himalayan linguistics

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

Himalayan Linguistics

Tibetan trisyllabic light verb construction recognition

Weina Zhao **Lin Li**
Qinghai Normal University

Huidan Liu **Jian Wu**
Chinese Academy of Sciences

ABSTRACT

The Tibetan trisyllabic light verb construction is a type of widely used verb phrase that is composed of a disyllabic noun or adjective and a light verb. A large number of Tibetan trisyllabic light verb constructions are widely found in Tibetan. Successfully recognizing this type of phrase greatly contributes to Tibetan information processing, however, thorough and systematic academic research in this field has not yet been launched. Therefore, we propose a model for the recognition of Tibetan trisyllabic light verb constructions based on an integrated strategy in this paper. Firstly, we extract all trisyllabic light verb construction candidates from a Tibetan corpus. In this step, light verbs are used as retrieval marks. Secondly, we filter candidates using a statistics-based model, rule-based model, and integrated model separately. Experimental results show that the integrated model performs much better than the other strategies, which proves that linguistic features contribute a lot to the automatic recognition of Tibetan trisyllabic light verb constructions by computers.

KEYWORDS

Tibetan, NLP, light verbs

This is a contribution from *Himalayan Linguistics, Vol. 15(1): 137–148.*
ISSN 1544-7502

Himalayan Linguistics, Vol. 15(1). © Himalayan Linguistics 2016
ISSN 1544-7502

Tibetan trisyllabic light verb construction recognition

Weina Zhao **Lin Li**
Qinghai Normal University

Huidan Liu **Jian Wu**
Chinese Academy of Sciences

1 Introduction

Already in Old Tibetan, trisyllabic light verb constructions (hereafter referred to as TTLVCs) are to be found in historical documents, however, in small quantity. As economy and culture develop day by day in Tibetan society, TTLVCs are widely distributed in modern Tibetan texts (Jiang 2005). TTLVCs consist of two parts: one is a disyllabic noun or adjective, the other is a light verb. TTLVCs recognition plays an important role in fields including text segmentation, chunking, and parsing for Tibetan. Until now, there is no effective strategy to recognize and extract TTLVCs from Tibetan due to their particular distinguishing features including openness, productivity, quantitative uncertainty, and structural instability. Therefore, it is worthy to explore a practical and effective method to recognize TTLVCs.

Many studies have been conducted in collocation extraction, chunking, and multi-word expression identification. All these research achievements provide a valuable foundation for TTLVC recognition. Sun (2002) employs four different methods including mutual information, t-test, Chi-squared test, and likelihood ratio in collocation extraction and compares their performance. Church and Hanks (1990) extracts English collocations from a corpus by calculating their mutual information. Wang (2005) proposes a Chinese verb-object phrase recognition algorithm that combines mutual information with entropy. Sun (1997) builds up a quantitative evaluation system with three dimensions to systematically analyze Chinese phrases that consist of the specific word "能力 (ability)". Qu (2004) extracts Chinese collocation candidates from a large corpus with a statistical method, and then filters correct collocations from candidates according to language rules. Jiang (2007) focuses on chunking in Chinese using a combination of statistic-based and rule-based methods, introducing a rule bank to his algorithm.

Current studies in TTLVCs mainly take a morphological perspective. A few researchers have studied TTLVC extraction based on their linguistic knowledge. Jiang (2005) analyzes the composition, structure, and syntactic function of TTLVCs. Hu (2005), Wang (1994), Gesang (1987), and Hu (1994) study the classification and structure of TTLVCs. According to these linguistic studies in Tibetan, Long (2007) proposes an algorithm to recognize TTLVCs that consists of three particular light verbs. In this paper, we propose an integrated model to extract TTLVCs from

unprocessed Tibetan corpora. Our model adopts a rule-based strategy and a statistic-based strategy including mutual information and maximum entropy methods.

