# Tibetan Multi-word Expressions Identification Framework Based on News Corpora

Minghua Nuo[1(✉)], Congjun Lun[2,3], and Huidan Liu[3]

[1] College of Computer Science-College of Software Engineering, Inner Mongolia University, Hohhot, China
nuominghua@163.com
[2] Institute of Ethnology and Anthropology, Chinese Academy of Social Sciences, Beijing, China
[3] Institute of Software, Chinese Academy of Sciences, Beijing, China
{congjun,huidan}@iscas.ac.cn

**Abstract.** This paper presents an identification framework for extracting Tibetan multi-word expressions. The framework includes two phases. In the first phase, sentences are segmented and high-frequency word-based n-grams are extracted using Nagao's N-gram statistical algorithm and Statistical Substring Reduction Algorithm. In the second phase, the Tibetan MWEs are identified by the proposed framework which based on the combination of context analysis and language model-based analysis. Context analysis, two-word Coupling Degree and Tibetan syllable inside word probability are three strategies in Tibetan MWE identification framework. In experimental part, we evaluate the effectiveness of three strategies on small test data, and evaluate results of different granularity for Context analysis. On small test corpus, F-score above 75% have been achieved when words are segmented in pre-processing. On larger corpus, the P@N (N is 800) overcomes 85%. It indicates that the identification framework can work well on larger corpus. The experimental result reaches acceptable performance for Tibetan MWEs.

**Keywords:** Tibetan Multi-word expression · Two-word coupling degree · Inside word probability

## 1 Introduction

In real-life human communication, meaning is often conveyed by word groups, or meaning groups, rather than by single words. Such word groups or multi-word expressions (MWE hereafter) can be described as *a sequence of words that acts as a single unit at some level of linguistic analysis*. MWEs are frequently used in everyday language, usually to precisely express ideas and concepts that cannot be compressed into a single word. As a consequence, their identification is a crucial issue for applications that require some degree of semantic processing (e.g. machine translation, summarization, information retrieval). Very often, it is difficult to interpret human speech word by word. Consequently, for an MT system, it is important to identify and interpret accurate meaning of such word groups, or multi-word expressions, in a source


C.-Y. Lin et al. (Eds.): NLPCC-ICCPOL 2016, LNAI 10102, pp. 16–26, 2016.
DOI: 10.1007/978-3-319-50496-4_2

language and interpret them accurately in a target language. However, accurate identification and interpretation of MWEs still remains an unsolved problem in Tibetan natural language processing research.

The Tibetan alphabet is syllabic, like many of the alphabets of India and South East Asia. Each letter has an inherent vowel /a/. Other vowels can be indicated using a variety of diacritics which appear above or below the main letter. A syllable contains one or up to seven character(s). Syllables are separated by a marker known as "tsheg", which is simply a superscripted dot. Linguistic words are made up of one or more syllables and are also separated by the same symbol, "tsheg", thus there is a lack of word boundaries in the language. Consonant clusters are written with special conjunct letters. Figure 1 shows the structure of a Tibetan word which is made up of two syllables and means "show" or "exhibition".


**Fig. 1.** Structure of a Tibetan word

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | སྐྱེ་ | སྐྱག་པོ་ | འདི་ | ཁང་པ | གོང་ཆེན་པོ་ | ཞིག་ | གཟིགས་ | སོང་ | ། |
| terday | man | rich | this | house | expensive | an | bought | did | . |

terday this rich man bought an expensive house.

**Fig. 2.** A Tibetan sentence and its translation

Tibetan sentence consists of one or more words, phrases or multi-word units. Another marker known as "shad" indicates the sentence boundary, which looks like a vertical pipe. Figure 2 shows a Tibetan sentence. It is segmented in line 2 and word by word translation is given in line 3.

In this paper, we present Tibetan MWE identification framework. The rest of this paper is organized as follows. In Sect. 2 we recall related work on multi-word expression extraction methods. Section 3 describes the outline of our framework. We propose the details of framework in Sect. 4. Then, in Sect. 5 we make experiments for evaluation. Section 6 concludes the paper.

