

UC Santa Barbara

Himalayan Linguistics

Title

Algorithmic description of the decomposition and checking of a Classical Tibetan syllable

Permalink

<https://escholarship.org/uc/item/70z8069f>

Journal

Himalayan Linguistics, 17(1)

Authors

Roux, Élie
Hildt, Hélios Drupchen

Publication Date

2018

DOI

10.5070/H917135529

Copyright Information

Copyright 2018 by the author(s). This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed

himalayan linguistics

A free refereed web journal and archive devoted to the study of the
languages of the Himalayas

Himalayan Linguistics

Algorithmic description of the decomposition and checking of a Classical Tibetan written syllable

Elie Roux

Buddhist Digital Resource Center

Hélios Hildt

Buddhist Digital Resource Center

ABSTRACT

This document presents our research on the the correct formation of a Classical Tibetan syllable. It was triggered by attempts at defining the boundaries of well-formed syllables in Classical Tibetan for spell checking purposes.

Formalizing the formation of the syllable led us to inspect the small differences among grammar books, both in Western and Tibetan language. We then checked these differences against the Tibetan dictionaries we consider reliable, and also against the Kangyur.

Our inquiry finally led us to study the way to decompose a syllable, discussing the ambiguous cases, as well as the formation of the Dzongkha syllable.

KEYWORDS

Tibetan, spell checking, NLP

This is a contribution from *Himalayan Linguistics*, Vol. 17(1): 50-66.

ISSN 1544-7502

© 2018. All rights reserved.

This Portable Document Format (PDF) file may not be altered in any way.

Tables of contents, abstracts, and submission guidelines are available at
escholarship.org/uc/himalayanlinguistics

Algorithmic description of the decomposition and checking of a Classical Tibetan written syllable

Elie Roux
Buddhist Digital Resource Center

Hélios Hildt
Buddhist Digital Resource Center

1 Introduction

This document presents our research on the orthography of a Classical Tibetan syllable. The study of this question was triggered by attempts at defining the boundaries of well-formed written syllables in Classical Tibetan for NLP purposes. What we are trying to assess is the extend of the possibilities of which a modern Tibetan person would think as valid if they wanted to invent a new Classical Tibetan word. This is a necessary first step to be able to discriminate valid Classical Tibetan from other forms such as transliterated Sanskrit, or just mistakes. Of course, the syllable level is not enough for a full feature spell checker; but we hope our work will provide a basis for future work in this area, as well as a complete description of the various corner cases of syllable construction. This research has already been useful, allowing us to create

- a basic spell checker for Classical Tibetan using the *hunspell* library
- rules for collation of Classical Tibetan

Formalizing the formation of the syllable led us to inspect the small differences between grammar books, both in Western and Tibetan language¹. We then checked these differences against some chosen Tibetan dictionaries² and the Derge Kangyur³.

Our inquiry finally led us to study the way to decompose a syllable, discussing the ambiguous cases, as well as the formation of the Dzongkha syllable.

¹ Our main readings in Western languages were (Tournadre and Dorje, 2005) and (Beyer, 1992), but we also consulted (Bacot, 1946) and (Das, 1915); the Tibetan sources were (འཇུ་པ་མཚན་མོས་གྱུ་འབྲུང་གནས།, XVIIIth c.) and (ཚེ་རྟན་ཞབས་བྲུང་།, 2003). The bibliography used in this document can be downloaded on <https://github.com/eroux/tibliography/>.

² The dictionaries we consulted are (Dorje et al., 2003), (མོན་ལམ།, 2016), (དུང་དཀར་རྩོམ་པུང་འབྲིན་ལམ།, 2002), and in a lesser extent (Yisun, 1985), (ཚོས་སྐད་རྫོང་ཁ་ཚོགས་མཚན་མོ།, 2010) and (Negi, 1993). We also consulted (Hill, 2010).