2 Features of Tibetan trisyllabic light verb constructions

From a structural point of view, TTLVCs contain two parts a disyllabic word (hereafter referred to as *Dw*) and a light verb (hereafter referred to as *Lv*). It is worth noticing that some elements are inserted into TTLVCs in some circumstances. These inserted elements (hereafter referred to as *Ie*) include adverbs, classifier and so on. As a whole, a TTLVC can be expressed as formula 1.

$$TTLVC = Dw + Ie + Lv \quad (1)$$

In most cases, a *Dw* is a disyllabic noun or an adjective, and a *Lv* is a monosyllabic light verb. The meaning of a TTLVC is largely determined by the *Dw*.

2.1 Monosyllabic light verbs

According to their syllable length, Tibetan verbal phrases[1] can be classified into three categories: monosyllabic verb, trisyllabic verb phrase, and multisyllabic verb phrase. In Tibetan, the majority of verbs are monosyllabic, which has long history and high frequency. Lexical meanings of several monosyllabic verbs have slowly evolved, so these verbs are referred to as monosyllabic light verbs. To be specific, *Lv* only retains its general meaning as an action such as meanings of do, make, commit, or conduct.

Based on origin, monosyllabic light verbs can be divided into four categories.

(1) Light verbs that evolve from volitional verbs.

Some widely used light verbs such as གཏོང་། ཆེད།, and རྒྱག originate in volitional verbs and form TTLVCs like ཁ་རྡུང་གཏོང་། (to denounce) or གནང་སློལ་ཆེད། (to reclaim).

(2) Light verbs that evolve from nonvolitional verbs.

Light verbs such as ཟེབས།, རོར།, ཡོང་།, and འཐོབ evolve from nonvolitional verbs that present objective results such as a TTLVC འཛེལ་རྒྱས་འཐོབ (to develop).

(3) Light verbs that originate from honorific morphemes. For instance, སློབ་སྦྱོང་གནང་། (to learn) consists of སློབ་སྦྱོང་ (learn) and a honorific morpheme གནང་། (to do).

(4) Aside from the aforementioned *Lv*, light verbs that originate from humilific morphemes like ཞུ also can form TTLVCs, for instance ཕྱག་རོགས་ཞུ (to help).

Tibetan light verbs exhibit stem alternation according to tense[2], which persists when they are used as part of TTLVCs. For instance, གནང་སློལ་ཆེད། (to reclaim) have three forms that are གནང་སློལ་བྱ།, གནང་སློལ་བྱེས།, and གནང་སློལ་བྱས།. These changes of tense of TTLVC are taken into consideration in our recognition model.

---

[1] In the current context we use 'verb phrase' as a more simply worded equivalent of 'verbal syntagma'; we do not intend 'verb phrase' in the generative sense.

[2] Tibetan verb stems are capable of distinguishing up to four forms: present (*lda da ba*), past (*'das pa*), future (*ma 'ongs pa*), and imperative (*skul tshig*). Terminological comfort might dictate that we refer to the imperative as a 'mood', but from the perspective of Tibetan morphology there is no reason to do so.

2.2 Disyllabic words

Hu (2002) analyzes the composition and structure of disyllabic words that form TTLVCs. He proposes a nine category morpheme pattern of disyllabic words that are V+V, N+V, D+V, A+V, V+N, R+V, N+N, N+A, A+A. V represents a light verb, N represents a noun morpheme, D presents an adverb morpheme, and A presents an adjective morpheme. The pattern V+V means that the disyllabic word is composed by two light verbs. For instance, in a TTLVC འཛེལ་རྒྱས་འགྲོ (to develop), འཛེལ and རྒྱས are both light verbs.

2.3 Inserted elements

The structure of TTLVCs are unstable because it is possible for some elements to be inserted between *Dw* and *Lv*. Inserted elements can be adverbs, adjectives, case-auxiliary words, enumerating auxiliary words, imperative auxiliary words, and so on. For instance, ཨ་ལོག་ཅིག་བརྒྱབ (to turn over once) consists of an indefinite article ཅིག and a TTLVC ཨ་ལོག་བརྒྱབ (to turn over), and བསམ་སྤྱོ་ཡག་པོ་བཏང་། (to think it over) consists of an adjective ཡག་པོ and a TTLVC བསམ་སྤྱོ་བཏང་། (to consider or to think).