## 2 Related Work

The issue of MWE processing has attracted much attention from the Natural Language Processing (NLP) community, including [1–13]. Study in this area covers a wide range of sub-issues, including MWE identification and extraction from monolingual and multilingual corpora, classification of MWEs according to a variety of viewpoints such as types, compositionality and alignment of MWEs across different languages. Directly related to our work is the development of a statistical MWE tool at Lancaster for searching and identifying English MWEs in running text [14, 15] Trained on corpus data in a given domain, this tool can automatically identify MWEs in running text or extract MWEs from corpus data from the similar domain. It has been tested and compared with an English semantic tagger [16] and was found to be efficient in identifying domain-specific MWEs in English corpora, and complementary to the

semantic tagger which relies on a large manually compiled lexicon. Extraction of Chinese multi-word expressions from corpus resources as part of a larger research effort to improve a machine translation (MT) system is reported in [17].

However, Tibetan MWE processing still presents a tough challenge, and it has been receiving increasing attention. In Tibetan information processing, the shortage of Tibetan language resource leads to the fact that most of the techniques related text processing are still developing. Recently, the focus of Tibetan information processing is gradually transferred from word processing to text processing. The Tibetan text processing started in the early 1990s, mainly analyze statically at the beginning. Since 2003, research on Tibetan syntactic chunks [18–20] is reported. Since 2010, Nuo et al. do research on chunk, multi-word equivalence for Chinese-Tibetan machine translation system. Nuo et al. [21] construct Chinese-Tibetan multi-word equivalence dictionary for Chinese-Tibetan computer-aided translation system. They present an identification framework for extracting Tibetan base noun phrase in [22]. So far, there is no Tibetan parser. We have built large scale Tibetan text resources recently, and we are tagging Part-Of-Speech and labeling role right now, these corpora can form our training set and test data. This paper presents identification of Tibetan MWEs using statistical methods.

## 3 Brief Description of Tibetan MWE Identification Framework

The proposed Tibetan MWE identification framework consists of three main steps: pre-processing step, context analyzing step, and language model-based analysis for candidate n-grams, which are in boldface in Fig. 3. The two-word coupling degree dictionary and Tibetan syllables inside word probability dictionary are trained from annotated training corpus.

In pre-processing step, Tibetan corpus is word segmented and stored one sentence per line. High-frequency strings are extracted using Nagao's algorithm [23] and Substring Reduction Algorithm [24]. They are initial candidate MWE. These candidates determined to be a MWE based on their internal structure, pragmatic environment in the text and semantic features.

In the context analyzing step, we use adjacent characteristic to capture pragmatic environment in the text. We will calculate adjacent features such as adjacent categories, adjacent pair categories, adjacent entropy etc., if the result is lower than threshold, the candidate n-gram will be filtered as a noise; if higher, goes to the next step.

The final step is language model-based analysis step. Coupling Degree is used to measure internal formation of a MWE; it can help us to examine whether high-frequency string has a complete semantics or not. In this step, firstly, we scan the candidate n-gram string word by word, and search Coupling Degree of pair of adjacent words, if the result is less than the threshold; the word pair regarded as not a MWE but a noise and be removed. Secondly, find inside word probabilities to determine whether candidate string is started with or ended with common function words (i.e. stop words). We combine Coupling Degree of adjacent words with inside word probabilities to analyze candidate n-grams and remove the noises. Then output the remaining meaningful strings to a file, they are MWEs.

