

Towards a standardisation of Tibetan transliteration for textual studies

Joanna Bialek (Humboldt Universität zu Berlin)

Introduction¹

In the past few decades the so-called ‘Wylie-system of transliteration’² has been more and more commonly used in scholarly literature. It owes its popularity mainly to the fact that it can be written with any Latin keyboard and does not require coding of special signs or adding of diacritics. Accordingly, the Wylie-system’s greatest advantage is that it is easy to operate and comfortable for those not primarily interested in language, be it written Tibetan or modern spoken vernaculars.³ Well that’s nice, but not enough. Its

¹ I would like to acknowledge financial support provided by grant BI 1953/1-1 of Deutsche Forschungsgemeinschaft in years 2017-2020 that enabled me to prepare this paper. I would also like to express my gratitude to Nathan Hill and especially Michael Balk for stimulating discussions on practical and theoretical aspects of transliteration. Furthermore, I would like to thank all those who engaged in the discussion on a previous version of the paper within a session on Academia. The paper uses the following conventions: Latin letters proposed for transliteration of Tibetan letters are enclosed in pointy brackets < >; examples of transliteration are provided in italics; and the IPA transcription in square brackets.

² Put forward in Wylie (1959).

³ Wylie has defined the main criteria for a “standard system of Tibetan transcription (sic)” as: 1. minimal complexity; and 2. capability of reproduction on a standard typewriter (1959: 263). These of course have nothing to do with academic standards required for a transliteration system. It is also not clear whether Wylie knew the difference between ‘transliteration’ and ‘transcription’ and deliberately devised a ‘Tibetan transcription system’, or whether he ignored the distinction. The ‘Wylie transliteration’ first gained in popularity in the 1980s and 1990s. The immediate reason might have been delivered with digital typesetting and the use of text editors by authors. The basic tool was there but in the beginning it lacked the functionality and precision of traditional printing, at least in the range of characters available. These limitations are now mostly overridden by character encoding of Unicode and we can again demand a better transliteration system.

most serious disadvantage is that it has not been based on any serious considerations of the relations between the characters used. For instance, it uses the letter 'h' with three different functions:

- For the 29th letter of the Tibetan alphabet: ཧྲ *h*;
- For 'aspiration': *kh, ch, th, ph, tsh*;
- For 'palatalisation': *zh, sh*.⁴

This 'logic' makes the transliteration useless for any language-related studies. Moreover, the paramount principle of transliteration – 'one letter for one letter' – was not observed by Wylie. The use of an apostrophe for transliterating འ and the lack of transliteration for the last letter of the alphabet, ཨ, have already been criticised in Hill (2012: 103) and Balk (2005: 2).⁵

On the other hand, publications on Tibetan languages make use of very diverse systems. In fact, it seems as if every scholar has been using her own transliteration.⁶ The table in the Appendix presents a selection of transliterations used in publications on Tibetan languages.

It is not the aim of this paper to convert everybody to one transliteration system, or to argue that the system herewith presented is the only correct one. My basic motivation is to suggest general *rules* any transliteration system should accord to in order to be internally coherent and logically structured. In an ideal case, a transliteration system is based on a one-to-one relation between the letters of the source script and the letters of the Latin alphabet. This however is seldom possible alone because many scripts make use of more than 26 letters that are inherent to the Latin alphabet. The result is that one has to either use more than one letter of the Latin alphabet to transliterate the source alphabet or resort to diacritic signs. The latter solution is conceived of as a better standard for a *transliteration*. In devising the transliteration put forward in the paper I attempted to follow these rules.

⁴ This argument has been formulated in Hill (2012: 103).

⁵ Other shortcomings of the 'Wylie system' are discussed in Balk (2005: 1-2).

⁶ Regrettably, even in linguistic studies the 'Wylie transliteration' finds its adherents.

'Librarian Romanisation' vs. 'academic transliteration'⁷

The idea of writing a paper on a new transliteration system for the Tibetan འགྲུབ་མཚན་ script arose from my own work on the origins of the Tibetan script⁸ and on a new textbook for Classical Tibetan⁹. My work in this field gained additional motivation from stimulating exchanges with Michael Balk, long-standing curator (now retired) of the Central Asian Collection of the State Library in Berlin and expert on the Romanisation of non-Latin scripts such as Tibetan and Mongolian. Most importantly, it occurred to me from discussions with Michael Balk that the wish for a unitary system of transliteration for all purposes cannot be fulfilled for the time being. Librarians, for instance, require a system that ensures efficient managing of large amounts of foreign literature that has to be properly catalogued. The system should allow them to ascribe a letter of the Latin script to a letter of a foreign script without the demand of knowing the language of the foreign script, and to interpret not only titles, but also proper names, including toponyms. On the other hand, library users do not always (e.g., in public spaces) have easy access to non-Latin letters or symbols with diacritics that they could enter in their search query. These specific requirements run counter to the needs of academic community, especially of those scholars who work with languages. Tibetologists preparing critical editions or text-linguistic analyses require more precision and, first of all, consistency in using written conventions. Comparative and historical linguists (who use written Tibetan sources but do not necessarily have a good command of the language) expect transparency and intelligibility from the system. Furthermore, a system that is well-motivated and comprehensible can facilitate the learning process for students who wish to acquire skill in reading Classical Tibetan texts. Therefore, it seems essential that 'academic transliteration' be distinguished from 'librarian Romanisation', at least until tools are available that would considerably facilitate the

⁷ I distinguish between 'Romanisation', which is a broader concept, and 'transliteration'. The former is understood as a conversion of one writing system to Latin script without observing the stringent rules of transliteration. According to my definition, 'Wylie transliteration' would be a Romanisation, albeit an infelicitous one.