Aside from the discontinuous TTLVCs discussed immediately above, there is another type of discontinuous TTLVC, namely one that is composed of more than one disyllabic word and one light verb. For instance, the TTLVC སློབ་སྦྱོང་དང་ཞིབ་འཚུག་བྱེད (to learn and study) is composed of two nouns སློབ་སྦྱོང (learn) and ཞིབ་འཚུག (study), in the middle of these two nouns is a conjunction དང (and), and a light verb བྱེད (to do).

3 TTLVC candidates extraction

In this section, we introduce how to extract TTLVC candidates from raw Tibetan corpora according to the linguistic features of TTLVC discussed above. We start with building up a light verb list. A Tibetan and Chinese bilingual Lhasa Spoken Language Dictionary (Yu 1983) contains a number of TTLVCs. We extract light verb constructions from this dictionary, and integrate the results with light verb constructions listed in other works (viz. Jiang 2005; Wang 1994; Hu 1994; Zhou 2003). As result, we acquire a light verb list with 875 entries.

Our light verb list includes as information the light verbs themselves and the number of occurrences of each light verb. A few examples are shown in Table1, and please see Appendix 1 at the end of this paper for a complete light verb list adopted in this work.

ID	Light Verb	Occurrence Number	ID	Light Verb	Occurrence Number
1	བྱེད	1027	2	རྒྱག	569
3	ཞུ	134	4	བཟོ	111
5	ཤོར	109	6	བཤད	84
7	ལེན	70	8	འགྲོ	61
9	སློག	49	10	བཞག	46

Table 1. Examples of Light verbs

According to the light verb list, we extract TTLVC candidates from a corpus. With regard to the inserted elements in TTLVCs, we work out TTLVC candidates using an extraction algorithm, the main procedures of which are listed in Tabel2.

(1) Split a Tibetan sentence by *tsheg* punctuation mark into syllables.
(2) To scan syllables obtained in step 1.
(3) If a *Lv* is found go to step 4, otherwise go back to step 1.
(4) If the syllable on the left of the *Lv* is an inserted element, extract two syllables on the left of the inserted element. These two syllables and the *Lv* compose one TTLVC candidate.
(5) If the syllable on the left of the *Lv* is not an inserted element, extract two syllables on the left of the *Lv*. These two syllables and the *Lv* compose one TTLVC candidate.
(6) If the *Lv* is at the end of a sentence go back to step 1, otherwise go back to step 2.

Table 2. TTLVC candidates extraction procedures

4 Statistic-based TTLVC recognition model

TTLVCs have two features: (1) although a TTLVC might be split by an inserted element, the elements of a TTLVC have a higher than average co-occurrence in corpus; (2) there is a clear boundary between the TTLVC and its context. Therefore, we propose two measures to grade a candidate. One is an internal measure and the other is an external measure. Several statistical methods are used to grade TTLVC candidates (Jiang 2007).

4.1 Internal Measure

From a statistical perspective, if the three syllables of a TTLVC candidate have a high co-occurrence frequency, they possess a strong degree of internal connectivity. Based on this theory, the higher the co-occurrence frequency the more likely they are to be a word or phrase, i.e. a true TTLVC. Mutual Information is usually applied to evaluate the level of co-occurrence probability of two words. The higher the mutual information of the two words is, the stronger internal connectivity they have. In this paper, we propose an internal measure formula based on the mutual information definition shown in forumal2.

$$InMeasure(W) = \log_2 \frac{P_{XY}}{P_X P_Y} \tag{2}$$

P_{XY} presents the co-occurrence probability of *Dw* and *Lv*. P_X means the occurrence probability of *Dw* and P_Y is the occurrence probability of *Lv*.

