Linguistics of the Tibeto-Burman Area
Volume 35.2 — October 2012

A NEW TRANSCRIPTION SYSTEM FOR OLD AND CLASSICAL TIBETAN [*]

Guillaume Jacques
CNRS-EHESS

Abstract: This paper proposes a new transcription for classical and old Tibetan based on IPA. This new system presents the advantage of being directly readable by non-specialists, without prior acquaintance with the Tibetan script.

Keywords: Tibetan script; Old Tibetan; phonology

1. INTRODUCTION

Tibetan is nowadays almost universally transcribed using Wylie's (1959) system or modifications thereof (Imaeda 2011). Given the fact that Wylie straightforwardly represents the letters of the Tibetan script, it may not be obvious why anyone would ever consider devising a new transcription.

The Wylie system is indeed perfectly suited for philologists working on Classical and Old Tibetan texts. However, as a tool for linguists working on Tibetan dialectology and comparative linguistics, this system presents several inconveniences, and is highly misleading for non-specialists. This has the unfortunate consequence that works on Tibetan historical phonology are difficult to read not only for general historical linguists specialised in a different language family, but also for Sino-Tibetanists working on a different branch of the family.

The Wylie transcription has four main defects to be addressed. First, the phonetic value of several letters and digraphs can be unclear, especially the alveolo-palatal obstruents *c j zh sh*. Second, the problematic letter འ (sometimes incorrectly called *ɦa-tɕʰuŋ*) is transcribed by the apostrophe ', a symbol that does not reflect its real pronunciations, and which cannot be capitalised in person or place names. Third, the two clusters ཇ- *gya-* and གྱ- *g.ya-* have to be distinguished by the addition of a dot. Fourth, the last letter of the alphabet, ཨ, is not represented in the Wylie transcription.

Given the fact that Unicode-compatible systems are available on nearly all computer systems, there is no need anymore for a type-able 'practical' transcription, as may have been the case in earlier times (Hill 2012). Since the pronunciation of Old Tibetan is relatively better known in comparison to that of many other old languages (Hill 2010), it seems more sensible to represent the Tibetan letters by their IPA equivalents. This system has the advantage of limiting to the minimum the preliminary explanations when discussing Old Tibetan data in articles dealing with historical phonology.

[*] I wish to thank Nathan Hill, Randy LaPolla, Jackson T.S. Sun, Nicolas Tournadre and an anonymous reviewer for comments on this article. I am responsible for any remaining errors.

2. THE BASIC CORRESPONDENCES

The following table represents the basic correspondences between Tibetan letters, Wylie transliteration, and the new transliteration system proposed in this paper:

Tibetan Alphabet	*Wylie*	*Present system*
ཀ	*k*	*k*
ཁ	*kh*	*k^{h}*
ག	*g*	*g*
ང	*ng*	*ŋ*
ཅ	*c*	*tɕ*
ཆ	*ch*	*tɕ^{h}*
ཇ	*j*	*dʑ*
ཉ	*ny*	*ɲ*
ཏ	*t*	*t*
ཐ	*th*	*t^{h}*
ད	*d*	*d*
ན	*n*	*n*
པ	*p*	*p*
ཕ	*ph*	*p^{h}*
བ	*b*	*b*
མ	*m*	*m*
ཙ	*ts*	*ts*
ཚ	*tsh*	*ts^{h}*
ཛ	*dz*	*dz*
ཝ	*w*	*w or ɦ^{w}*
ཞ	*zh*	*ʑ*
ཟ	*z*	*z*
འ	‘	*ɦ or ^{n}*
ཡ	*y*	*j or ^{j}*
ར	*r*	*r*
ལ	*l*	*l*
ཤ	*sh*	*ɕ*
ས	*s*	*s*
ཧ	*h*	*h*
ཨ	-	*ʔ*

Table 1. Correspondences between transcription systems (consonants)

The vowels are transcribed as in the Wylie system:

Tibetan Alphabet	*Wylie*	*Present system*
ཨ	*a*	*ʔa*
ཨི	*i*	*ʔi*
ཨུ	*u*	*ʔu*
ཨེ	*e*	*ʔe*
ཨོ	*o*	*ʔo*

Table 2. Correspondences between transcription systems (vowels)

The present system differs from Wylie in the following ways:

(i) Aspirated consonants are transcribed by ^{h} instead of simple h.
(ii) The letter ཨ is transcribed with a distinct letter *ʔ*
(iii) Alveolo-palatal stops and fricatives are transcribed with their IPA symbols
(iv) The palatal and velar nasals are transcribed using the IPA symbols ɲ and ŋ
(v) The letters ཀ and ཡ are transcribed in different ways depending on their position in the word (see the following sections).