⁸ See Bialek (Forthcoming a).

⁹ See Bialek (Forthcoming c).

use of Unicode characters in search queries, including Internet search engines.

Transliteration vs. transcription

Theoretically there exist no regulations concerning the issue of which letter of the source alphabet should be given which equivalent in the Latin alphabet. Thus, we can also think of the following transliteration system:

ཀ	ཁ	ག	ང	ཅ	ཆ	ཇ	ཉ	ཏ	ཐ	ད	ན	པ	ཕ	བ	མ	ཙ	ཛ	ཎ	ཤ	ཞ	ཟ	ང
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w
ཡ	ར	ལ	ཤ	ས	ཏ	ཨ		ོ	ི	ུ	ེ	ོ										
x	y	z	A	B	C	D		E	F	G	H	I										

Accordingly, the Tibetan syllable ཅེན་ would be transliterated as *oxHk*. We could also agree on a system in which the same syllable would be transliterated as *mfeD* | (with the vertical line transliterating the *cheg* < ` >). The only true restriction is that the assignment of the letters must be predictable.

A transliteration is *not* a transcription. The latter encodes *sounds* of a language in script.¹⁰ The most accurate system of transcription is the International Phonetic Alphabet (IPA) devised to represent sounds that are part of oral language. As with any written form of communication, the IPA is based on conventions. For instance, that the sound described phonetically as ‘velar voiced stop’ is ascribed the letter < g > and its IPA notation is [g]. The sound could likewise have been ascribed the letter < n >, theoretically. But the convention was established and there is wide agreement on its foundations.

There is a good reason that transliterations have some resemblance to the representation of the sounds in script based on widely accepted conventions – like writing < g > for [g]. This reason is most obvious to those working with written languages like Old Tibetan (OT) or the so-called Classical Tibetan (CT). Of course, one can read texts just by looking at them without speaking aloud, or with the ‘sounds of the letters’ kept ‘in mind’. But as soon as one has to teach

¹⁰ Funnily enough, Wylie calls his system “a standard orthographic *transcription*” (1959: 263; my emphasis).

the language to students, saying that “the pronunciation does not matter” won’t suffice.¹¹ Because there is no way to ascertain the actual pronunciation of Old Tibetan – or, more generally, written Tibetan – so that it could be transcribed in the IPA,¹² it is helpful to create a tool that would support students in their efforts to learn Tibetan. Another, perhaps more important, reason for having one transliteration system is that it can become a platform between the written language of yore (whose orthography was established almost 1400 years ago) and the modern varieties of Tibetan, that all go back to a Central Tibetan language of the first half of the 7th century for which the script was devised (cf. Bialek 2018b).

Sometime between the years 630s and 648, the Tibetan script was invented, and we have every reason to assume that its shape was very close to the script we know from the oldest dated Old Tibetan text: the *Žol* inscription from the year 764.¹³ In the meantime we know that the OT orthography rather faithfully mirrored the pronunciation of the Central Tibetan language at the time of the script invention (cf. Bialek 2018b). Moreover, Nathan Hill has succeeded in reconstructing the Old Tibetan phoneme inventory (2010). Therefore, we are now in a very comfortable position to devise a transliteration system of the Tibetan script that can be roughly based on the recon-

¹¹ Reading old texts in a modern pronunciation is even more problematic, but that’s an issue for a separate discussion.

¹² It is only possible to reconstruct the phonetic values ascribed to the letters at the time of the script invention (see below). This is however not synonymous with being able to transcribe syllables or words of Old Tibetan. We know that in the 7th century dialectal differentiation had already occurred (cf. Bialek 2018b). In fact, from then on we can’t reasonably speak of the ‘pronunciation’ of Old Tibetan, not to mention an oxymoron such as ‘pronunciation of Classical Tibetan’. Any attempt at devising a transcription for any of these languages (like, for instance, the one proposed by Jacques (2012)) must therefore be rejected as ill-founded and misleading. Jacques’ argument that such a transcription would facilitate the work of historical linguists (*ibid.*, p. 95) is likewise delusive. A transcription of a fifteenth century text with a system based on the pronunciation of the early seventh century is founded on a grave misunderstanding. By way of example, later texts contain vocabulary largely unknown in Old Tibetan, part of which was certainly also coined much later. How can one ascribe them a pronunciation from the period in which they did not even exist?

¹³ The recently discovered bell inscription from Dgay-ldan-byin-chen (cf. Lhamchog-rgyal 2011) dates to the reign of Khri Lde-gcug-brcan (704–54) and is certainly older than the *Žol* inscription (cf. Bialek Forthcoming b). However, till now no concrete date could be proposed for its composition.

structured phonetic values ascribed by the Tibetans to the single letters of the alphabet at the time of its invention.