³ (ཚོས་གྱུ་འབྲུང་གནས།, 1721), input by Esukhia, <https://github.com/Esukhia/derge-kangyur>

2 Elements of a Tibetan syllable

2.1 Description and vocabulary

2.1.1 Basic description

For the description of the different elements of a Tibetan syllable, we will take the vocabulary used in (Tournadre and Dorje, 2005). As an example, decomposing the syllable འཇུངས་ results in radical letter འ, prefix འ, superscribed འ, subscribed འ, vowel accent འ, first suffix འ and second suffix འ. Among these six categories, only the radical letter is mandatory, all the others are optional⁴. This constitutes a good framework for the description of the elements of a syllable but lacks a few elements.

2.1.2 Missing elements

First, some syllables (such as འཇུངས་) contain a wasur (འ). These are traditionally considered as subscribed letters, with the ability to combine with other subscribed (as in འཇུངས་). This last case breaks our initial description because it requires two optional subscribed letters instead of one, and the second can only be a wasur. Our proposal is to treat the wasur separately from the subscribed letters in order to keep things simple. So འཇུངས་ would be decomposed as radical letter འ, wasur, first suffix འ, second suffix འ; and འཇུངས་ as radical letter འ, subscribed འ and wasur.

A second problem is the འ “suffix” (as in འཇུངས་, “chapter”, /leu/, 2 syllables): it not considered as a suffix in traditional grammars (that are usually completely silent about it), and does not have normal suffix properties. We will call it a *special suffix*. At most one can appear in a syllable.

The third missing feature is affixed particles. These are appended to syllables with no suffixes or replace suffix འ. For instance འཇུངས་ is decomposed as radical letter འ and affixed particle འ. Two affixed particles can even be combined, as in འཇུངས་.

In order to describe the different elements, we will call *final part* what is written on the right side of the radical letter; *main stack* the radical and everything written above and below except vowel accents (so subscribed, superscribed letters and wasur).

We also would like to introduce the notion of *root*, which is constituted of everything preceding the vowel accent or final part. This notion will come handy in some parts of this document, and is crucial for collation purposes.

2.1.3 Formalization

Now we can propose a formalization of the elements of the syllable with some symbols coming from the regular expressions: “?” means an element that can either be omitted or appear once, and “|” simply means an exclusive “or”:

$$\begin{aligned}
 \text{syllable} &= [\text{root}][\text{vowel}]?[\text{final part}]? \\
 \text{root} &= [\text{prefix}]?[\text{main-stack}] \\
 \text{main-stack} &= [\text{superscribed}]?[\text{radical}][\text{subscribed}]?[\text{wasur}]? \\
 \text{final-part} &= ([\text{special-suffix}][\text{affixed-particles}]?) \mid [\text{suffixes}] \mid [\text{affixed-particles}] \\
 \text{affixed-particles} &= [\text{affixed-particle}][\text{affixed-particle}]?
 \end{aligned}$$

⁴ For the sake of simplicity, we will not consider that the “a” vowel accent or the first suffix འ are implied when no vowel accent or first suffix is present.

suffixes = [first-suffix][second-suffix]?

Note that although this is the way a syllable is built, some syllables can have multiple decomposition possibilities according to this scheme. For instance བར་ can be decomposed as either:

- radical བ + suffix ར (/bar/, “between”)
- radical བ + affixed particle ར (either /war/ nominalizer+oblique mark, or /bar/ “cow”+oblique mark)

This schema also doesn’t account for implied suffix འ: for instance if དཀའི་ did not have its affixed particle, it would be དཀའ་, and for some purposes it may also be useful to retain this information, but we chose the most simple schema for readability purposes.

2.2 Constraints on the construction of a syllable

In this part we will study the different possibilities of construction of these elements, first for each element taken independently, then the constraints of their relations.

2.2.1 Simple constraints

All consulted sources agree on all the constraints on the different elements except the special suffix. These constraints are:

- prefix can be ཀ ད བ མ OR འ
- first suffix can be ཀ ང ཅ ཆ ཇ ཈ ཉ ཐ ད དྷ ན པ ཕ བ བྷ མ ཙ འ ར ལ OR ས
- second suffix can be ས⁵
- superscribed can be ར ལ OR ས
- subscribed can be འ ར OR འ ལ
- radical letter can be ཀ ཁ ག གྷ ང ཅ ཆ ཇ ཈ ཉ ཐ ད དྷ ན པ ཕ བ བྷ མ ཙ འ ར ལ ཤ ཥ ས ཏ OR མ
- vowel accent can be འ ཡ ར ལ OR འ ར
- wasur is འ
- affixed particle can be འ ར ལ ཤ ཥ ས OR ས