4.2 External measure

Normally, we can identify the external boundary of a TTLVC by observing its context. In our work, a maximum entropy model is chosen to measure the external independence of a TTLVC. We use the model to evaluate the left and right boundary of a candidate separately. According to information entropy theory, a TTLVC candidate with higher information entropy is free to occur in various contexts. If a candidate stably appears in a variety of context, it has a higher probability of being a TTLVC.

In this paper, *Le(W)* and *Re(W)* are proposed to evaluate the external measure of a TTLVC, as shown in formula 3 and formula 4.

$$Le(W) = -\sum_{\forall a \in A} P(aW|W) \times \log_2 P(aW|W) \tag{3}$$

$$\mathrm{R}e(W) = -\sum_{\forall b \in B} P(Wb|W) \times \log_2 P(Wb|W) \tag{4}$$

In formula 3, *W* is a TTLVC candidate, *Le* is the left boundary entropy; *Re* is the right boundary entropy; and *A* is a collection of syllables that appear on the left of candidates; *a* is a specific syllable on the left of the candidate; *B* is a collection of syllables that appear on the right of our candidates; *b* is a specific syllable on the right of the candidate.

We can combine formula 3 and formula 4, to yield a more general external measure as in formula 5.

$$ExMeasure(W) = \sqrt{(1 - Le(W))(1 - \mathrm{Re}(W))} \tag{5}$$

4.3 Integrated algorithm based on internal measure and external measure

To improve precision and accuracy of a single statistic-based approach employing both the internal and external measures discussed so far, we propose a comprehensive formula (formula 6) that integrates internal measure and external measure. Additionally, a candidate has a higher probability of being a TTLVC, if it has a relatively high frequency in corpus. Therefore the frequency of the syllable sequence in question *F(W)* has been introduced into our algorithm.

$$UniMeasure(W)=(1-1/F(W)) \times InMeasure(W) \times ExMeasure(W) \tag{6}$$

5 Rule-based TTLVC recognition model

Observing the Tibetan corpus, we find that a high frequency sequence of syllables may not be a TTLVC. Take the syllable sequence བར་ངོ་ཆོལ། as an example. In our corpus, བར་ངོ་ཆོལ། appears in various contexts, for instance, it is often used in ཡར་ཐོན་རྒྱུན་འཕྲོངས་ཀྱིས་ལྡག་ལོག་རྒྱག་བར་ངོ་ཆོལ། (Persist in progress and oppose retrogression!) Due to its high frequency, it is identified as a TTLVC by our statistic-based model. Actually, it is just a sequence of three syllables, not even respecting word boundaries, not to mention being a TTLVC. In fact, it is quite easy to rule out the possibility of བར་ངོ་ཆོལ། being a TTLVC according to linguistic knowledge. The structure of བར་ངོ་ཆོལ། is "ར། (nominalization marker) + ར་ (case marker) + ངོ་ཆོལ།". Because nominalization markers appear at the end of words, there cannot be a word break before a nominalization marker, so we know that བར་ངོ་ཆོལ། is not a TTLVC. Our statistic-based model does not perform very well in such cases; however, rule-based approach can deal with them effectively. As a result, we build up a linguistic knowledge rule bank to make up for the deficiency.

We make a rule to exclude the candidates such as བར་ངོ་ཆོལ།. For instance, if a *Dw* contains a case marker or a nominalizations marker, then the candidate should be removed. According to

Tibetan grammar, *Dw* are often nouns. If case markers appear in a *Dw*, there is a low chance the syllable sequence is a TTLVC. The rules are relative rather than absolute, thus the rule bank can be added, deleted, and modified according to needs. Several rules of our rule bank are listed in Table 3.

If a candidate meets one of the following rules, remove it from candidate list.
(1) The first syllable of a TTLVC candidate is a case marker, a nominalizations marker, an adjective suffix, or a negative adverb.
(2) The second syllable of a TTLVC candidate is a case marker.