Transliteration system for textual studies

Based on the above general considerations Fig. 1 provides the reconstructed phonetic values (in square brackets) of the corresponding Tibetan letters and the proposed transliteration of the latter (in italics).¹⁴

ཀ	<i>k</i> [k]	ཁ	<i>kh</i> [k ^h]	ག	<i>g</i> [g]	ང	<i>ñ</i> [ŋ]
ཅ	<i>č/ć</i> [tɕ]	ཆ	<i>čh/ćh</i> [tɕ ^h]	ཇ	<i>ǰ/j</i> [dʒ]	ཉ	<i>ñ/n</i> [ɲ]
ཏ	<i>t</i> [t]	ཐ	<i>th</i> [t ^h]	ད	<i>d</i> [d]	ན	<i>n</i> [n]
པ	<i>p</i> [p]	ཕ	<i>ph</i> [p ^h]	བ	<i>b</i> [b]	མ	<i>m</i> [m]
ཅ	<i>c</i> [ts]	ཆ	<i>ch</i> [ts ^h]	ཇ	<i>ǰ/j</i> [dʒ]	ཉ	<i>w</i> [w]
ཞ	<i>ž</i> [z]	ཟ	<i>z</i> [z]	འ	<i>y/h/h̄</i> [ɣ]		
ཡ	<i>y</i> [j]	ར	<i>r</i> [r]	ལ	<i>l</i> [l]		
ཤ	<i>ś</i> [ɕ]	ས	<i>s</i> [s]	ཧ	<i>h</i> [h]	ཨ	<i>q</i> [ʔ]/[Ø]

Fig. 1

A quick look at the table in the Appendix suffices to see that the transliteration proposed here shares many of the 'new' features with

¹⁴ If more than one transliteration is proposed for a Tibetan letter, the first one is preferred. Letters added after slash present an alternative transliteration. For their discussion see below.

the 'old French system' as used by Bacot and the transliteration of the first edition of Hahn's *Lehrbuch* (1971).

My first consideration in devising the transliteration was that the system should be internally coherent and accord with generally acknowledged transliteration standards. Furthermore, it should give the student a rough impression about the pronunciation of the letters at the time of their invention, or at least not be misleading in this regard. In the following I will discuss the most controversial elements of the proposed transliteration.

ཅ — ཆ — ཇ — ཉ — འ — ཡ

Because most of the previous systems that used diacritics transliterated letters འ and ཡ as < ž > and < ś >¹⁵, one could propose using the acute accent < ´ > (U+0301) to mark the quality of palatalisation in general. This would yield the following equivalents for all the letters that represented palatal sounds:

ཅ < č > ཆ < čh > ཇ < j > ཉ < ě >¹⁶ འ < ž > ཡ < ś >¹⁷

This transliteration is reinforced by the fact that the acute accent, when used with consonantal letters, unanimously represents alveolo-palatals in various orthographies (cf. https://en.wikipedia.org/wiki/Acute_accent; accessed 29.10.2019). This solution, however, has one serious flaw: it has to use the combination of < j > (U+0237) and the acute accent < ´ > (U+0301) in < j´ > to transliterate ཇ. Therefore, instead of the acute accent one could opt for the caron < ˇ > (U+02C7) that was sometimes applied for transliteration of ཅ, ཆ, and ཇ:

ཅ < č̇ > ཆ < č̇h > ཇ < j̇ > ཉ < ě̇ > འ < ž̇ > ཡ < ṥ >

However, a closer examination has yielded that the caron diacritic is generally used to represent retroflex or palatalo-alveolar sounds in world orthographies (cf. <https://en.wikipedia.org/wiki/Caron>; accessed 29.10.2019).¹⁸ Moreover, the letters < ž̇ > and < ṥ > have only

¹⁵ Only Zeisler consequently uses caron to transliterate the letters (see Appendix). Curiously, Bacot (1946-8: 9) transliterated འ with a caron but ཡ with an acute accent (see Appendix).

¹⁶ Instead of the more popular < ě̇ >, Beckwith used < ě̇ > (see Appendix).

¹⁷ It came as surprise to me that the same transliteration was already used in Kaschewsky (1987). The only exception is his < ě̇ > for ཉ.

¹⁸ By the way, Old Tibetan did not have palatalo-alveolar consonants therefore the frequently encountered transcriptions [ʃ], [ʒ] and others based on them are inaccurate. This is also obvious from the dialectal material collected in CDTD. [ʃ] and

seldom been used, whereas < ñ > is completely absent from previous transliterations. In order to keep in with the established conventions the following combination can be proposed:

ཅ < č > ས < čh > ཇ < j > ཉ < ñ > འ < z > ཡ < ś >

ཉ, འ, and ཡ are transliterated as in the majority of previous transliteration systems, whereas ཅ, ས, and ཇ acquire Latin equivalents that allow to relate them more easily to another series of affricates, namely č, čh, and j (see below). I deem this transliteration a temporary solution until a Unicode character for < j > with the acute accent has been developed (which, however, may never happen...).

č — čh — j

Old Tibetan had two sets of affricates: alveolo-palatal and alveolar. Because letters of all alveolo-palatals are transliterated either with an acute accent or with a caron, it follows that the letters representing alveolar affricates can be transliterated by the same Latin letters minus the diacritics. Therefore, the transliteration of all the affricates would be:

ཅ < č > ས < čh > ཇ < j > ཇ < c > ས < ch > ཇ < j >¹⁹

Compare hereto the sibilant sets:

འ < z > ~ ར < z >

and

ཡ < ś > ~ ལ < s >²⁰

[3] attested in WAT and SMu are areal features and resulted from direct and long-standing contact with languages like Urdu or Hindi that don't have alveolo-palatals but only palatalo-alveolars (cf. Bialek Forthcoming a).