When inspecting the different constraints on the relation between elements, we can see some consistent lists among all sources:

- ཀ ཁ ག གྷ ང ཅ ཆ ཇ ཈ ཉ ཐ ད དྷ ན པ ཕ བ བྷ མ ཙ འ ར ལ for superscribed ར + radical letter
- ཀ ཁ ག གྷ ང ཅ ཆ ཇ ཈ ཉ ཐ ད དྷ ན པ ཕ བ བྷ མ ཙ འ ར ལ for superscribed ལ + radical letter

⁵ Old Tibetan also has ད.

3 Decomposing a syllable

3.1 Finding the main stack

What we propose here is to find the main stack in any valid syllable. Once the main stack is found, a set of rules described in a later section can be applied to find the different elements. Some rules found in (Tournadre and Dorje, 2005) describe a general way for doing that but are not precise enough to be implemented. For instance they would fail in cases such as བའི་, ལེན་, བའམ་, etc. Based on the information we gathered from our research, we propose to apply the following rules in this particular order¹⁷:

1. if the syllable contains a subscript, superscript or wasur then the main stack is what contains it (ex: དབྱེས་ → རྩ)
2. if a letter other than འ carries a vowel then it is the main stack (གཞི་ → འ)
3. if a vowel is carried by འ and འ is the first letter then འ is the main stack (འོད་ → འ)
4. if a vowel accent is carried by an འ which is not the first letter, the main stack is before the first འ with a vowel¹⁸ (མའིའོ་ → མ)

The following rules deal with syllables with no vowel, superscribed, subscribed nor wasur, only composed of letters that could be radical letters:

5. if the syllable has three or four letters and ends with འང or འམ, then the main stack is right before འང or འམ¹⁹ (བའམ་ → བ)
6. if the syllable is composed of only one letter this letter is the main stack
7. if the syllable contains two letters, then the first is the main stack (བར་ → བ)
8. if the syllable contains four letters, the main stack is the second letter (བཟབས་ → ཟ)

The following rules deal with syllables composed of a sequence of 3 simple letters:

9. if the final letter is not ས, then the main stack is the second letter (བཟའ་ → ཟ)
10. if the first letter cannot be a prefix to the second letter when it has no superscribed nor subscribed (see above for conditions on prefixes), then the main stack is the first letter (ཐགས་ → ཐ)
11. if ས cannot be second suffix after the second letter if it was a first suffix (meaning the second letter is not ཀ, ང, བ nor མ), then the main stack is the second letter (གནས་ → ན)

3.2 Ambiguous syllables

If the syllable doesn't match any of the above rules, then it is ambiguous. These syllables are those with no explicit vowel, no superscribed, no subscribed, no wasur, three letters, a final ས, the first letter can be a prefix to the second letter with no superscribed nor subscribed, and the second letter

¹⁷ This means that the first rule the syllable will match will be the one determining the main stack. This has the advantage to be transcribed easily into computer code.

¹⁸ This rule also works for Old Tibetan using འམ. The only exception to this rule is the very unlikely མའིའོ་(O), but the legality of this syllable is not clear.

¹⁹ This rule could be extended for Old Tibetan with འང.

is ག, ར, བ or མ. It is easy to find the 9 corresponding cases: མརས་, མགས་, དབས་, དངས་, དགས་, དམས་, བགས་, འབས་ and འགས་.

For each of these cases we have three possible structures:

1. ས is a second suffix (e.g. མརས་ can be decomposed as radical མ, suffix ར, second suffix ས, we note this decomposition $མ|རས$)
2. ས is a suffix (e.g. མརས་ is radical ར, prefix མ, suffix ས, noted $མར|ས$)
3. ས is an affixed particle (e.g. མརས་ radical ར, prefix མ, suffix འ replaced by affixed particle ས, noted $མར|འ+ས$)

Case 2 and 3 have equivalent pronunciations but are important to distinguish because they will imply a different analysis in terms of lemmatization and part of speech tagging.