Note that as in the Tibetan script and in Wylie transcription, the final stops and the stops occurring in clusters are always transcribed as voiced, although they were probably unvoiced in many cases. For instance, in forms such as *bkab* 'to cover.PST' and *ltɕags* 'iron', the stops /b/ and /g/ occurring next to an unvoiced obstruent or word-finally were certainly devoiced. However, it seems unreasonable in the system of transcription to attempt to represent the phonetic reality of Old Tibetan. The voiced stops here represent archiphonemes, as the voicing contrast is neutralized syllable-finally and as the first element of a cluster whose main consonant is unvoiced. Aside from this, some clues from morphophonological alternations in modern dialects suggest that the final stops are underlyingly voiced (see Sun 1986: 35-36).

3. THE LETTER ཀ

The exact pronunciation of the letter ཀ is the topic of an ongoing controversy (see Sun 1986: 112-115, Coblin 2002, 2006; Hill 2005, 2009, 2010). Coblin argues that it represents prenasalisation when used pre-consonantally, a voiced fricative [ɦ] or [ɣ] word-initially, but that it could also be a diacritical symbol in transcription of foreign words (especially from Chinese). Hill on the other hand claims that this letter was pronounced [ɦ] / [ɣ] in all contexts, including pre-consonantally and syllable-finally.

A transcription system cannot address in detail this complex issue, but we suggest the following solution: to use a different symbol of pre-consonantal, plain initial and syllable final positions:

Tibetan Alphabet	*Wylie*	*Present system*	*Meaning*
འགྲོ་	*'gro*	*ⁿgro*	'to go'
འཕུར་	*'phur*	*ⁿpʰur*	'to fly'
དགའ་	*dga'*	*dga(ɦi)*	'to like'
བཀའ་	*bka'*	*bka(ɦi)*	'word' (honorific)
འོད་	*'od*	*ɦod*	'light'

Table 3. Examples of words containing the letter འ

In pre-consonantal position, we represent འ by the prenasalisation symbol ⁿ (which can be encoded as a distinct Unicode glyph). In all other contexts, we use *ɦi*.

In final position, the letter *ɦi* is visibly a mere symbol for disambiguating syllabification: without འ, the combination <d+g> དག would represent /dag/ since in the Indic writing systems the short /a/ is not indicated by any symbol. Hill (2010) points out however that the distribution of འ is not entirely predictable in Old Tibetan texts, and proposes that it might have been a genuine consonant.

We suggest therefore representing this letter as *ɦi* syllable-finally when transcribing texts or quoting sentences, but to omit it when discussing individual words (especially their evolution into modern dialects).

4. THE WA-ZUR AND THE YA-BTAGS

The symbols *wa-zur* and *ya-btags* which appear under the main consonant letter in the Tibetan script are represented in the present system with the IPA labialisation and palatalisation symbols ʷ and ʲ. This solution has the advantage of distinguishing between the two groups ཇྱ- *gʲa-* and གྱ- *gja-* without any need for a separating dot as in the Wylie system:

Tibetan Alphabet	*Wylie*	*Present system*	*Meaning*
ཕྱོགས་	phyogs	*pʰʲogs*	'side'
གཡག་	g.yag	*gjag*	'yak'
གྱང་	gyang	*gʲaŋ*	'wall'
ཚྭ་	tshwa	*tsʰʷa*	'salt'

Table 4. Examples of words containing ya-btags and wa-zur

5. ORTHOGRAPHIC PECULIARITIES OF OLD TIBETAN TEXTS

This transliteration system would not be complete without taking into account some of the peculiarities of the Old Tibetan orthography. The most important feature is the inverse *gi-gu* vowel, a symbol of unknown phonetic value for which no provision exists in the Wylie system, and which is generally represented as upper-case I. Since its phonetic value is unknown, and since even its status as a genuine phoneme of Old Tibetan is uncertain, we suggest the symbol *ï* for this vowel: ཨི for instance is represented as *?ï*.

