Tibetan lexicography

Edward Garrett, Nathan W. Hill (SOAS, University of London)

Lexical Characteristics of Tibetan

Most researchers see Tibetan as a member of a language family which also includes Burmese and Chinese; this family is known by names including 'Tibeto-Burman', 'Sino-Tibetan', and 'Trans-Himalayan', of which the last is the most neutral and accurate (cf. van Driem, 2012). In 650 Tibetan was reduced to writing as an administrative exigency of running the Tibetan empire; the earliest extant documents date from a century later (Hill, 2010b, pp. 110-112). Tibetan linguistic history is conventionally divided between Old Tibetan (eleventh century and earlier) and Classical Tibetan (later texts). Tibetan boasts a vast literature with a wide variety of genres, and the family of Tibetan languages spoken today is comparable in size and diversity to the Romance languages (Tournadre, 2008, pp. 282-283).

Old Tibetan did not have tone and the tonal systems of those modern Tibetic languages that posses them derive transparently from segmental phonology. Tibetan has agglutinative morphology and ergative alignment (Tournadre, 1996); it exhibits *Gruppenflexion*, with ten morphological cases (cf. Hill, 2012). Tibetan lacks any agreement systems, but verbal suffixes indicate switch reference (Andersen, 1987; Zadoks, 2000; 2002; Haller, 2009). Tibetan verbal inflection is complex, with four verb stems showing a variety of ablaut, stem alternation, prefixes and suffixes (e.g. present *ḥdzin*, past *bzuṅ*, future *gzuṅ*, imperative *zuṅs* 'take').

Tibetan has its own alphabetic order, which serves as the organizational principal for all Tibetan dictionaries. The Tibetan alphabet distinguishes 30 consonants (k, kh, g, ṅ, c, ch, j, ñ, t, th, d, n, p, ph, b, m, ts, tsh, dz, w, ź, z, ḥ, y, r, l, ś, s, h, ʔ) and five vowels ([a], i, u, e, o); the alphabet is a good, but not perfect match to Old Tibetan phonology (cf. Hill, 2010b). Alphabetization is complex; letters are arranged both vertically and horizontally, and a word is not necessarily alphabetized by either the left-more or the uppermost letter in a syllable. A syllable has a graphic structure that may be represented $C_2C_3C_1G_4V_5C_6C_7$, of which the sequence $C_3C_1G_4V_5$ is represented vertically, e.g. བསྒྲུབས་ $b_2(s_3g_1r_4u_5)b_6s_7$. In terms of alphabetization, a dictionary entry is placed according to its first syllable; the first syllable is placed in a relevant section according to C_1 and within this section it is placed in a relevant subsection according to C_2 and within this subsection it is placed in a sub-subsection according to C_3, etc. The absence of an element in the relevant position precedes all possible letters in order, as if there were a null consonant first among consonants. In the case of vowels, the absence of vowel marking is interpreted as /a/. These abstract principles lead in practice to a relative alphabetical order such as *ka, kun, kyaṅ, kyi, klu, klub, dkag, dkor, dkyu, bkaṅ, rked, skad, skyon, bskaṅ, bskyuṅs*, all of which occur before any syllable build with 'kh' as C_1. The reader with time on his hands will be able to confirm that this list is correctly ordered given the order of the 30 consonants and five vowels. The task is made a bit easier by presenting all null consonants and the numbers for syllable position: $Ø_2Ø_3k_1Ø_4a_5Ø_6Ø_7$, $Ø_2Ø_3k_1Ø_4u_5n_6Ø_7$, $Ø_2Ø_3k_1y_4a_5\dot{n}_6Ø_7$, $Ø_2Ø_3k_1y_4i_5Ø_6Ø_7$, $Ø_2Ø_3k_1l_4u_5Ø_6Ø_7$, $Ø_2Ø_3k_1l_4u_5b_6Ø_7$, $d_2Ø_3k_1Ø_4a_5g_6Ø_7$, $d_2Ø_3k_1Ø_4o_5r_6Ø_7$, $d_2Ø_3k_1y_4u_5Ø_6Ø_7$, $b_2Ø_3k_1Ø_4a_5\dot{n}_6Ø_7$, $Ø_2r_3k_1Ø_4e_5d_6Ø_7$, $Ø_2s_3k_1Ø_4a_5d_6Ø_7$, $Ø_2s_3k_1y_4o_5n_6Ø_7$, $b_2s_3k_1Ø_4a_5\dot{n}_6Ø_7$, $b_2s_3k_1y_4u_5\dot{n}_6s_7$.