In order to decide where the main stack is, the only way is to take the decomposition with the highest probability according to our knowledge of the Tibetan language.

Let's review the different possibilities for the 9 cases, prefixing by * a form unattested in the dictionaries we consulted:

- མརས་: $མ|རས$, $*མར|ས$, $མར|འ+ས$ → ambiguous, མ as main stack is more intuitive to our Tibetan informants
- འབས་: $*འ|བས$, $*འབ|ས$, $འབ|འ+ས$ → བ is the main stack
- མགས་: $*མ|གས$, $*མག|ས$, $*མག|འ+ས$ → ambiguous, མ as main stack is more intuitive to our Tibetan informants
- བགས་: $བ|གས$, $*བག|ས$, $*བག|འ+ས$ → བ
- འགས་: $*འ|གས$, $འག|ས$, $འག|འ+ས$ → ག
- དབས་: $*ད|བས$, $དབ|ས$ (OT), $དབ|འ+ས$ → བ
- དགས་: $*ད|གས$, $དག|ས$, $དག|འ+ས$ → ག
- དངས་: $ད|ངས$ (misspelling of དངས་²⁰), $*དང|ས$, $*དང|འ+ས$ → ད
- དམས་: $*ད|མས$, $དམ|ས$, $དམ|འ+ས$ → མ

3.3 Decomposition of the syllable

Once the main stack is found, the prefix, vowel accent and wasur are immediate to find. If a superscript or subscript is present, they can be immediately found with the rules exposed in “Simple constraints”.

Suffixes can be immediately classified between special suffix, first suffix, second suffix and affixed particle; the exceptions are ས and ར, plus the syllables ཀའི་, ཟའོ་, ལྷའོ་. The latter are very rare and the context should make it obvious, but they are not decidable at syllable level.

²⁰ Attested in (Negi, 1993) (vol. 6) and (Duff, 2000).

	བ	པོ	བཤམ་	ལོ	མ/མོ
ག	X	X			X
ང	X	X			X
ད	X	X	X		X
ན	X	X			X
བ	X				X
མ	X	X			
འ			X	X	X
ར	X	X	X		X
ལ	X	X	X		X
ས	X	X	X		X
-	X	X	X	X	X

At the time of the redaction of this article, some questions are still pending, like the possibility to affix one of these particles after the special suffix ཟ, or the case of the following syllables found in the official lists: རྫོང་ལྷ་ རིལ་ལྷ་ རྫོང་ (we are not sure if these are errors or if other fusion with ཟ or other syllables are possible).

4 Conclusion

We have described all the possible combinations of a Classical Tibetan syllable, listing all constraints and exceptions we found, resulting in a complete set of rules easy to implement in a computer language.

An immediate application has been to implement the rules in a spell checker running with the *hunspell* library,²³ freely available on <https://github.com/eroux/hunspell-bo>. Our formalization of the ཟ and འ endings allowed us to implement stricter rules in our spell checker and detect more potential mistakes.

We have also managed to build solid rules for the decomposition of a Classical Tibetan syllable, listing ambiguities and possible disambiguation.

REFERENCES

- Bacot, Jacques. 1946. *Grammaire du Tibétain littéraire*. Paris: Librairie d'Amérique et d'Orient.
 Beyer, Stephan. 1992. . Albany (NY): State University of New York Press.

²³ Hunspell (<https://hunspell.github.io/>) is the most popular spell checking library, used in all free software (LibreOffice, Firefox, etc.), but also in many closed source software such as the Adobe Suite, Mac OSX, etc. this makes our spell checker easily usable.