The Sanskrit loanwords in Tibetan however, can be transliterated using the standard system of Sanskrit Romanisation, without any attempt at an IPA representation of these forms.

The complex abbreviations used in Tibetan texts (such as ཨོ྅ན་ for ཨོ་རྒྱན་ *o.rg^{j}an* "Oddiyana", Imaeda 2011:40) and the graphical variants (བཀྲ་ for བཀའ་ *bkaf* "word (honorific)") will however not be taken into account in this transliteration, as they are only relevant for philologists working on the spelling of Old Tibetan manuscripts, not for linguists describing the evolution from Old Tibetan to modern dialects or comparing Tibetan to other Sino-Tibetan languages.

6. SYLLABLE BREAKS

In the Tibetan writing system, syllables are consistently separated by the symbol ་, called *ts^{h}eg*. Syllable separation is necessary in the Tibetan script to distinguish the first elements of clusters or the final consonants from independent syllables. Thus <g+s+*ts^{h}eg*> གས་ is read as *gas*, while <g+*ts^{h}eg*+s+*ts^{h}eg*> ག་ས་ is read *ga sa*, and <g+s+r+*ts^{h}eg*> གསར་ is read as *gsar*. Most Tibetologists either separate syllables in the transcription by a space, or use a hyphen, hence writing *ga-sa*. However, this practice is problematic when it comes to glossing texts and breaking words into morphemes: some morpheme breaks occur within the syllable. For instance, the past tense of *sgrug* 'to pick up', *bsgrugs*, must be analysed as:

(1) *b-sgrug-s*
PST-pick.up-PST

Clearly, the use of a colon to separate syllables would be problematic in glossed texts. Using a space however is equally problematic for two reasons. First, many morphemes (for instance *me loŋ* "mirror") are polysyllabic. Also, the syllable boundaries do not always reflect the morpheme boundaries. For instance, the word *p^{h}ru gu* 'child' derives from *p^{h}rug* "child" by addition of the diminutive suffix *-u*, which originates from the noun *bu* "son". The correct morpheme break should be *p^{h}ru g-u*; if spaces are used to separate syllables, one part of the morpheme becomes stranded on the other side of the space.

To avoid these problems, Zeisler (2006) suggested using a space only to reflect genuine word boundaries, and to ignore the syllable boundaries in the transcription. This approach raises a different problem however, as syllabification

is not straightforward in Tibetan. For instance, applying Zeisler's method the word ལག་ཅུབས་ 'glove' would be transcribed as *lagɕubs*. However, a reader unfamiliar with Tibetan (and a computer program even more) would have trouble determining whether this word should be syllabified as *lag+ɕubs* or *la+gɕubs*: both solutions would be equally probable from the point of view of Tibetan phonotactics. For this reason, Zeisler's solution should not be adopted.

Since however we do not use the dot to distinguish between *gj-* and *g-j-*, a relatively straightforward solution offers itself: separating syllables belonging to the same word by a dot. Hence 'child' would be transcribed as *pʰru.g-u* and 'glove' as *lag.ɕubs* in our system. The Tibetan phrase break marker *ཪad* being usually transcribed with a slash / rather than with a dot, there is no risk of ambiguity with punctuation marks either.

The following text example illustrates how the envisioned system would be used in glossing Tibetan texts:[1]

(2) *sbal-ⁿdre* *de* *gɲis* *rdziŋ* *de-r* *ʑug-s-pa-nas*
frog-demon that two pond that-LOC enter-PST-NMLZ-ABL

b-zuŋ / *ʑiŋ.pa-rnams-la* *tɕʰu* *b-kag-pa-s*
PST-seize farmer-PL-DAT water PST-block-NMLZ-ERG

lo *re.bʑin* *sbal-ⁿdre* *de* *gɲis-la* *bu*
year each frog-demon that two-DAT child

gʑon.nu *gɲis* *mtɕʰod* *dgos* /
young.man two offer have.to
'From the moment that the two frog-demons invested the pond, they blocked the farmers' water on them and each year, two young men had to be offered to these two frog-demons.' (Robin and Klu rgyal 2005:86)

Alternatively, the past forms *bzuŋ* and *bkag* can be left unanalysed and glossed as PST:seize and PST:block without internal morpheme breaks.