Since digital Tibetan text is now preferentially encoded as Unicode, it is desirable to sort Tibetan in conformance with the requirements of the Unicode Standard. To this end, the Unicode Collation Algorithm (UCA) should be employed (http://unicode.org/reports/tr10/). On its own, the UCA supplied Default Unicode Collation Element Table (DUCET) will not sort Tibetan words correctly. However, language-specific collation elements, that is, clusterings of one or more Unicode characters to be treated as single items for the purpose of determining sort weight, can be defined and included in customised collation rules which specify those cases where the sort order for a language differs from the default (http://www.unicode.org/reports/tr35/tr35-collation.html#Rules). Using this approach, Pema Geyleg and Robert Chilton devised a collation rule set for Dzongkha, a language which shares the same script and sort order as Tibetan. Chris Tomlinson's open-soure implementation of Tibetan sorting (https://github.com/tibetan-nlp/sorting-and-conversion), which is based on the International Components for Unicode for Java (ICU4J), exploits this rule set in order to correctly sort Tibetan text.

Tibetan syllables are distinguished with explicit punctuation, but word breaks are not overtly marked. The limitation of onset clusters to word initial syllables provides a possible definition for a language specific phonemic word; most lexemes so defined would be disyllabic. However, the lexemes that head noun phrases and function as syntactic constituents in a sentence, i.e. syntactic words, are often much longer. Because of the lack of explicit word delimitation, dictionaries normally include entries that consist of anything from individual bound morphemes up to entire phrases or conventional expressions without distinction. The absence of explicit word breaking creates at least two hurdles for Tibetan NLP. First, some word breaking must be imposed on the data, both an intellectual and a practical challenge. Second, however one defines a word, a page break may bisect a word. Thus, the use of a page-driven structure in electronic texts poses a challenge to the explict encoding of word breaks. The analysis of Tibetan part-of-speech categories has scarcely begun and no Tibetan dictionary gives a part-of-speech label to each of its entries. For the treatment of word breaking and the analysis of part-of-speech categories in the project 'Tibetan in Digital Communication', the first project to publicly release a part-of-speech tagged Tibetan corpus, see Hill & Garrett, 2017a.

History of Tibetan Lexicography

Methodologies of dictionary compilation divide heuristically into three types. First, some dictionaries lack explicit methodology and assemble words in an *ad hoc* manner. Second, there are dictionaries that are compiled over very long periods of time on the basis of collections of slips recording attestations of words as used in context. Third, more recent dictionaries are compiled on the basis of electronic text corpora. These methods may be called respectively the 'informal method', the 'traditional method', and the 'modern method'. The overwhelming majority of Tibetan dictionaries were compiled with the informal method. Only a very few Tibetan dictionaries use the traditional methodology. No Tibetan dictionary yet compiled makes use of the modern method.

In the land of snows lexicography enjoys an august history. After the official conversion of Tibet to Buddhism circa 779, the imperium found it useful to standardize terminology to facilitate the translation of Buddhist works, mainly in Sanskrit, into Tibetan. Three lexicographical works assisted this translation work: the *Bye brag tu rtogs byed chen po* (*Mahāvyutpatti*), the *Bye brag tu rtogs byed ḫbriṅ po*, and the *Bye brag tu rtogs byed chuṅ ṅu*. The second work is better known under the title *Sgra sbyor bam po gñis pa*. The third work is no longer extant. The two extant works were in circulation at least by 814 (Uray, 1989; Scherrer-Schaub, 2002; Hermann-Pfandt, 2008). Sanskrit-Tibetan bilingual lexicography continued form that time until our day (cf. Ruegg, 1998).