¹⁹ If caron is used for transliterating the 'alveolo-palatal letters', then one can adopt the common < j > (U+006A) for ཇ.

²⁰ Alternatively, one could propose the following transliterations: ཅ < ts >, ས < tsh >, ཇ < dz >, ཇ < ts >, ས < tsh >, ཇ < dz >. It has two advantages over the first option: its letters are coded in Unicode (i.e. there is no need to add special diacritics) and, depending on one's linguistic background, it may be more intuitively pronounced without confusion. Its disadvantage is that it uses up to three letters (< tsh >, < tsh >). More importantly, letters ཅ, ས, and ཇ have virtually never been transliterated with < ts >, < tsh >, and < dz > or similar, whereas for the transliteration of č, čh, and j with < c/č >, < ch/čh >, and < j/j > there is a precedence case in the 'old French tradition' and Hahn (1971). The only author known to me who used ཅ < ts >, ས < tsh >, ཇ < dz >, is Richter (see Appendix). Likewise, Beyer used the latter letter combination (1993: 66) but his system cannot be considered a

ꠄ

The greatest bone of contention among Tibetologists has been the transliteration of the letter ꠄ. Most frequently one decided either for < h > with a diacritic or for various forms of 'apostrophe'. Nathan Hill convincingly reconstructed the phonetic value represented by the letter as voiced velar fricative [ɣ] (2009).²¹ This reconstruction makes the use of the letter < h > for transliterating ꠄ problematic, because < h > is unanimously associated with voiceless sounds in various orthographies. Thus, < h >, < h̃ >, < h̄ >, < h̅ >, < h̆ >, < ḣ >, < ḧ >, < h̉ >, or < h̊ >, if used at all in linguistic contexts, always mark a voiceless sound.²² Chinese scholars since Yu Daoquan (于道泉) use the letter < v > for transliterating ꠄ. This convention was also adapted by Kolmaš in his cataloguing work (<http://katalog.orient.cas.cz/tibet/tibet.htm>; Michael Balk, p.c. 24.10.2019 & 03.11.2019). The only motivation for choosing < v > seems to have been the fact that < v > was otherwise not assigned to any Tibetan letter. Theoretically, one could have chosen < f > or < x > or < q > instead. The choice of < v > has no phonetic rationale but, as mentioned before, a transliteration system does not need to account for phonetics. The assignment of < v > to ꠄ has however unpleasant consequences for didactics; students learn that ꠄ was originally devised for a voiced velar fricative [ɣ] (realised as such even today in some dialects, cf. Hill 2009: 117ff.), but for unknown reasons it is represented in the transliteration as < v > which usually stands for bilabial or labiodental fricatives [ɸ], [ɸ], [v]. There is not the least concurrence between the Tibetan letter and its Latin representation. Apart from the < h >-letters, 'apostrophes', and < v > no other Latin character has ever been proposed for ꠄ. I put forward three options for a new transliteration of ꠄ, discussing pros and cons for each:

1. < ɣ > (U+0263) is Latin letter gamma based originally on the Greek gamma < γ > (U+03B3) and used in the IPA to represent voiced velar fricative [ɣ].²³ I started using this transliteration in

transliteration.

²¹ This reconstruction is confirmed in my forthcoming study (cf. Bialek Forthcoming a).

²² The use of < h̄ > may also lead to confusion when Tibetan transliterations/transcriptions of originally Sanskrit words are being transliterated.

²³ The letter < ɣ > is used in various orthographies (most commonly of African languages) to represent [ɣ].

my publications some time ago. In a private communication (email 26.10.2019), Nathan Hill remarked that one should avoid using the same symbols in transliteration and transcription. I agree with this argument, in general, although one notices that the transliteration of the majority of Tibetan letters is done by means of characters that are likewise used in transcription, cf. ཀ < k > [k], ད < d > [d], or མ < m > [m], etc.

2. < ġ > (U+0121) or < ĝ > (U+011F). In an ongoing study I have reconstructed the origins of the Tibetan letter ར as going back to an Indian g with the same diacritic as the one used to disambiguate alveolar affricates from alveolo-palatal ones: ར̣.²⁴ The two reconstructions (of the phonetic value [ɣ] and of the shape based on an Indian letter g) would speak for the transliteration of ར with a < g > letter + a diacritic. The overdot in the transliteration would relate < ġ > to another velar: ས < ñ >. In writing systems of the world, < ġ > is preponderantly used to transliterate letters that represent voiced sounds. With this proposal the overdot would be associated in the transliteration of the Tibetan script with the quality of ‘velarity’: ར < ġ >, ས < ñ >²⁵. However, the unpleasant consequence of the use of a transliteration based on the letter < g > would be forms like ġgaġ or ĝgaĝ for རགའ. Any combination ར + ཀ (and these are many) would yield the ‘g-cluster’ ġg-/ĝg- in onset. This is aesthetically not a good solution and might be difficult for students to handle.
3. < h > with a diacritic might be a compromise if one does not want to use < ɣ > due to its application in the IPA. As I said, it is not an optimal solution because any letter < h > is commonly associated with voiceless sounds and ར represented a voiced sound. On the other hand, both letters are associated with fricative values. Because < ɦ > is used to transliterate Sanskrit *visarga*, a distinct diacritic should be used for ར in order to avoid

²⁴ In Bialek (Forthcoming a) I provide paleographical evidence from Indian inscriptions and early Tibetan writings which supports the hypothesis that was formulated for the first time by Francke (1912: 270). The relics of the original shape of ར can be encountered in OT texts in which ར is added a small hook on the top.