- Das, Sarat Chandra. 1915. An introduction to the grammar of the Tibetan language. Darjeeling: Darjeeling Branch Press.
- Dorje, Pema; and Hanfutin; and Drakpa. 2003. དག་ཡིག་གསར་བསྐྱོག། (The new spelling-dictionary). Xining: མཚོ་ཕོ་མོ་མི་རིགས་དཔེ་སྐྱུར་ཁང་།.
- Duff, Tony. 2000. The Illuminator Tibetan-English encyclopedia dictionary. Kathmandu: Padma Karpo Translation Committee.
- Dzongkha Development Commission. 2010. ཚོས་སྐད་རྫོང་ཁ་ཚོག་མཛོད་ཆེན་མོ། (A comprehensive Chöke-Dzongkha dictionary), at <http://www.dzongkha.gov.bt/en/publications/title/a-comprehensive-ch-ke-dzongkha-dictionary>
- Dzongkha Development Commission. 2011. བསྐྱོག་ཡིག་གསལ་གྱི་ཨ་འོང་། (A Handbook of Dzongkha and Chöké Abbreviations), at <http://www.dzongkha.gov.bt/en/publications/title/a-handbook-of-dzongkha-and-ch-k-abbreviations>
- Dzongkha Development Commission. 2010. གཞི་རིམ་རྫོང་ཁའི་བརྗོད་གཞུང་། (Basic Level Dzongkha Grammar Textbook), at <http://www.dzongkha.gov.bt/dz/publications/title/basic-level-dzongkha-grammar-textbook>
- Hildt, Hélios. 2016. Towards describing Tibetan syntax: From word segmentation to rewrite rules through a semi-automatic workflow. *Himalayan Linguistics* 15: 78–112.
- Hill, Nathan. 2010. A lexicon of Tibetan verb stems as reported by the grammatical tradition. Munich: Bayerische Akademie der Wissenschaften.
- Kirtivajra, 1982. བོད་ཉལ་གྱི་བརྗོད་ཡིག་མིན་ཚོག་དོན་གསུམ་གསལ་བྱེད། (Tibetan-Mongol dictionary, the elucidation of the three configurations of words) [Reprint in: Four Tibetan-Mongolian Lexicon. New Delhi: Sharada Rani], at <https://www.tbrc.org/#!rid=W00KG09211>
- Negi, J.S. 1993. བོད་སྐད་དང་ལེགས་སྐྱར་གྱི་ཚོག་མཛོད་ཆེན། (Tibetan-Sanskrit dictionary). Sarnath: Central Institute of Higher Tibetan, at <http://tbrc.org/link?RID=W1KG5422>
- Tournadre, Nicolas; and Dorje, Sangda. 2005. Manual Of Standard Tibetan: Language and civilization. Ithaca (NY): Snow Lion.
- Tsering, Tashi. 2006. Standardizing Tibetan terms of information technology. The China Tibetology Research Center, at http://digitaltibetan.org/images/e/e0/Standardizing_Tibetan_Terms_of_IT-China_Tibetology_Research_Center-2006.pdf
- Yisun, Zhang. 1985. བོད་རྒྱ་ཚོག་མཛོད་ཆེན་མོ། (Great Sino-Tibetan Dictionary). Beijing: མི་རིགས་དཔེ་སྐྱུར་ཁང་།, at <http://tbrc.org/link?RID=W29329>