Modern bilingual Tibetan-Sanskrit dictionaries include some of the finest works of Tibetan lexicography. Lokesh Chandra compiled a 12 volume Tibetan-Sanskrit dictionary on the basis of canonical Buddhists texts available in both languages (1958-61). This work was continued with seven supplementary volumes (1992-1994) and a one volume Sanskrit-Tibetan index (2007). Attestations are given for each entry. In addition, Negi (1993-2004) compiled another Tibetan-Sanskrit dictionary, this one in sixteen volumes. Negi includes extensive quotations in addition to citations and made reference to a larger number of texts than Chandra. In addition to these two Tibetan-Sanskrit dictionaries, there are bilingual indices available for a number of Tibetan translations of Sanskrit Buddhist texts, including: *Abhidharmakośabhāṣya* (Hirakawa, 1973-1978), *Bodhicaryāvatāra* (Weller, 1952-1955), *Kāśyapaparivarta* (Weller, 1933), *Mahāyānasūtrālaṅkāra* (Nagao, 1958-1961), *Meghadūta* (Chimpa et al., 2011), *Nyāyabindu* (Obermiller, 1970 [1927-28]), *Prasannapadā Mādhyamakavṛtti* (Yamaguchi, 1974), *Yogācārabhūmi* (Yokoyama, 1996), *Laṅkāvatārasūtra* (Suzuki 2000), *Sukhāvatīvyūhasūtra* (Inagaki, 1984), and *Saddharmapuṇḍarīkasūtra* (Ejima et al., 1985-1993), among others. Apart form works treating Sanskrit, a highlight in the history of Tibetan multilingual lexicography is the inclusion of Tibetan as one of the five languages in the monumental pentaglot dictionary of the Qianlong period (cf. Corff, et al., 2013).

As is common across the world, monolingual lexicography has more recent origins than the compilation of multilingual works. As Tibetan changed through time a genre arose which explained archaisms with newer terms. The earliest of these 'old-new-terminologies' (*bdra-gsar-rñiṅ*) is the *Li śi gur khaṅ* by Rin chen bkra śis written in 1536 (cf. Taube, 1978). The writing out of verb paradigms, which had been phonetically leveled in many dialects, dates to the late eighteenth century, the earliest author of this genre being A kya yoṅs ḥdzin dbyaṅs can dgaḥ baḥi blo gros (1740-1827, cf. Hill 2010a, p. xxiii). Chos kyi grags pa (1980[1949]) wrote the first monolingual Tibetan dictionary to be organized alphabetically. Until recently this was used very widely by Tibetan as well as Western scholars. A Tibetan-Tibetan dictionary of lasting importance is that edited by Blo mthun bsam gtan (1979). This excellent dictionary includes carefully written definitions and a more sophisticated and reliable handling of verbs than found in most dictionaries. Its relatively small size means that obscure words are not to be found, but it has a strength in colloquial words and eastern dialect forms. The methodological high water mark of monolingual works is probably Ṅag dbaṅ tshul khrims' (1997) dictionary of difficult and archaic words. The author provides attestations and cites the works they are found in, but does not specify page and line numbers and has an inadequate bibliography; consequently, these citations are not easily verified.

The first Tibetan dictionary by a western author is a manuscript Tibetan-Latin dictionary by the Cappucian missionaries Giuseppe da Ascoli, Franceso Maria da Tours and F. Domenico da Fano (1674-1728), compiled between 1708 and 1713. This dictionary unfortunately remains unpublished but according to Simon (1964, p. 85) an extract is held at the Bibliothèque Nationale (Fonds Tibétain No. 542). A Tibetan-Italian dictionary was compiled by F. Francesco Orazio della Penna (1680-1745), a student of da Fano. The text of this work was translated into English and considerably mangled. The English version became the first published Tibetan dictionary (Schroeter, 1826) but the original remains unpublished. Schroeter, died while revising the work and learning Tibetan; the editors who saw the work through publication knew no Tibetan (cf. Simon 1964; Bray 2008).