²⁵ The association of the overdot with the velar quality would necessitate the question about the use of < j > for ས. To preserve the coherence of the system it would be advisable to use < j > (U+0237) instead. On the other hand, < j > would be a more user-friendly option.

confusion when transliterating Tibetan transliterations/transcriptions of Indian words. In previously proposed transliterations two other diacritics occurred: a macron below < ħ > (U+1E96) and a breve < ħ > (U+1E2B). One can remark however that the choice of < h > with whatever diacritic would result in clusters with two *h*'s that might be difficult to handle by students: *ħkh-*, *ħčh-*, *ħth-*, *ħph-*, and *ħch-*.

None of the suggested transliterations seems flawless. The transliterations with < g > and < h > could cause problems in processing of certain clusters for students learning written Tibetan. I may only express my wish that new considerations or arguments will be put forward in favour of one of the characters in later studies.

ཧ

The letter < h > is used for transliterating the 29th letter of the Tibetan alphabet and as an additional grapheme of letters that represented aspirated sounds. The sound value represented by the letter ཧ has been reconstructed as voiceless glottal fricative [h] (Hill 2010) and so its transliteration as < h > is likewise appropriate for the quality of aspiration.

ཨ

The final controversial issue concerns the last letter of the Tibetan alphabet: ཨ. Previous transliteration systems were usually unanimous in ignoring it. That is, they treated the letter as non-existent and transliterated only the vowel value associated with it, e.g. ཨ *a*, ཨི *i* etc.²⁶ The phonetic value associated originally with the letter could not yet be conclusively reconstructed; the letter might have been a mere place holder for marking a vocalic onset or represented glottal stop [ʔ]. In a recent publication Nathan Hill proposed transliterating ཨ with < q > (Hill 2019: 6). It is a good practice to transliterate every letter of a writing system with a letter of the Latin alphabet. Whatever its original phonetic value might have been, we cannot deny that the letter ཨ does exist in the Tibetan alphabet and seems to have been there from the very beginning. If one strives not to introduce non-Latin letters into a transliteration and to keep the system distinct

²⁶ For a more detailed discussion of the issue, see Balk 2005: 2-3.

from a transcription (e.g., < ? >) then < q > seems to be the least irritating choice. The comparison with other writing systems shows that the letter is usually associated with voiceless uvular or velar stops (although voiced equivalents are also represented; cf. https://en.wikipedia.org/wiki/Q#Use_in_writing_systems; accessed 29.10. 2019). Accordingly, I propose accepting Hill's transliteration of ལ as < q >.

Additional letters. For transliterating 'reversed' letters ར̣ < ṭ >, ཐ̣ < tḥ >, ང̣ < ḍ >, ཡ̣ < dḥ >, ས̣ < ṅ >, and ས̣ < ṣ > the underdot is used. The diacritic then represents the quality of 'reflexivity' and the transliteration makes such words resemble their Sanskrit origins more faithfully. Capital letters, suggested instead by Imaeda (2011: 42), unnecessarily blur the picture; a danger of misinterpretation might occur when proper names are capitalised as well. In Old Tibetan another form of the vowel sign for [i] was used: ི. As I argue in Bialek (Forthcoming a), this sign originally represented the lengthened [i:] and therefore I suggest to transliterate it as < ī > (by analogy with the use of the macron in transliterations of Indian alphabets).

In conclusion, the proposed transliteration exclusively uses Latin letters, consistently adding the following markers to represent particular qualities:

- < h > for aspiration: < kh >, < čh >, < th >, < ph >, < ch >;
- acute accent < ˊ > (U+0301), caron < ˇ > (U+02C7), and tilde < ˜ > (U+0303) for palatalisation: < č/ć >, < čh/čh >, < j̄/j̄ >, < ñ >, < ž >, < ś >;²⁷
- overdot < ˙ > (U+0307) for velarity: < ṅ >;
- underdot < ̣ > (U+0323) for retroflexity: < ṭ >, < tḥ >, < ḍ >, < dḥ >, < ṅ >, < ṣ >;
- macron < ̄ > (U+0304) for vowel length: < ī >;
- (for the lack of a better solution) macron below < ̅ > (U+0331) or breve below < ̆ > (U+1E2B) for the velar fricative ʁ.

These are of course traits of phonetics which do not need to be considered in a transliteration system. However, I think it is important to use diacritics that are not usually associated with other qualities. I

²⁷ With the introduction of the precomposed character < j > in Unicode the caron and the tilde could be replaced by the acute accent.

also mentioned the possibility of using double and triple letters for affricates (< ts >, < dz > etc.). This would add more transparency for historical linguists, as the relations between, e.g., < ts > [tʂ] and < ś > [ɕ] or < dz > [dʂ] and < ź > [ʐ] would be straightforwardly marked in the transliteration. But this is a function of a transcription and therefore I decided against this option as it violates the rule 'one letter for one letter'.