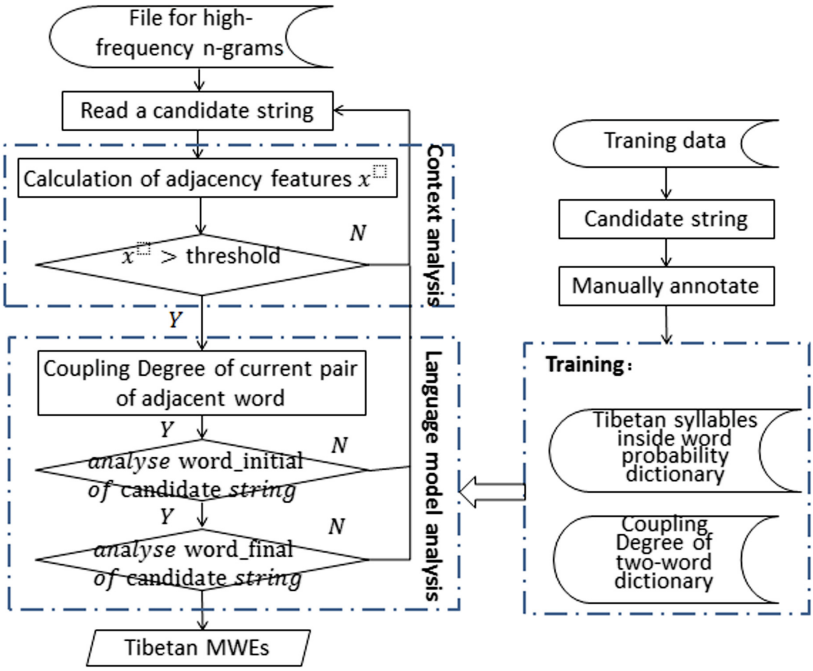


**Fig. 3.** Flow chart of Tibetan MWE identification framework

In next section, we describe in detail how to identify Tibetan MWEs. Different methods are evaluated, and we will select the method with best performance to generate referable Tibetan MWE.

## 4 Tibetan MWE Identification Based on the Combination of Context Analysis and Language Model-Based Analysis

In pre-processing stage, corpora text has been formatted and segmented. High-frequency repeated strings from large-scale corpus contain meaningful strings (i.e. MWE) as well as disturbance term (i.e. noises). The essence of extracting MWEs from corpus is to remove those noises from candidate n-grams. This section will detail the core steps of Tibetan MWE identification framework.

### 4.1 Context Analysis

Acting as a single unit, internal words in MWE are tightly related; external (or context) words of MWE are loosely related. Meaningful string as an independent language unit has a variety of different contexts in the real text. In order to describe the flexibility of the string *S*'s context, we define a series of adjacent feature measures.

**Definition 1: Adjacent Set** (abbreviated as **NS)**
Adjacent Set of a MWE are divided into Left Adjacent Set LNS and Right Adjacent Set RNS, left or right adjacent words of the string *S* in corpus constitute LNS or RNS respectively.

**Definition 2: Adjacent Categories**
Adjacent categories are divided into left and right either, respectively refer to the number of elements in LNS and RNS.

**Definition 3: Pair of Adjacent Set** (abbreviated as **PNS)**
Each occurrence of left and right context word of the string *S* constitutes an adjacent pair $<L_i, R_i>$, all adjacent pairs of the string *S* in corpus form PNS. Pair of adjacent set can indicate the complete pragmatic environment of a string.

**Definition 4: Categories of Adjacent Pair**
It denotes the number of elements in the set PNS.

**Definition 5: Adjacent Entropy**
We name entropy of the adjacent pair of string *S* as Adjacent Entropy; Entropy is the basic unit of information measure, represents the overall statistical characteristics of the uncertainty. Frequency $n_i$ denotes occurrence of each pair of adjacent $<L_i, R_i>$ in Tibetan corpus; the sum of frequencies denoted as *N*, the entropy of adjacent pair can be formulated by the following:

$$E_L = -\sum_{i=1}^{|V_L|} \frac{n_i}{n} \log(\frac{n_i}{n}) \tag{1}$$

The greater adjacent entropy is, the more flexible pragmatic environment of string *S* is; so that it is more likely to be a meaningful string. When the corpus smaller, types of adjacent is relatively small, entropy's ability to distinguish become poor.