These first two dictionaries and others of the 19th and early 20th century are well discussed by Simon (1964). Jäschke's dictionary from this period is the first Tibetan dictionary of real caliber and as a work of lexicography is almost unrivaled to this day. Subsequent years have witnessed the publication of scores of other Tibetan dictionaries (cf. Simon 1964, Viehbeck 2017). Hundreds of Tibetan dictionaries are now available; these include bilingual dictionaries, both to and from such languages as English, French, German, Latin, Japanese, etc. and specialized dictionaries focusing on medicine, plants, dialects, archaic terms, neologisms, etc. (cf. Walter, 2006; McGrath, 2008). None of these works matches the methodological rigor or sophistication of Jäschke, and many are directly derivative of his work.

The single most impressive work of Tibetan lexicography is the ongoing *Wörterbuch der tibetischen schriftsprache* published by the Bayerische Akademie der Wissenschaften (Francke et al., 2005-). Helmut Hoffmann founded the project in 1954; the first fascicle was published in 2005. The thirty four fascicles published by 2016 cover from *ka* until *dharma*. Each entry gives copious citations of original sources precisely cited to page and line number. The use of previous dictionaries is carefully distinguished from the evidence of textual attestations. In addition, very thorough reference to previous scholarship is given when relevant. The compilation of the dictionary is discussed by Uebach & Panglung (1998), to which Maurer & Schneider (2007) and Schneider & Maurer (2012) provided a more recent perspective.

Tibetan electronic corpora

The Tibetan language is served by a number of electronic text corpora, but to-date only one such corpus includes word breaking and part-of-speech tagging. The largest electronic corpus is by far the ever expanding e-text library of the Tibetan Buddhist Resource Center (www.tbrc.org), which as of December 27, 2014 consisted of 959,020 pages of text. These texts are encoded in Unicode and stored in XML files. The material for this collection comes from two sources: OCRed modern printed texts and the digital files of publishers of Tibetan texts. The TBRC provides a dedicated search interface, but the corpus itself is not available for download.

The Old Tibetan Documents Online (OTDO) is a collection of 109 Old Tibetan texts (http://otdo.aa.tufs.ac.jp/ and http://otdo.aa-ken.jp/). The texts include documents discovered at the library cave at Dunhuang and imperial inscriptions form central Tibet. These materials are not included in any other digital corpus. OTDO texts are encoded in a purposed designed Roman transcription. The OTDO includes a search interface; the corpus is downloadable.

Otani Tibetan E-Texts (http://web1.otani.ac.jp/cri/twrpw/results/e-texts/) consists of 14 texts input from xylographs held at the Otani University library. The bulk of this collection is historical and biographical classics. These texts, in Unicode, are available for download. The collection is not searchable online.

Since 1988 the Asian Classics Input Project (ACIP) has manually transcribed texts from the Buddhist Canon into a purpose designed Roman transcription. According to a now dead link that is cited on Wikipedia (en.wikipedia.org/wiki/Michael_Roach#cite_note-BiA2-13 accessed 29 December 2014) in 2011 the project had input over 8,500 texts, circa 500,000 pages. More recent information is not available on the ACIP homepage (www.asianclassics.org). Despite a complex editorial procedure designed to reduce copying errors, their texts are not universally regarded as reliable.

A digital version of the Derge Kanjur (an edition of the Tibetan Buddhist canon), prepared by the British Library and SOAS, University of London is hosted by the Tibetan and Himalayan Digital Library of the University of Virginia (www.thlib.org/encyclopedias/literary/canons/kt/catalog.php#cat=d/k). The data are in Unicode and stored in XML. There is a search facility. Unfortunately, the edition currently online contains many typos. The TBRC in collaboration with Eusukhia (esukhia.org) have proofread these materials, but the corrected version is not yet available for public download or consultation.

The one currently available part-of-speech tagged Tibetan corpus was compiled as part of the research project 'Tibetan in Digital Communication' funded by the UK's Arts and Humanities Research Council and based at SOAS, University of London. In addition to the corpus, the project developed a number of digital tools allowing the corpus to be employed in many areas of humanities research, and enabling other researchers to more easily develop their own corpora or software tools. These tools included an online corpus management system, a word tokenizer, and a part-of-speech tagger (https://github.com/tibetan-nlp and Hill & Garrett, 2017a-c).