- མཁམ་སྟོན་རྗེ་དབང་ལྷན་པུ་. 1979. ཏུས་གསུམ་རེའུ་མིག་ལྷ་མིའི་དགོངས་གཏེར། (The Treasure of Thomi's Insight, The Complete Verbal Forms). New Delhi: C.T. Khrto, at <http://tbrc.org/link?RID=W1KG14490>
- ཚེ་ཏན་འབས་བྱུང་།. 2003. བོད་གངས་ཚན་གྱི་སྐྱེ་རིག་པའི་བསྟན་བཅོས་ལེ་ཚན་འགའ་ཕྱོགས་བསྟུན། (A compilation of a few themes from the grammatical treatises of the Land of Snow). Xining: མཚོ་ཕྱོན་མི་རིགས་དཔེ་སྟེན་ཁང་།, at <http://tbrc.org/link?RID=W1KG25350>
- ཚོས་ཀྱི་འབྲུང་གནས། (Ed.), 1721. བཀའ་འགྲུའ། (Tibetan Buddhist canon), ཟླ་དགེ། [Reproduced: Esukhia, 2012-2018], at <https://github.com/Esukhia/derge-kangyur>
- དུང་དཀར་སློབ་ཐབས་འཕེན་ལས།, 2002. དུང་དཀར་ཚིག་མཚོ་ཆེན་མོ། (Dungkar's great dictionary). Beijing: China Tibetology Publishing House, at <http://tbrc.org/link?RID=W26372>
- ཕྱོན་ལམ།. 2016. ཕྱོན་ལམ་ཚིག་མཚོ་ཆེན་མོ། (Monlam Grand Tibetan dictionary), at <http://www.monlamit.com/>
- ས་ཚན་ཀུན་དགའ་སྟོང་པོ།. 1092. ཡི་གེའི་བཞག་ཐབས་བྱེས་པ་བདེ་ལྷག་ཏུ་འཇུག་པ། (A reading manual, an accessible introduction for children) [in ས་སྐྱ་བཀའ་འབྲུམ། (The collected works of Sakya masters), ཟླ་དགེ།. 1736. Reprint in: Dehradun. 1992], at [http://tbrc.org/link?RID=O01CT0026|O01CT00264CZ121847\\$W22271](http://tbrc.org/link?RID=O01CT0026|O01CT00264CZ121847$W22271)
- སི་ཏུ་པམ་ཆེན་ཚོས་ཀྱི་འབྲུང་གནས།, XVIIIth Century. ལུས་གངས་ཚན་པའི་བརྒྱུད་ལུང་དག་པར་སྦྱོར་བའི་བསྟན་བཅོས་ཀྱི་བྱེ་བྲག་སུམ་རྩ་བ་དང་ཉུགས་ཀྱི་འཇུག་པའི་གཞུང་གི་རྩམ་པར་བཤད་པ་མཁམ་པའི་མགུལ་རྒྱན་སྐྱེ་ཏེག་ཐེང་མངོས། (The beautiful rosary of pearls ornamenting the throat of scholars, a complete presentation of the thirty root verses and the guide to signs, two treatises showing the correct application of Tibetan grammar) [Reprint in: Xining: མཚོ་ཕྱོན་མི་རིགས་དཔེ་སྟེན་ཁང་།. 2001), at <http://tbrc.org/link?RID=W22787>

Elie Roux
elie.roux@telecom-bretagne.eu

Hélios Hildt
hhdрупchen@gmail.com

APPENDIX: The list of valid roots and exceptions

The following pages contain the complete lists of valid roots and exceptions, easily deducible from this article. They are in a simple form:

RootOrSyllable/PropertySuffix

Where PropertySuffix is:

- A if any vowel + suffix or affixed particle can appear after the root, with the exception of the suffix འ
- NB if any vowel + suffix or affixed particle can appear after the root, but at least one has to appear
- C if only affixed particles can appear after the root or syllable

Note that we take special suffixes into account separately.

As an example, ལ/A means that all the following possibilities are valid:

ལ, ལག, ལགས, ལང, ལངས, ལད, ལན, ལབ, ལབས, ལམ, ལམས, ལཔ, ལཔའི, ལཔའོ, ལཔོ, ལའང, ལའམ, ལར, ལས,
 ལེ, ལེག, ལེགས, ལེང, ལེངས, ལེད, ལེན, ལེབ, ལེབས, ལེམ, ལེམས, ལེཔ, ལེཔའི, ལེཔའོ, ལེཔོ, ལེའང, ལེའམ, ལེར, ལེས,
 ལུ, ལུག, ལུགས, ལུང, ལུངས, ལུད, ལུན, ལུབ, ལུབས, ལུམ, ལུམས, ལུཔ, ལུཔའི, ལུཔའོ, ལུཔོ, ལུའང, ལུའམ, ལུར, ལུས,
 ལོ, ལོག, ལོགས, ལོང, ལོངས, ལོད, ལོན, ལོབ, ལོབས, ལོམ, ལོམས, ལོཔ, ལོཔའི, ལོཔའོ, ལོཔོ, ལོའང, ལོའམ, ལོར, ལོས,
 ལོ་, ལོ་ག, ལོ་གས, ལོ་ང, ལོ་ངས, ལོ་ད, ལོ་ན, ལོ་བ, ལོ་བས, ལོ་མ, ལོ་མས, ལོ་པ, ལོ་པའི, ལོ་པའོ, ལོ་པོ, ལོ་འང, ལོ་འམ, ལོ་ར, ལོ་ས

དལ/NB implies the following valid possibilities:

