short, the word al-Turk is grammatically singular, but as it is semantically non-distinctive for plural marking it can be used to represent singular, collective, or plural number. The same appears to be true for Sogdian.

50 Harmatta (1972) argues that the Sogdian can transcribe *kit because it is associated with the same rhyme as 月, and the Turkic name Ch’u-yueh 處月 MChi *cieguar transcribes *cīgl. This is not correct. The second syllable of the Chinese could hardly transcribe *gil (or *gīl). However, Ch’u-yueh could perhaps represent a foreign *cīgl.

52 The immediately following string in the Bugut Inscription is read by Kljaštornyj and Livšic (1972: 85) as c(y)n’t*n ‘China’ and by Yoshida and Moriyasu (1999) as (’)ñ-y-n’s ‘Ashinas’, but it is partly damaged and unclear; the reading of the same name in the Karabalgasun Inscription is also uncertain. In view of the nonexistence of the word ‘Turk’ in the extant text of the Bugut Inscription, Yoshida and Moriyasu’s translation, “Kings of the Turkish Ashinas tribe have established [this] stone of law,” should be revised.

53 Clauson (1972: 438) remarks, “Tavvaç [i.e., Tañjçu] was a frequent component in Karakhanid royal titles.”

54 The Jou-jan and early Türk were also under very heavy Indo-Buddhist and Chinese influence. Several rulers had Indic names or titles, including P’o-lo-men 喃鋩門 ‘Brahman’, a paternal relative of the last Jou-jan ruler, A-na-kuei 阿那瓊 ‘Anakwai Qaghan, active in the 520s (TT 196: 5381). A well-known Indic-titled ruler among the Türk is Muqan Qaghan’s third successor 擊圖 She-tu, usually called Sha-po-lięh 沙錦略 in the sources and transcribed as ‘Išbara’ or ‘Ishbara’ by scholars, namely Mo-ho Shih-po-lo K’o-han 莫何始波羅 (TT 197: 5404) *ma-ha-si-par-la, ‘Mahēvära’ Qaghan. ‘Išbara’ has previously been identified as a loan from Sanskrit śāvara, but I have not been able to identify where or by whom the identification has been made. While Mo-ho 莫何 represents theoretical MChi *ma-pa, the many attested forms in both Tibetan and Khotanese Brahmi are unanimous in transcribing 何 as ha (Tak. 304-305). This is undoubtedly a title, since Mahēvära (from mahā ‘great’ + śāvara ‘sovereign’) means ‘great king’; it is a title of Siva, ‘king of the gods’, and is also equated with Brahma, ‘the Lord of the world’ (cf. mahāśāvara, with mahī ‘earth, land,
sionaries to China long before the rise of the T’o-pa. It occurs independently as 越 *wat ~ *wač ‘ruler’ in many solidly attested examples in Buddhist texts from the third century on. The development of the affricate final č ~ ě, explainable within Mongolic, was evidently transcribed as an affricate by both the Turks and the Chinese.55 The first embassy to the Türk was sent in 545, when the T’o-pa were still ruling large parts of North China. Use of a T’o-pa word for ‘ruler’ would make sense in a diplomatic situation where people would have been speaking T’o-pa or Jou-jan.

T’u-chüeh 突厥 is thus a clear transcription of a foreign *türk-wač, a compound consisting of the name Türk plus the word *bač ‘ruler, lord’. It means ‘Rulers of the Türk’ or ‘the Türk Rulers’.58

REFERENCES


55 The Chinese transcription, considered purely internally, is ambiguous with respect to the final. Three reflexes of patti—*wač, *bat ~ *bač, and *pat—are attested in the Central Eurasian-North Chinese milieu, though perhaps not in the same areas, so an underlying Chinese reading *turk-wač would be possible if it were not for the confirmation of the affricate provided by the Old Turkic and Greek transcriptions. In order to determine more precisely the sequence of borrowing—and in particular, the development of the ‘western’ names with *pat, T’u-fa and T’u-fan—it would be necessary to establish the chronology and areal location of Chinese transcriptions of different forms of the word patti ‘ruler’.

58 The exact parallelism between the names *tay-wač ‘T’o-pa’, *tay-pat ‘Tibet’, etc., which mean ‘Rulers of the Earth’, and *türk-wač ‘Rulers of the Türk’ (or ‘Türk Rulers’), and the somewhat similar phonetic shape of the first syllables, might suggest the possibility that the word türk, the etymology of which is unknown, could also mean ‘Earth’ in its original language. All these names would then mean the same thing, ‘Rulers of the Earth’. Cf. Pelliot’s remark in note 17.


forthcoming a. The Sonority Sequencing Principle and Old Tibetan Syllable Margins.


Ling-hu, Teh-fen 1971. 周書 Chou shu. Peking: Chung-hua shu-chê. (= CS)

Liu, Hsi 1975. 輯唐書 Chi T’ang shu. Peking: Chung-hua shu-chê. (= CTS)


Pan, Ku and Pan, Chao 1962. 漢書 Han shu. Peking: Chung-hua shu-chê. (= HS)


Wei, Shou 1974. 魏書 Wei shu. Peking: Chung-hua shu-chü. (= WS)