Corpus-based lexicography

While the size and coverage of Tibetan's digital corpus is extraordinary, until now its lexicographic utility has been limited. Without a part-of-speech tagged corpus, it can be very laborious to navigate through vast volumes of data. For example, a search for the syllable *gyis* will invariably flag up the agentive case marker *gyis* as well as the imperative form of the verb *bgyid* ('to do'). If one is studying the imperative of the verb *bgyid*, then one has no choice but to look through hundreds of examples of the agentive case marker. A part-of-speech tagger solves this problem by using rules or statistics to distinguish homonyms.

The SOAS project created a part-of-speech tagger which applies a sequence of tag-removing rules to arrive at an analysis of a sentence. First implemented using regular expressions and subsequently rewritten in Constraint Grammar (http://beta.visl.sdu.dk/constraint_grammar.html), the part-of-speech tagger consists of a series of contextual rules. For example, the tagger includes a number of rules designed to distinguish between negation and nominals, including correctly categorising *ma* as either [neg] or "mother" [n.count], and *mi* as either [neg] or "person" [n.count]. Some of these rules are shown below; for the sake of non-Tibetan readers the Tibetan script is written in Roman bold:

```
#056: Isolating ma [neg] in the phrase skad cig ma gcig 'one moment'
    REMOVE (n.count) (-2 ("<skad>")) (-1 ("<cig>")) (0 ma) (1 ("<gcig>")) ;


#063: Identifying ma [neg] in the prohibitive
    SELECT (neg) (0 ma) (1C (v.pres)) (2 (cv.imp)) ;
    REMOVE (d.indef) (-2 ma) (-1C (v.pres)) (0 (cv.imp)) ;

#066: Isolating ma [n.count] and mi [n.count] before case markers
```

```
    REMOVE (neg) (0 mami) (1 case.xxx LINK NOT 0 v.xxx) ;
```

Rule #056 says that *ma* must be negation when occurring in a certain fixed phrase (*skad cig ma gcig*), that is, when preceded by two specific words (*skad cig*) and followed by another (*gcig*). The first part of rule #063 says that if the first word after *ma* is a certain (hence, "C") [v.pres], and the second word after *ma* is a possible [cv.imp], then assign [neg] to *ma*; while the second part of the rule makes sure that in this same context, homonymous [cv.imp]/[d.indef] should be assigned [cv.imp]. Finally, rule #066 says that a *ma* or *mi* should be a nominal if it's followed by a possible case marker that cannot also be a verb.

The SOAS part-of-speech tagger achieves > 99.8% accuracy. That is, the tagger almost never removes a tag for a word if the tag is correct. However, the tagger is often unable to decide on a single tag for a word. The average word has 1.41 tags, which means that while many words are assigned a single (and almost always correct) tag, others are left with 2, 3 or more possible tags.

The SOAS corpus consists of no more than 1 million words, making the hand-tagged Tibetan corpus rather small by the standards of corpus linguistics, with many infrequent words and senses simply not occurring in the sample. To expand the corpus and thereby provide a more secure footing for informed lexicographic investigations, the SOAS part-of-speech tagger has been unleashed on the additional corpora mentioned above. To the extent that these corpora share features in common with the hand-tagged corpus, the exercise has been successful.

Future prospects

The previous section discussed a part-of-speech tagger which facilitates Tibetan lexicographic reserach through the disambiguation of homophones. However, part-of-speech tagging alone has limited payoff; other techniques from computational linguistics will also need to be developed or adapted for Tibetan.

One obstacle is that despite the existence of numerous dictionaries organised and alphabetised into a list of entries by head word, no serious attempt has yet been made to uncover and articulate principles of lemmatisation for Tibetan, that is, the systematic grouping of related word forms under the same lexical entry. The part-of-speech tagger for Tibetan does not yet tag the variant forms of a word under the same lemma. For example, the stems of "cut" in Classical Tibetan are *gcod*, *bcad*, *gcad*, and *chod*; all of these forms should be listed under the same lemma. Old Tibetan poses further lemmatisation challenges. For example, syllable boundaries are not as consistently marked as they are in Classical Tibetan; we find *rdzogso* instead of *rdzogs-so*, *phulo* instead of *phul-lo*, and so on. Once the conditions of these mergers is understood, rules can be written to expand the merged syllables into full forms that refer to the correct lemma: if *phulo* expands to disyllabic *phu-lo*, the first *phu* must be classed as a variant form of *phul*. Since it would be absurd to create a dictionary without at the very least cross-referencing the variant forms of a word, work on lemmatisation, whether automatic or manual, must be prioritized.