Table 3. TTLVC filter rule sample

6 Experiments and Analysis

6.1 Experiments and Results

We test our TTLVC recognition model proposed in this paper using a corpus that composed of Tibetan publications. Our corpus contains fifty thousand sentences, whose genre includes government documents, news reports, legal texts, and so on. Based on the same corpus, we conduct three groups of experiments that adopt different methods shown as follow.

· Statistic-based Method 1: Mutual Information

· Statistic-based Method 2: Statistic-based Method1 + Left/Right Entropy

· Integrated Method: Statistic-based Method2 + Rule Bank

Considering the importance of light verbs, we build up a couple of light verb lists according to their number of occurrences (*F*). To choose an optimal light verb list, we conduct several experiments based on different lists, and the precision (*P*) of experiments are shown in Table 4.

P / *F*	Statistic-based Method1	Statistic-based Method2	Integrated Method	Average Precision
>100	0.791	0.858	**0.918**	0.855
>50	0.787	0.819	0.906	0.837
>10	**0.758**	**0.803**	**0.879**	**0.813**
>0	0.700	0.717	0.865	0.760

Table 4. Results of TTLVC recognition experiments based on different light verb lists

The results of the experiments suggest that the scope of the light verb list plays a significant role in TTLVC recognition. When we employ a relatively small light verb list, we acquire the best precision 91.8% in the TTLVC extraction experiment. In this experiment we only employ the light verbs whose number of occurrence is greater than 100. But a large number of TTLVC candidates have been discarded because of the smaller light verb list. Thus we choose light verbs whose frequency is more than ten to build up a more thorough list of light verbs. In addition, all tenses of these light verbs are also added into our list. Our final list contains 210 entries.

Based on the final list, 65,764 TTLVC candidates have been extracted from the corpus. By morphological analysis and removing candidates whose occurrence frequency is lower than three, we retain 11,243 candidates.

According to the experimental results, the precision of Method1 is 75.8%, and it reaches 80.3% by introducing Left/Right Entropy. When the integrated method is adopt, the precision increases 7.6% and reaches 87.9%. These results prove that the integrated method performs the best in the TTLVC task. TTLVC samples recognized by our model are listed in Table 5.

Examples	Meaning	Number of Occurrence
དོ་ཚོལ་བྱེད།	to resist	1,490
རྒྱལ་སྤེལ་གཏོང་།	to stimulate	240
སེམས་ཧག་གཅོད།	to determine	199
འཕེལ་རྒྱས་འགྲོ	to make progress	561
འཕེལ་རྒྱས་གཏོང་།	to develop	138
དབྱེ་བ་འབྱེད།	to distinguish	79
རྩལ་ར་འདོན།	to play a part in	351
རྩོར་འཁྲུལ་ཤོར།	to make a mistake	48
བྱ་ལོན་འཇལ།	to repay a debt	10

Table 5. TTLVC examples

6.2 Error analysis

In general, our integrated algorithm performs well in TTLVC recognition as shown in table 4. Table 5 lists actual TTLVC identified by our system. However, our algorithm makes some mistakes, most of which fall into three categories.

1. Misidentifying multi-syllabic verb phrases as TTLVCs. Aside from TTLVCs, there exist a large number of multi-syllabic verbal phrases such as four syllabic or five syllabic verbal phrases. Our algorithm falsely extracts the last three syllables of these phrases as TTLVCs. For instance, གནས་ཚུལ་ལེགས་འགྱུར་ (the situation has been changed) is a four syllable verbal phrase, and our algorithm extracts ཚུལ་ལེགས་འགྱུར་ and take it as a TTLVC. Because these multi-syllabic verbal phrases have a high frequency in our corpus, and they conform with linguistic rules related to TTLVC, it is very difficult to correct this type of error for our model.

2. Misidentifying some multi-word expressions as TTLVCs. There are many tri-syllabic multi-word expressions with high occurrence frequency in Tibetan, but which are not TTLVCs. For instance, ཇི་ཚུར་བྱེད་ (How should something be done?) is a sentence and it occurs 19 times in our corpus.