If we aspire to call something a 'system' then it has to be internally coherent and logically structured. I dare to state that the above transliteration fulfils these criteria. Moreover, it is based on the first principle in devising transliteration systems: one letter for one letter. The exception is made only for Tibetan letters that represented aspirated sounds. The remaining letters are expressed by simple symbols or a symbol plus a diacritic.

Punctuation marks

Tibetan script makes use of three basic punctuation marks: ཚྲིག་ < ' >, འདྲི་ < | >, ཡིག་མགོ་ < ། > (sometimes also called དབྱ་ or དབྱ་ལྗོངས་) and manifold combinations thereof. There exist almost unlimited ornamental variants of the basic signs but they have little or no relevance to the text discourse and therefore can be omitted from the discussion.

Because each sign of the original script should have its representation in the transliteration, a consensus has been reached to transliterate < ' > with a space (U+0020) and < | > with a slash < / > (U+002F).²⁸ In Bialek (Forthcoming a) I propose transliterating ཡིག་མགོ་ as section sign < § > (U+00A7). I argue that the ཡིག་མགོ་ and the section

²⁸ Beckwith, Walter (e.g., in 2010 & 2015), and Zeisler (e.g., in 2011) have been using the inconsistent transliteration of a ཚྲིག་ once as a space, once as Ø, joining the neighbouring syllables into 'one word'. This method was also followed, although not consequently and using the Wylie 'transliteration', by Tournadre in (2010). I agree with Jacques (2012: 93–4) that, apart from producing strings of letters difficult to disambiguate, this approach fuses transliteration with transcription and should not be followed in a pure transliteration system. Due to the unsettled character of the Old Tibetan orthography, diplomatic transliteration of an Old Tibetan text may sometimes deviate from the one used here. If necessary, one can use the following signs: ཚྲིག་ = interpunct < · > (U+00B7), 'double ཚྲིག་' = colon < : > (U+003A), lack of ཚྲིག་ = space (U+0020). Richter suggested an interpunct as an alternative transliteration of a ཚྲིག་ (1964b: 177), which is also an option worth considering for Classical Tibetan texts.

sign have parallel functions in the writing systems of Tibetan and Latin: i.e. they introduce a new section of a text (or at least this seems to have been the original function of ཡིག་མཐོག་ in OT).²⁹

Conclusions

The transliteration system proposed in the paper should not be understood as an ultimate solution. Rather it is intended as an invitation to a discussion in which other alternatives can be introduced and considered. It is conceivable that more than one Romanisation system for dealing with written Tibetan is actually necessary. Depending on the purpose of the Romanisation and the target group one can think of three independent systems:

- Professional transliteration for textual and text-linguistic studies (like the one proposed here);
- Less rigorous and devoid of diacritics system for non-textual studies, catalogues, and (web) search engines similar to ‘Wylie’ but adding handleable transliterations for འ and ཨ. As an example, the ABEC system developed for the online Old Tibetan Dictionary can be quoted.
- IPA-based transcription for linguistic studies and historical reconstructions.

The paper presents tentative proposals and is certainly not comprehensive in terms of the problems discussed.³⁰ However, I deem it important that a few basic rules are observed when devising a new transliteration system for textual and text-linguistic studies:

- Use of one letter for one letter (plus diacritics);³¹

²⁹ As kindly remarked by Nicola Bajetta (session on Academia; November 2019), the section sign < § > was used by Orofino to transliterate *sbrul śad* (Orofino 2007: 99).

³⁰ I think it is necessary to first agree on the basic elements of the system before we extend it with the purpose of including elements that are used only, e.g., in Tibetan transcriptions/transliterations of other writing systems or in intricate ‘clippings’ like ཨྲྀ.

³¹ An exception can be made for Tibetan letters that represent aspirated sounds. The use of an apostrophe instead (e.g., འཇམ་ གཙམ་) would blur the transliteration, making single letters indistinguishable from each other, and so unnecessarily

- Use of Latin letters;
- Coherence;
- Consistency;
- Avoidance of misleading symbols.³²

The advantage of living in digital times is that we have unlimited access to data on all documented languages and orthographies of the world. We can make use of the data in devising a tool that will meet scholarly standards of a transliteration system and be used by a wider community than the previous scholarly transliterations could reach.

Abbreviations

CDTD	Bielmeier et al. (see References)
CT	Classical Tibetan
IPA	International Phonetic Alphabet
OT	Old Tibetan
SMu	Southern Mustang
U	Unicode
WAT	Western Archaic Tibetan
WTS	Franke et al. 2005– (see References)

References

- Bacot, Jacques. 1946–8. *Grammaire du Tibétain littéraire*. 2 vols. Paris: Librairie d'Amérique et d'Orient. Reprint, 1981.
- Balk, Michael. 2005. "On Letters, Words, and Syllables. Transliteration and Romanization of the Tibetan Script." https://staatsbibliothek-berlin.de/fileadmin/user_upload/zentrale_Seiten/ostasien_abteilung/bilder/Letters__Words_and_Syllables.pdf (accessed 02.11.2019).

complicate the reading for students.