A second obstacle towards further progress is the problem of Tibetan word segmentation. As with Chinese, Tibetan text does not use whitespace or other mechanisms to mark word boundaries. As with Chinese, the automatic determination of word boundaries by computer is a "hard problem". Various solutions to this problem have been explored. One approach has followed Huidan et al (2011) by re-casting Tibetan word segmentation as a syllable tagging problem, with each syllable in search of an appropriate word-internal position label. For example, the only syllable of a monosyllabic word is tagged with 'S' for "single syllable", and the first, middle, and end syllables of multisyllabic words are tagged with "B", "M", and "E", respectively. The machine then applies the syllable tagging patterns it learns from a training corpus to the new texts it is exposed to. Another approach leaves less to chance, exploiting simultaneous left-to-right and right-to-left maximal dictionary-based matching using the Aho-Corasick algorithm (https://github.com/tibetan-nlp). The urgency of the word segmentation problem is underscored by two facts: first, that automatic part-of-speech tagging currently performs better for Tibetan than automatic word segmentation; and second, that mistakes in word segmentation tend to feed mistakes in part-of-speech tagging, since the latter process requires a segmented corpus. One direction for future research then would be to find a way to improve both processes by allowing them to work in tandem, each to the benefit of the other.

A third obstacle relates to the challenges presented by new, unseen texts. Unknown words and named entities can wreak havoc for dictionary-based methods, and further problems are introduced by the consideration of data representing diverse genres, text types, and linguistic epochs. Whether existing tools can be shown to be successful in the face of such diversity remains to be established.

No Tibetan dictionary has yet been compiled which benefits from the advances in corpus linguistics which have revolutionized the lexicography of better studied languages. The challenge for Tibetan lexicography is to transition to the modern method of lexicography by exploiting the vast collections of digital Tibetan materials now available online. With Tibetan computational linguistics in its infancy, and generally not a priority for commercial or governmental funding, progress has necessarily been slow. However, the path forward is clear, and the obstacles to surmount evident. Future prospects for Tibetan lexicography are bright.

References

Andersen, P. K. (1987). “Zero-anaphora and related phenomena in Classical Tibetan”. *Studies in Language,* 11, 279-312.

Bray, J. (2008). "Missionaries, officials and the making of the 1826 Dictionary of the Bhotanta, or Boutan Language." *Zentralasiatische Studien,* 37, 33-75.

van Driem, G. (2012). “The Trans-Himalayan phylum and its implications for population prehistory.” *Communication on Contemporary Anthropology,* 5, 135-142.

Garrett, E., Hill, N.W., Kilgarriff, A., Vadlapudi, R. & Zadoks, A. (forthcoming). “The contribution of corpus linguistics to lexicography and the future of Tibetan dictionaries.” *Revue d'Etudes Tibétaines*.

Haller, F. (2009). "Switch-reference in Tibetan." *Linguistics of the Tibeto-Burman Area,* 32(2), 45-106.

Hermann-Pfandt, A. (2008). *Die lHan kar ma: Ein früher Katalog der ins Tibetische übersetzten buddhistischen Texte Kritische Neuausgabe mit Einleitung und Materialien*. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.

Hill, Nathan W. (2010a). *A Lexicon of Tibetan Verb Stems as Reported by the Grammatical Tradition*. Munich: Bayerische Akademie der Wissenschaften.

Hill, Nathan W. (2010b). “An overview of Old Tibetan synchronic phonology.” *Transactions of the Philological Society,* 108(2), 110-125.

Hill, Nathan W. (2012). “Tibetan -las, -nas, and -bas.” *Cahiers de Linguistique—Asie Orientale,* 41(1), 3-38.