3. Misidentifying an adjective or adverb and a light verb as TTLVCs.

It is possible that TTLVCs are separated by inserted elements such as an adjective or an adverb. In some circumstances, our algorithm incorrectly recognizes inserted element and the light verb behind them as TTLVCs. The reason why our algorithm makes this kind of mistake is inserted elements probably co-occur with light verbs in corpus several times. Therefore, their co-occurrence makes our statistically based algorithm unable to get rid of these mistakes.

For instance, གཙེག་སྡུད་ཀ་ཅང་ཆེན་པོ་བྱེད་ (to intensively concentrate on) consists of a TTLVC གཙེག་སྡུད་བྱེད་ (to concentrate on) and an adverb ཀ་ཅང་ཆེན་པོ་ (intensively). Our algorithm, however, misidentifies ཆེན་པོ་བྱེད་ as a TTLVC.

7 Conclusion

In this paper, we propose an integrated approach to identify Tibetan trisyllabic light verb constructions from a raw corpus. Experimental results show that a statistics-based strategy that combines mutual information and entropy performs well in TTLVC recognition. Accuracy and precession of our model have been largely improved when a rule-based method is introduced into our algorithm. The result proves that a multi-pronged strategy is more effective than a unitary strategy. In the future, we plan to apply our algorithm to other types of Tibetan phrase structure extraction problems such as the identification of noun phrases.

Acknowledgements

This research was supported by the Qinghai Natural Science Foundation under Grant 2015-ZJ-923Q，Ministry of Education under Grant Z2015066 and Z2015067, and National Natural Science Foundation of China under Grand 61550004.

References

Church, Kenneth Ward; and Hanks, Patrick. 1990. "Word association norms, mutual information, and lexicography". *Computational Linguistics* 16.1: 22-29.

Gesang Jumian. 1987. *Practical Tibetan grammar*. Chengdu: Sichuan Minzu Press. (格桑居冕.实用藏文文法[M]. 成都: 四川民族出版社 1987.)

Hu, Tan. 2002. *Tibetan studies.* Beijing: China Tibetology Publishing House. (胡坦.藏语研究论文[M]. 北京: 中国藏学出版社, 2002.)

Jiang, Binggui; Zhang, Qinlong et al. 2007. "Chinese Multi-word Chunks Extraction for Computer Aided Translation". *Journal of Chinese Information Processing* 21.1: 9-16. (姜柄圭, 张秦龙, 谌贻荣,等. 2007. "面向机器辅助翻译的汉语语块自动抽取研究". 中文信息学报, 21(1): 9-16.)

Jiang, Di; Kong, Jiangping. 2005. *New progress of Chinese minority language engineering*. Beijing: Social Sciences Literature Press. (江荻,孔江平.中国民族语言工程研究新进展[M]. 北京: 社会科学文献出版社, 2005.)

Long, Congjun. 2007. "Tibetan trisyllabic verb analysis and detection". *Proceedings of Research on the information technology of the national language and character*, 548-555. (龙从军. "藏语三音动词分析及自动识别方法.民族语言文字信息技术研究". 第十一届全国民族语言文字信息学术研讨会论文集.中国云南西双版纳傣族自治州. 2007: 548-555.)

Qu, Weiguang; Chen, Xiaohe; and Ji, Genlin. 2004. "A Frame-based Approach to Chinese Collocation Automatic Extracting". *Computer Engineering* 30.23: 22-24. (曲维光, 陈小荷, 吉根林. 2004. "基于框架的词语搭配自动抽取方法". 计算机工程 30(23): 22-24.)

Hu, Shujin. 1994. *Concise Tibetan grammar*. Kunming: Yunnan Minzu Press, 75-89. (胡书津.简明藏文文法.昆明:云南民族出版社. 1994: 75-89.)