³² A slightly different set of criteria was proposed by Richter: "1. Jedem tibetischen Schriftzeichen (Buchstaben) soll nach Möglichkeit nur ein Transliterationszeichen entsprechen; 2. Die Transliterationszeichen müssen: a. möglichst genau, b. gut unterscheidbar, c. einprägsam und d. ohne Druckschwierigkeiten wiederzugeben sein." (1964b: 172f.).

- Beckwith, Christopher I. 1993. *The Tibetan Empire in Central Asia: A History of the Struggle for Great Power among Tibetans, Turks, Arabs, and Chinese during the Early Middle Ages*. Princeton: Princeton University Press.
- Beckwith, Christopher I., and Michael L. Walter. 2015. "Dating and characterization of the Old Tibetan Annals and the Chronicle." In *From Bhakti to Bon. Festschrift for Per Kværne*, edited by Hanna Havnevik and Charles Ramble, 53–88. Oslo: Novus Press.
- Beyer, Stephan V. 1993. *The Classical Tibetan Language*. Delhi: Sri Satguru Publications.
- Bialek, Joanna. 2018a. *Compounds and Compounding in Old Tibetan. A Corpus Based Approach*. 2 vols. Marburg: Indica et Tibetica.
- Bialek, Joanna. 2018b. "The Proto-Tibetan clusters sL- and sR- and the periodisation of Old Tibetan." *Himalayan Linguistics* 17 (2): 1–50. doi: <https://doi.org/10.5070/H917238831>.
- Bialek, Joanna. Forthcoming a. *Old Tibetan Annals. A Comprehensive Text Grammar*.
- Bialek, Joanna. Forthcoming b. "Kinship terms: New indicators for dating the Old Tibetan documents."
- Bialek, Joanna. Forthcoming c (2020). *Classical Tibetan. A Textbook*.
- Bielmeier, Roland, Felix Haller, Katrin Häsler, Brigitte Huber, and Marianne Volkart, eds. 2013 (draft). *Comparative dictionary of Tibetan dialects*.
- Francke, August Hermann. 1912. "The Tibetan Alphabet." *Epigraphia Indica* 11: 266–72.
- Franke, Herbert, Jens-Uwe Hartmann, and Thomas O. Höllmann, eds. 2005–. *Wörterbuch der tibetischen Schriftsprache*. München: Verlag der Bayerischen Akademie der Wissenschaften.
- Hahn, Michael. 1971. *Lehrbuch der klassischen tibetischen Schriftsprache*. Hamburg: Michael Hahn.
- Hahn, Michael. 1996. *Lehrbuch der klassischen tibetischen Schriftsprache*. Swisstal-Odendorf: Indica et Tibetica.
- Hill, Nathan W. 2009. "Tibetan <h> as a plain initial and its place in Old Tibetan phonology." *Linguistics of the Tibeto-Burman Area* 32 (1): 115–40.
- Hill, Nathan W. 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'Études Tibétaines* 23: 103–5.

- Hill, Nathan W. 2019. *The Historical Phonology of Tibetan, Burmese, and Chinese*. Cambridge: Cambridge University Press.
- Imaeda, Yoshiro. 2011. "Towards a comprehensive and unambiguous transliteration scheme of Tibetan." In *New Studies of the Old Tibetan Documents: Philology, History and Religion*, edited by Yoshiro Imaeda, Matthew Kapstein and Tsuguhito Takeuchi, 39–43. Tokyo: Tokyo University of Foreign Studies.
- Jacques, Guillaume. 2012. "A new transcription system for Old and Classical Tibetan." *Linguistics of the Tibeto-Burman Area* 35 (2): 89–96.
- Kaschewsky, Rudolf, and Pema Tsering. 1987. *Die Eroberung der Burg von Sum-pa. Aus dem tibetischen Gesar-Epos*. 2 vols. Wiesbaden: Otto Harrassowitz.
- Lha-mchog-rgyal. 2011. "Bcan po khri lde gcug brcan skabs kyi ğag roñ dğay ldan byin ċhen gcug lag khañ gi dril buyi kha byañ gi yi ger dpyad pa." *Bod ljonš žib yjug* 118 (1): 1–9.
- Orofino, Giacomella. 2007. "From archaeological discovery to text analysis: The Khor chags monastery findings and the *Mañjuśrīnāmasaṁgīti* fragment." In *Discoveries in Western Tibet and the Western Himalayas: Essays on History, Literature, Archaeology and Art*, edited by Amy Heller and Giacomella Orofino, 85–127. Leiden: Brill.
- Richter, Eberhardt. 1964a. *Grundlagen des Lhasa-Dialektes*. Berlin: Akademie-Verlag.
- Richter, Eberhardt. 1964b. "Zum Problem der Schaffung einer einheitlichen Umschrift (Transliteration und Transkription) des Tibetischen." *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 114: 171–9.
- Schwieger, Peter. 2006. *Handbuch zur Grammatik der klassischen tibetischen Schriftsprache*. Halle: IITBS, International Institute for Tibetan and Buddhist Studies.
- Sommerschuh, Christine. 2008. *Einführung in die tibetische Schriftsprache. Lehrbuch für den Unterricht und das vertiefende Selbststudium*. Norderstedt: Books on Demand.
- Tournadre, Nicholas, and Sangda Dorje. 1998. *Manuel de Tibétain Standard, langue et civilisation: Bodkyi spyiskad slobdeb*. Paris: Mondes et Langue.
- Tournadre, Nicolas. 2010. "The Classical Tibetan cases and their transcategoriality: From sacred grammar to modern linguistics." *Himalayan Linguistics* 9 (2): 87–125.