Hill, Nathan W., & Garrett, Edward. (2017a). A part-of-speech (POS) tagged corpus of Classical Tibetan [Data set]. Zenodo. http://doi.org/10.5281/zenodo.574878

Hill, Nathan W., & Garrett, Edward. (2017b). A part-of-speech (POS) lexicon of Classical

Tibetan for NLP [Data set]. Zenodo. http://doi.org/10.5281/zenodo.574876

Garrett, Edward, & Hill, Nathan W. (2017c). A rule based Tibetan part-of-speech (POS) tagger for the creation of gold standard training data [Data set]. Zenodo. http://doi.org/10.5281/zenodo.574882

Huidan L., Nuo, M., Ma, L., Wu, J. & He, Y. (2011). "Tibetan Word Segmentation as Syllable Tagging Using Conditional Random Field." 25th Pacific Asia Conference on Language, Information and Computation: 168–177.

Maurer, P. & Schneider, J. (2007). "Neues Datenbanksystem für das Wörterbuch der tibetischen Schriftsprache." *Akademie Aktuell* 22.3: 23.

McGrath, Bill (2008). "Tibetan Dictionaries." http://www.thlib.org/reference/dictionaries/tibetan-dictionary/dictionary-biblio.php (accessed, 5 March 2013)

Seyfort Ruegg, D. (1998). "Sanskrit-Tibetan and Tibetan-Sanskrit Dictionaries and Some Problems in Indo-Tibetan Philosophical Lexicography", in B. Oguibénine (ed.), *Lexicography in the Indian and Buddhist Cultural Field*, Munich: Bayerische Akademie der Wissenschaften (Studia Tibetica Band IV), 115-142.

Scherrer-Schaub, C. A. (2002). "Enacting Words: A Diplomatic Analysis of the Imperial Decrees (bkas bcad) and Their Application in the sGra sbyor bam po gñis pa Tradition," *Journal of the International Association of Buddhist Studies,* 25(1-2), 263-340.

Schneider, J. & Maurer, P. (2012). "Ein Wörterbuch des Tibetischen." *Akademie Aktuell,* 40(1), 50-51.

Schroeter, F. (1826). *A Dictionary of the Bhotanta or Boutan Language*. Serampore.

Simon, W. (1964). "Tibetan Lexicography and Etymological Research." *Transactions of the Philological Society,* 63(1), 85-107.

Tournadre, N. (1996). *L'ergativité en tibétain: approche morphosyntaxique de la langue parlée.* Louvain: Peeters.

Tournadre, N. (2008). "Arguments against the Concept of 'Conjunct'/'Disjunct' in Tibetan." *Chomolangma, Demawend und Kasbek. Festschrift für Roland Bielmeier zu seinem 65. Geburtstag*. B. Huber, et al., eds. Vol 1. Halle (Saale): International Institut for Tibetan and Buddhist Studies, 281–308.

Uebach, H. & Panglung, J. L. (1998). "The Project "Dictionary of Written Tibetan" : An Introduction." *Lexicography in the Indian and Buddhist cultural field*. Boris L. Oguibénine, ed. Munich: Kommission für Zentralasiatische Studien, Bayerische

Akademie der Wissenschaften. 149-163.

Uray, Géza (1989). “Contributions to the date of the Vyutpatti-treatises.” *Acta Orientalia Academiae Scientiarum Hungaricae,* 43(1), 3-21.

Walter, Michael (2006). “A bibliography of Tibetan dictionaries.” *Bibliographies of Mongolian, Manchu-Tungus, and Tibetan dictionaries*. H. Walravens, ed. Wiesbaden: Harrassowitz, 174-235.

Viehbeck, Markus (2017). “Coming to terms with Tibet: scholarly networks and the production of the first 'modern' Tibetan dictionaries.” *Ancient Currents, New Trends: Papers Presented at the Fourth interational Seminar of Young Tibetologists*. F.-X. Erhard, ed. Potsdam: edition tethys, 469-489.

Zadoks, Abel (2000). Switch Evidence in Old Tibetan: between Switch Reference and Evidentiality. Paper presented at the 9th Seminar of the IATS. Leiden University, The Netherlands. 24-30 June 2000.