### 4.2 Two-Word Coupling Degree

For each adjacent pair of words ($w_1$,$w_2$), the Coupling Degree (short for CD) is measured by the following formula:

$$CD(w1, w2) = \frac{VMI(w1, w2)}{H(w1) + H(w2)} \tag{2}$$

where *VMI* is a variant of average mutual information; $w_1$, $w_2$ represent occurrence of words. *VMI* is defined as follows:

$$\begin{aligned} VMI(w1, w2) = {} & P(w1, w2) \log \frac{P(w1, w2)}{P(w1)P(w2)} + P(\overline{w1}, \overline{w2}) \log \frac{P(\overline{w1}, \overline{w2})}{P(\overline{w1})P(\overline{w2})} \\ & - P(w1, \overline{w2}) \log \frac{P(w1, \overline{w2})}{P(w1)P(\overline{w2})} - P(\overline{w1}, w2) \log \frac{P(\overline{w1}, w2)}{P(\overline{w1})P(w2)} \end{aligned} \tag{3}$$

In this formula, $P(w1, w2)$ is the probability of sentences where both $w_1$ and $w_2$ adjacently occur. $P(\overline{w1}, \overline{w2})$ is the probability of sentences where both $w_1$ and $w_2$ won't occur. $P(w1, \overline{w2})$ is the probability of sentences where $w_1$ occur with other right-hand adjacent word but not $w_2$. $P(\overline{w1}, w2)$ is the probability of sentences where $w_2$ occur with other left-hand adjacent word but not $w_1$.

The denominator in $CD(w_1, w_2)$ is a smoothing factor. A high $VMI(w_1, w_2)$ value shows that $w_1$ and $w_2$ have strong tendency to appear together. It is possible that one or both of them are highly frequency words, where $H(w_1)$ and/or $H(w_2)$ have high values. Divided by this denominator, coupling degree of word pairs is decreased.

$H$ refers to the entropy of a Tibetan syllable, defined as following formula:

$$H(s) = -[P(s)\lg P(s) + P(\overline{s})\lg P(\overline{s})] \tag{4}$$

### 4.3 Tibetan Syllable Inside Word Probability

Tibetan each syllable has its own unique word-formation usage; certain syllables are often in one or a few specific location (word-initial, word-medial, word-final) on compound words. This paper focuses on word-initial and word-final syllable and their probabilities to be a word.

**Definition 6: inside word probability (short for IWP)**
IWP is the probability of a sequence of two or more Tibetan syllables being a sequence of independent MWE. IWP is defined as follows:

$$P_{word}(c, pos) = \frac{N(c, pos)}{N(c, word)} \tag{5}$$

where value range of pos is 0 and 1; 0 indicates word-initial and 1 indicates word-final.

Makes statistics for $N$, $N_1$, $N_2$ of each syllable on word segmented corpus. $N$, $N_1$, $N_2$ denotes the total number, the number of occurrence in the word-initial and word-final position respectively; then word-initial IWP is the ratio of $N_1$ and $N$, word-final IWP is the ratio of $N_2$ and $N$.

Generally, a MWE begins with word-initial syllable of one word and must ends with word-final syllable of another word. When too low word-initial IWP is detected for the first syllable of a string, it might be noise. Similarly, when too low word-final IWP is detected for the last syllable of a string, we can regard it as a noise. This rule can effectively filter out disturbance term.

This comprehensive statistical filtering measure for n-gram syllable string is able to extract more correct MWE. The performance of different measures, including context analysis and language model-based analysis, on Tibetan MWE identification is given in experimental parts.

# 5 Experiments

## 5.1 Experimental Data

We conduct following experiments, on one hand, to validate effectiveness and feasibility of context analysis and the language model-based analysis; on the other hand, to test the ability of the framework on large-scale corpus. We have built 326,062,576-bytes Tibetan news corpus over the internet via an automatic crawler. They are from three web sites, that are, *Tibet Daily*, *People's Daily* and *Qinghai Daily*. We will utilize this Tibetan News Corpus to evaluate extracted Tibetan MWEs in Sect. 5.2.3. Part of this News Corpus is used in Sect. 5.2.1 and 5.2.2, which is randomly selected and has MWE manual checking results. The two-word coupling degree dictionary and Tibetan syllables inside word probability dictionary are trained from annotated training corpus (58 MB). Parameters (i.e. Thresholds) used in the experiment are listed in Table 1.