དལའ, དལག, དལགས, དལང, དལངས, དལད, དལན, དལབ, དལབས, དལམ, དལམས, དལཔ, དལཔའི, དལཔའོ, དལཔོ, དལའང, དལའམ, དལར, དལས, དལེ,
 དལེག, དལེགས, དལེང, དལེངས, དལེད, དལེན, དལེབ, དལེབས, དལེམ, དལེམས, དལེཔ, དལེཔའི, དལེཔའོ, དལེཔོ, དལེའང, དལེའམ, དལེར, དལེས, དལུ,
 དལུག, དལུགས, དལུང, དལུངས, དལུད, དལུན, དལུབ, དལུབས, དལུམ, དལུམས, དལུཔ, དལུཔའི, དལུཔའོ, དལུཔོ, དལུའང, དལུའམ, དལུར, དལུས, དལོ,
 དལོག, དལོགས, དལོང, དལོངས, དལོད, དལོན, དལོབ, དལོབས, དལོམ, དལོམས, དལོཔ, དལོཔའི, དལོཔའོ, དལོཔོ, དལོའང, དལོའམ, དལོར, དལོས, དལོ་,
 དལོ་ག, དལོ་གས, དལོ་ང, དལོ་ངས, དལོ་ད, དལོ་ན, དལོ་བ, དལོ་བས, དལོ་མ, དལོ་མས, དལོ་པ, དལོ་པའི, དལོ་པའོ, དལོ་པོ, དལོ་འང, དལོ་འམ, དལོ་ར, དལོ་ས

And ལུ/C implies ལུའ, ལུའི, ལུའོ, ལུའང, ལུའམ, ལུའར, ལུའས.

These lists constitute the basis of the spell checker we built and are available on our git repository.

We have tested our spell checker against the data of (Hildt, 2016) and have only found expected discrepancies, due to the treatment of syllables built on ཅ, ས, སྐ, སྒ as exceptions.

མ/A	ར/A	འ ར ལ	འལ/C	ལ/C
མལ/A	མ/A		འལ/C	ལ/C
མལ/A	བམ/A	Specials suffix:	འལ/C	ལ/C
མལ/A	པ/A		འལ/C	ལ/C
མ/A	པ/A	འལ/C	འལ/C	ལ/C
གམ/NB	གམ/NB	འལ/C	འལ/C	ལ/C
བམ/NB	བམ/NB	འལ/C	འལ/C	ལ/C
མ/A	མ/A	འལ/C	འལ/C	ལ/C
མ/A	མ/A	འལ/C	འལ/C	ལ/C
བམ/A	མ/A	འལ/C	འལ/C	ལ/C
བམ/A	གམ/NB	འལ/C	འལ/C	ལ/C
མ/A	བམ/NB	འལ/C	འལ/C	ལ/C
མམ/NB	བམ/A	འལ/C	འལ/C	ལ/C
འམ/NB	བམ/A	འལ/C	འལ/C	ལ/C
འ/A	མ/A	འལ/C	འལ/C	ལ/C
མའ/NB	མ/A	འལ/C	འལ/C	ལ/C
འའ/NB	མ/A	འལ/C	འལ/C	ལ/C
འ/A	མ/A	འལ/C	འལ/C	ལ/C
བའ/A		འལ/C	འལ/C	ལ/C
མ/A	Exceptions:	འལ/C	འལ/C	ལ/C
མ/A		འལ/C	འལ/C	ལ/C
གམ/NB	དམའ	འལ/C	འལ/C	ལ/C
བམ/NB	མའ	འལ/C	འལ/C	ལ/C
མ/A	མ/C	འལ/C	འལ/C	ལ/C
མ/A	མའ	འལ/C	འལ/C	ལ/C
གམ/NB	མ/C	འལ/C	འལ/C	ལ/C
བམ/NB	མ/C	འལ/C	འལ/C	ལ/C
བམ/A	མའ	འལ/C	འལ/C	ལ/C
འ/A	མ/C	འལ/C	འལ/C	ལ/C
ལ/A	མའ	འལ/C	འལ/C	ལ/C
གལ/NB	མའ	འལ/C	འལ/C	ལ/C