- Walter, Michael, and Christopher I. Beckwith. 2010. "The Dating and Interpretation of the Old Tibetan Inscriptions." *Central Asiatic Journal* 54 (2): 291-319.
- Wylie, Turrell. 1959. "A standard system of Tibetan transcription." *Harvard Journal of Asiatic Studies* 22: 261-7.
- Zeisler, Bettina. 2004. *Relative Tense and Aspectual Values in Tibetan Languages: A Comparative Study*. Berlin: Mouton de Gruyter.
- Zeisler, Bettina. 2011. "For love of the word: a new translation of Pt 1287, the Old Tibetan Chronicle, chapter I." In *New studies of the Old Tibetan Documents: philology, history and religion*, edited by Yoshiro Imaeda, Matthew T. Kapstein and Tsuguhito Takeuchi, 97-213. Tokyo: Tokyo University of Foreign Studies.

Internet sources

Academia: <https://www.academia.edu/>

Tibetská knihovna: <http://katalog.orient.cas.cz/tibet/tibet.htm>

Wikipedia: https://en.wikipedia.org/wiki/Main_Page

Old Tibetan Dictionary: <http://otdict.com/>

Appendix

The table presents a selection of transliterations used in publications on Tibetan languages. The shadowed rows of the table mark the transliterations on which there is little or no agreement among scholars.

	Bacot 1946-8: 9	Wylie 1959:267 ³³	Richter 1964a: 7	Hahn 1971: 1	Beckwith 1993: xiii	Hahn 1996: 1 ³⁴	Zeisler 2004: 223	Jacques 2012:90 ³⁵	Bialek 2018b	Hill 2019: 6
ཀ	k	k	k	k	k	k	k	k	k	k
ཁ	kh	kh	kh	kh	kh	kh	kh	k ^h	kh	kh
ག	g	g	g	g	g	g	g	g	g	g
ང	ñ	ng	ñ	ñ	ñ	ñ	ñ ³⁶	ŋ	ñ	ñ
ཅ	č	c	tš	c	c	c	c	te	c	č
ཆ	čh	ch	tšh	ch	ch	ch	ch	te ^h	ch	čh
ཇ	j	j	dž	j	j	j	j	ɟe	j	j
ཉ	ñ	ny	ñ	ñ	ñ	ñ	ñ	ɲ	ñ	ñ
ཏ	t	t	t	t	t	t	t	t	t	t
ཐ	th	th	th	th	th	th	th	t ^h	th	th
ད	d	d	d	d	d	d	d	d	d	d
ན	n	n	n	n	n	n	n	n	n	n
པ	p	p	p	p	p	p	p	p	p	p
ཕ	ph	ph	ph	ph	ph	ph	ph	p ^h	ph	ph
བ	b	b	b	b	b	b	b	b	b	b
མ	m	m	m	m	m	m	m	m	m	m
ཚ	c	ts	ts	č	ts	ts	ts	ts	ts	ts
ཛ	ch	tsh	tsh	čh	tsh	tsh	tsh	ts ^h	tsh	tsh
ཞ	j	dz	dz	j	dz	dz	dz	ɟe	dz	dz
ཟ	v	w	w	w	w	v (w)	w	w/ɦ ^w	w	w
འ	ž	zh	ž	ž	ž	ž	ž	z	ž	ž
ཨ	z	z	z	z	z	z	z	z	z	z
ཨ	,	,	,	h	,	,	h	ɦ/ ⁿ	ɣ	h
ཨ	y	y	j	y	y	y	y	j/j	y	y ³⁷
ཨ	r	r	r	r	r	r	r	r	r	r
ཨ	l	l	l	l	l	l	l	l	l	l
ཨ	ś	sh	š	ś	ś	ś	š	e	ś	ś
ཨ	s	s	s	s	s	s	s	s	s	s
ཨ	h	h	h	h	h	h	h	h	h	h
ཨ	∅	∅	∅	∅	,	∅	∅	?	∅	q



³³ The 'Wylie transliteration system', apart from not being a transliteration, has in fact been entirely borrowed from Nebesky-Wojkowitz (Wylie 1959: 267). Also, recent language-oriented publications like Tournadre and Dorje (1998: 32), Schwieger (2006: 21f.), and Sommerschuh (2008: 5) use this 'transliteration'.

³⁴ In the fifth edition of his textbook (1995) Hahn first changed the transliteration. The change concerned the transliteration of the alveolar affricate series and the letter ཅ. The new transliteration has been followed in WTS (see fasc. 1, p. xx), CDTD, and Bialek (2018a: 52).

³⁵ The system proposed by Jacques is basically a transcription and as such cannot be compared with the remaining systems, but the author himself calls it "transliteration" (ibid., pp. 90, 93 (twice), 95). Similarly, the system used in Beyer (1993) is a mixture of transliteration and transcription and for this reason has been omitted from the table.

³⁶ In more recent publications Zeisler replaced *ñ* with *ŋ*; so, e.g., in Zeisler (2011).

³⁷ In Hill (2019: 6, Fig. 1.1) there is a typographical error: ཨ is transliterated as < j >, although in the book < y > is used.