**Table 1.** Value of parameters in following experiment.

| Parameter names | Function | Value |
|---|---|---|
| $C_{\max}$ | Two-word integration threshold | 0.9 |
| $C_{\min}$ | Two-word separation threshold | 0.3 |
| $P_{\text{inital}}$ | Tibetan syllable word-initial estimation | 0.4 |
| $P_{\text{final}}$ | Tibetan syllable word-final estimation | 0.5 |

## 5.2 Evaluation

We will evaluate the precision ($P$), recall ($R$), f-score ($F$) of Tibetan MWE identification in experimental part.

$$P = N_1/N_2 \tag{6}$$

$$R = N_1/N_3 \tag{7}$$

$$F = 2PR/(P+R) \tag{8}$$

where $N_1$ denotes the number of correctly segmented Tibetan MWEs; $N_2$ denotes total number of segmented Tibetan MWEs; $N_3$ denotes the total number of Tibetan MWEs in testing texts.

### 5.2.1 Evaluation for Different Strategies in Identifying Framework

Context analysis, two-word Coupling Degree and Tibetan syllable inside word probability are three strategies in Tibetan MWE identification framework. In this subsection, we will measure the different combination of these three strategies without segmentation for pre-processing. In Table 2, CNG indicates candidate n-grams, CA indicates context analysis, CD indicates Coupling Degree, IWP indicate inside word probability.

**Table 2.** Results for different combination of three strategies.

| Different combination | *P* | *R* | *F* |
|---|---|---|---|
| CNG | 0.03 | 1.0 | 5.83% |
| CNG + CA | 0.09 | 0.70 | 15.95% |
| CNG + CD | 0.04 | 0.94 | 7.67% |
| CNG + IWP | 0.05 | 0.89 | 9.47% |
| CNG + CA + CD | 0.13 | 0.70 | 21.93% |
| CNG + CA + IWP | 0.20 | 0.68 | 30.91% |
| CNG + CD + IWP | 0.05 | 0.87 | 9.46% |
| CNG + CA + CD + IWP | 0.41 | 0.67 | 50.87% |

Table 2 illustrates the comparison results for various combinations of three strategies. CNG is the baseline, the f-score of CA is the best when these strategies independently used. It means CA is most effective. IWP is better than CD; the recall of CD is the best. In pair-wise testing, combination with CA is better than without CA. It shows that context analysis prior to language model-based analysis is reasonable. CA can eliminate many noises, while language model-based analysis works as a supplement filter.

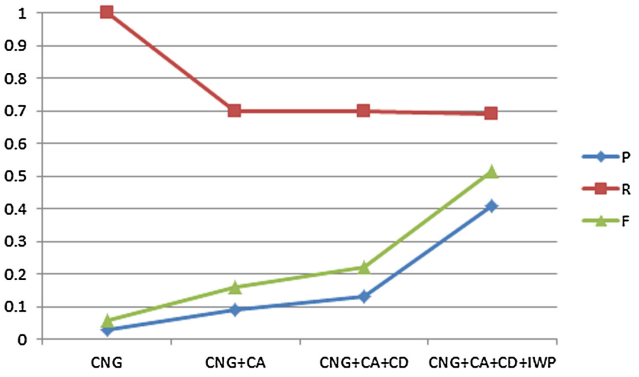


**Fig. 4.** Results of ascending series of the strategies

As we see from Fig. 4, each step of filtering operations greatly improved the precision, while reduced the recall smoothly. It means each filtering strategies works well. CA missed correct candidate MWE more due to the small size of test corpus, it leads to the reduction of the recall. On a large scale corpus, the problem can weaken.