The only case where such system could potentially lead to ambiguity is if a one-consonant morpheme appears between two other morphemes within a word, especially in the case of the negation and tense prefixes:

(3) *ma-b-lta-s-na*
NEG-PST-look-PST-if
If (I/you/he) do not look.

Glossed in such a way, there is ambiguity as to whether the syllabification should be *mab.lta.sna*, *ma.blta.sna*, *mab.ltas.na* or *ma.bltas.na* (the correct one). There are two solutions to avoid this problem. First, suppressing syllable-internal morpheme boundaries:

[1] Interlinear glosses follow the Leipzig Glossing Rules.

(4) *ma-bltas-na*
NEG-PST:look-if

Second, using distinct symbols for syllable-internal and syllable-external morpheme boundaries; since tautosyllabic verbal markers in Old Tibetan were probably clitics rather than true affixes, I suggest using the equation mark for these:

(5) *ma=b-lta-s=na*
NEG=PST-look-PST=if

7. CONCLUSION

The new transliteration system proposed in this paper does not aim at replacing the sophisticated Romanisation scheme used by philologists such as Imaeda to represent the orthographical peculiarities of Old Tibetan texts. Its purpose is more limited: to provide a self-explanatory system to represent Classical and Old Tibetan data in articles on historical linguistics, readable by linguists unfamiliar with the Tibetan script.

REFERENCES

Coblin, Weldon S. 2002. On certain functions of 'a-chung in early Tibetan transcriptional texts. *Linguistics of the Tibeto-Burman Area*, 25.2: 169-84.

Coblin, Weldon S. 2006. Two Notes on Táng-time Tibetan Transcriptions of Chinese. *Linguistics of the Tibeto-Burman Area*, 29.2: 133-36.

Hill, Nathan 2005. Once more on the letter ཀ. *Linguistics of the Tibeto-Burman Area*, 28.2: 111-141.

Hill, Nathan 2006. Tibetan *vwa* ‘fox’ and the sound change Tibeto-Burman **wa* → Old Tibetan *o*. *Linguistics of the Tibeto-Burman Area*, 29.2: 75-90.

Hill, Nathan 2009. Tibetan <ḥ-> as a plain initial and its place in Old Tibetan Phonology. *Linguistics of the Tibeto-Burman Area*, 32.1: 115-140.

Hill, Nathan 2010. An overview of Old Tibetan synchronic phonology. *Transactions of the Philological Society*, 108.2: 110-125.

Hill, Nathan W. 2012. ‘A note on the history and future of the 'Wylie' system.’ Revue d'Etudes Tibétaines, 23 . pp. 103-105.

Imaeda, Yoshiro 2011. Towards a Comprehensive and Unambiguous Transliteration Scheme of Tibetan, in Yoshiro Imaeda, Matthew Kapstein and Tsuguhito Takeuchi (eds) *New studies of the Old Tibetan Documents: Philology, History and Religion.* Tokyo: Research Institute for Languages and Cultures of Asia and Africa, 39-44.

Sun T.-S. Jackson. 1986. *Aspects of the phonology of Amdo Tibetan: Ndzorge Sháme Xra dialect* (Monumenta Serindica No. 16). Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Robin, Françoise and Klu rgyal tshe ring 2005. *Les contes facétieux du cadavre (mi ro rtse sgrung)*. Paris: l'Asiathèque.

Wylie, Turrell 1959. A Standard System of Tibetan Transcription. *Harvard Journal of Asiatic Studies*, 22, 261-267

Zeisler, Bettina 2006. The Tibetan understanding of karman: Some problems of Tibetan case marking. In: Christopher I. Beckwith (ed.), *Medieval Tibeto-Burman languages II, PIATS 2003: Tibetan studies: Proceedings of the Tenth Seminar of the International Association for Tibetan Studies, Oxford 2003*. Brills Tibetan Studies Library, 10. Leiden: Brill, 57-101.

Author's address:

Guillaume Jacques
CNRS-EHESS, CRLAO
131 Boulevard Saint-Michel,
75005 Paris
France

rgyalrongskad@gmail.com