Zadoks, Abel (2002). The Tibetan Connection: Switch Reference and Evidentiality from Old Tibetan to Middle Tibetan. Paper presented at the 8th Himalayan Languages Symposium. Bern University, Switzerland. 19-22 September, 2002.

Dictionaries

Blo mthun bsam gtan (1979). *Dag yig gsar bsgrigs*. Xining: Mtsho sṅon mi rigs dpe skrun khaṅ.

Chandra, L. (1958-1961). *Tibetan-Sanskrit dictionary, based on a closed comparative study of Sanskrit originals and Tibetan translations of several texts*. New Delhi: International Academy of Indian Culture.

Chandra, L. (1992-1994). *Tibetan-Sanskrit dictionary. Supplementary volumes*. New Delhi: International Academy of Indian Culture and Aditya Prakashan.

Chandra, L. (2007). *Sanskrit-Tibetan dictionary: being the reverse of the 19 volumes of the Tibetan-Sanskrit dictionary.* New Delhi: International Academy of Indian Culture and Aditya Prakashan.

Chimpa, L., Kumar, B. & Samten, J., eds. (2011). *Meghadūta: critical edition with Sanskrit and Tibetan index*. New Delhi: Aditya Prakashan.

Chos kyi grags pa (1980[1949]). *Brda dag miṅ tshig gsal ba*. Dharamsala: Damchoe Sangpo

Corff, Oliver, et al. (2013). *Auf kaiserlichen Befehl erstelltes Wörterbuch des Manjurischen in fünf Sprachen. „Fünfsprachenspiegel“.* Wiesbaden: Harrassowitz 2013, Ejima, Yasunori, et al. (1985-1993). *Index to the Saddharmapuṇḍarīkasūtra: Sanskrit, Tibetan, Chinese.* Tokyo: Hotoke no Sekaisha,

Francke, Herbert, et al. (2005-). *Wörterbuch der tibetischen Schriftsprache*. Munich: Verlag der Bayerischen Akademie der Wissenschaften.

Hirakawa, Akira (1973-1978). *Index to the Abhidharmakośabhāṣya*. Tokyo: Daizō Shuppan.

Inagaki, Hisao (1984). *A tri-lingual glossary of the Sukhāvatīvyūha sūtras: indexes to the Larger and Smaller Sukhāvatīvyūha sūtras*. Kyoto: Nagata Bunshodo.

Jäschke, H. A. (1881). *Tibetan English Dictionary*. London: Unger Brothers.

Ṅag dbaṅ tshul khrims (1997). *Brda dkrol gser gyi me long*. Beijing: Mi rigs dpe skrun khang.

Nagao, G. (1958-1961). *Index to the Mahāyāna-sūtrālaṁkāra*. Tokyo: Nihon Gakujutsu Shinkōkai.

Nagao, G. (1994). *An index to Asaṅga's Mahāyānasaṃgraha*. Tokyo: The International Institute for Buddhist Studies.

Negi, J. S. (1993-2004). *Tibetan-Sanskrit dictionary*. Sarnath: Dictionary Unit, Central Institute of Higher Tibetan Studies.

Obermiller, E. (1970). *Indices verborum Sanskrit-Tibetan and Tibetan-Sanskrit to the Nyāyabindu of Dharmakīrti and the Nyāyabinduṭīka of Dharmottara*. Osnabrück: Biblio-Verlag.

Suzuki, D. T. (2000). *An index to the Lankavatara sutra (Nanjio edition): Sanskrit-Chinese-Tibetan, Chinese-Sanskrit, and Tibetan-Sanskrit*. New Delhi: Munshiram Manoharlal Publishers.

Weller, F. (1933). *Index to the Tibetan translation of the Kāçyapaparivarta*. Cambridge: Harvard-Yenching Institute.

Weller, F. (1952-5). *Tibetisch-sanskritischer Index zum Bodhicaryāvatāra*. Berlin: Akademie-Verlag.

Yamaguchi, S. (1974). *Index to the Prasannapadā Madhyamaka-vṛtti*. Kyoto: Heirakuji-Shoten.

Yokoyama K. (1996). *Index to the Yogācārabhūmi, Chinese-Sanskrit-Tibetan*. Tokyo: Sankibō Busshorin.