Sun, Jian; Wang, Wei; Zhong, Yixin. 2002. "Methods of Finding the Collocation Based on Statistics." *Journal of the China Society for Scientific and Technical Information* 21.1: 12-16. (孙健, 王伟, 钟义信. 2002. 基于统计的常用词搭配(Collocation)的发现方法. 情报学报, 21(1): 12-16.)

Sun, Maosong; Huang, Changning; and Fang, Jie. 1997. "A preliminary study on the quantitative analysis of Chinese Collocation". *Studies of the Chinese Language* 1: 29-38. (孙茂松,黄昌宁,方捷.汉语搭配定量分析初探[J].中国语文,1997(1): 29-38.)

Wang, Suge; Yang, Junling; and Zhang, Wei. 2005. "Automatic Acquisition of Chinese Collocation". *Journal of Chinese Information Processing* 20.6: 31-37. (王素格, 杨军玲, 张武. 2006. "自动获取汉语词语搭配". 中文信息学报, 20(6): 31-37.)

Wang, Zhijing. 1994. *Lhasa Tibetan grammar*. Beijing: Minzu University of China press. (王志敬.藏语拉萨口语语法[M].北京:中央名族大学出版社.1994.)

Yu, Daoquan. 1983. *Tibetan and Chinese controlled Lhasa oral English dictionary.* Beijing: Minzu Press. (于道泉. 藏汉对照拉萨口语词典. 北京: 民族出版社. 1983.)

Zhou, Jiwen; and Xie, Houfang. 2003. *Tibetan Lhasa dialect grammar*. Beijing: Minzu Press, 50-59. (周季文,谢后芳. 藏语拉萨话语法. 北京: 民族出版社. 2003: 50-59.)

Weina Zhao
zhaoweina1999@qq.com

ACKNOWLEDGEMENTS

ID	Light Verb	Number of Occurrence	ID	Light Verb	Number of Occurrence	ID	Light Verb	Number of Occurrence
1	ཕྱེད།	1027	71	ཞེད།	12	141	སངས།	6
2	རྒྱག	569	72	གསོག	12	142	ཟད།	6
3	ཞུ	134	73	ལངས།	11	143	དག	6
4	བཟོ	111	74	འབབ།	11	144	ཉལ།	6
5	མོར།	109	75	བསྒྲུབ།	11	145	ཉན།	6
6	བཤད།	84	76	སྒྲོག	11	146	བསད།	6
7	ལེན།	70	77	སྐྱེར།	11	147	བཀོག	6
8	འཕྲོ	61	78	སྤྲེལ།	11	148	སྐྱེར།	6
9	སྐོག	49	79	སྤྲས།	11	149	བསྐོར།	6
10	བཞག	46	80	ཤེབས།	11	150	ལྷོན།	6
11	ཞེབས།	43	81	དབྱུག	11	151	ཁྱེད།	6
12	ཐུག	41	82	ཟླ	10	152	འགྱངས།	6
13	གཅོད།	40	83	བཀག	10	153	འཁྱུད།	6
14	ཚགས།	39	84	བཅག	10	154	འཕོ	6
15	འཇུག་ས།	39	85	ཁྱག	10	155	འཕོལ།	6
16	ཀྱ།	36	86	འཁེལ།	10	156	སྒྲུག་ས།	6
17	ལྷོན།	33	87	འགོད།	10	157	གཞེར།	6
18	སྐྱེད།	32	88	གཏོར།	10	158	གཏོད།	6
19	ཉོན།	30	89	ཉེར།	10	159	འདེད།	6
20	འབུལ།	29	90	འཛེལ།	10	160	འབྱུ	6
21	བསྡུག་ས།	28	91	ཚོག	10	161	འཁྱུག	6
22	འཚོར།	27	92	རེས།	10	162	བཙོངས།	6
23	འདོན།	26	93	འཕར།	9	163	རྩེས།	6
24	འཐེན།	25	94	དཀྲོལ།	9	164	ཚུད།	6
25	ཚོག	25	95	སྐོར།	9	165	བཞེངས།	6
26	འདེབས།	24	96	འཁྱོལ།	9	166	གཡོག་ས།	6

27	ཤོད།	24	97	འགེལ།	9	167	གསོད།	6
28	གཞུ།	24	98	ཁྱེའུ།	9	168	བཟང་ས།	6
29	བཅད།	23	99	བཙུན།	9	169	ཤར།	5
30	ཁྱེ།	23	100	འདི།	9	170	བདག་ས།	5
31	འཇིལ།	23	101	ཚོན།	9	171	བགད།	5
32	ཚན	22	102	འཛོད།	9	172	བཞད་ས།	5
33	འཁྲུལ།	22	103	ཚེ།	9	173	ཁྱེར།	5
34	ཁྱད།	22	104	ལེའུ།	9	174	ཁྱོད།	5
35	མོད།	22	105	འཇའ།	8	175	ཁྲག	5
36	ཁྱེའུ།	21	106	འཁབ།	8	176	བཀྲ་བ	5
37	ཐོད།	21	107	འཁྱག་ས།	8	177	ཏོག	5
38	ཁྲབ།	19	108	གོན།	8	178	བཙོའུ།	5
39	ཐེ།	19	109	ཆོལ།	8	179	བཙོམ།	5
40	ཡངས།	18	110	ཐོས།	8	180	འཁྱ།	5
41	འགོག	18	111	མཐུད།	8	181	འཇོག	5
42	ཚེ།	18	112	ཐོན།	8	182	ཉོ།	5
43	འཁྲེལ།	17	113	འདོ་ས།	8	183	ཉེད།	5
44	བསྡན།	17	114	ཁྱེར།	8	184	བསྡར།	5
45	ཁྲད།	17	115	ཁྱེ།	8	185	མཐུན།	5
46	གསོ།	17	116	འབྱུད།	8	186	འཇོར།	5
47	ཁྲོལ།	16	117	བད།	8	187	དྲན།	5
48	ཁྲག	15	118	ཁྲག	8	188	བརྡབ།	5
49	ཚོད།	15	119	འཇུལ།	8	189	ཏོག	5
50	བདབ།	14	120	བཞེར།	8	190	ནོར།	5
51	ཆུལ།	14	121	ཡའ།	7	191	གནོན།	5
52	བཙོས།	14	122	རཏ	7	192	སྐབས།	5
53	ཐོབ།	14	123	ཚོམ།	7	193	སྐོས།	5
54	འཁྱེད།	14	124	འཁལ།	7	194	འཕྲུད།	5
55	འཁྱི	14	125	བཟས།	7	195	ཙོད།	5

56	བཞེས།	14	126	དབྱུག	7	196	འཚོག	5
57	གཟེད།	14	127	ཐུམ་	7	197	འདེབས།	5
58	འཚལ།	13	128	ཐུམ་	7	198	ཁྱོལ།	5
59	ཚད།	13	129	ཉེངས།	7	199	བྱལ།	5
60	བཀོད།	13	130	གཙོག	7	200	འཁྱེག	5
61	ཀུཐ	13	131	བརྩེ།	7	201	ཁྱུར།	5
62	ཁོལ།	13	132	ཁྱག	7	202	རྩོམ།	5
63	བཀོན།	13	133	རྐུབ།	7	203	འཇོམས།	5
64	འདྲེན།	13	134	འཛེན།	7	204	རྐྱག།	5
65	བསྡད།	13	135	ཁྱུར།	7	205	གཟེགས།	5
66	ཐུར།	13	136	ཏོན།	7	206	གཡོལ།	5
67	ཚུགས།	13	137	ཉེས།	7	207	ལོངས།	5
68	འཇགས།	12	138	བཅར།	6	208	བཏེག	5
69	འཇུག	12	139	བཤལ།	6	209	གསོལ།	5
70	དོན།	12	140	ལན།	6	210	སྐེད།	5