### 5.2.2 Evaluation for the Effect of Context Analysis Granularity

Context analysis granularity is syllable or word. In this subsection, we will evaluate the different granularity of CA. In pre-processing step, sentences in test corpus are segmented or unsegmented will produce n-gram words or n-gram syllables respectively. Results are in Table 3.

**Table 3.** Comparison of different granularity.

| Context analysis granularity | *P* | *R* | *F* |
|---|---|---|---|
| Syllable (unsegmented) | 0.41 | 0.67 | 50.87% |
| Word (segmented) | 0.74 | 0.78 | 75.95% |

Table 3 shows that, both precision and recall significantly improved when word-segmented in pre-processing. The reason is word-segmentation can avoid the "semi-meaningless word".

### 5.2.3 Evaluation on Large Corpus

Preliminary experimental results, on small scale of corpus, illustrate the effectiveness of the combination of context analysis and language model-based analysis. The following test will be made on the whole corpus. The size of whole corpus is too large, manually check all extracted MWEs is impractical. In order to quantify the result, sort the results by the frequency or adjacent categories, and then P@N measure is used.

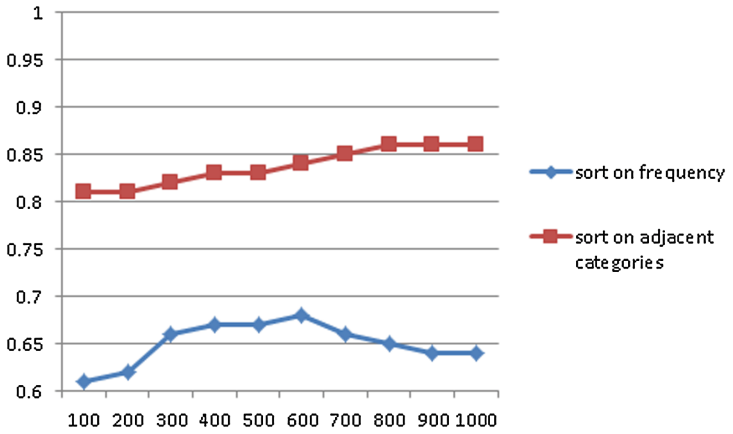


**Fig. 5.** Evaluation on large data corpus

Figure 5 shows the P@N results in two different sort order, the frequency and adjacent categories respectively. Comparative analysis of results found that sorting by adjacent categories is effective than the frequency. When N changes from 100 to 1000, results of adjacent-categories-based sorting keep steady above 80%. In terms of one curve, the P@N first increase and then decline. It is because of some high frequency stop-word list are in the identification results in 300 best.

The experimental results demonstrate that three strategies in framework can improve the precision of MWEs identification; the context analysis is indeed helpful to promote the accuracy and recall rates of Tibetan MWEs on large scale corpus.

## 6 Conclusion

We are in the initial stage of identification of Tibetan MWEs. On the basis of the existing resources of our group, we propose Tibetan MWE identification framework and implement all its components. As a result, it works on different scale of corpus. On small test corpus, the best F-score achieves 75.95%. On larger corpus, the P@N (N is 800) overcomes 85%. With only minor adjustment, it can be ported to other languages. Due to the lack of resources and previous technology, the result is acceptable. Further improvement is needed to become practically applicable for MT system.

**Acknowledgements.** We thank the reviewers for their critical and constructive comments and suggestions that helped us improve the quality of the paper. The research is partially supported by National Science Foundation (No. 61303165) and Informatization Project of the Chinese Academy of Sciences (No. XXH12504-1-10).

## References

1. Smadja, F.: Retrieving collocations from text: Xtract. Comput. Linguist. **19**(1), 143–177 (1993)
2. Dagan, I., Church, K.: Termight: identifying and translating technical terminology. In: Proceedings of 4th Conference on Applied Natural Language Processing, Stuttgart, German, pp. 34–40 (1994)
3. Daille, B.: Combined approach for terminology extraction: lexical statistics and linguistic filtering. Technical paper 5, UCREL, Lancaster University (1995)
4. McEnery, T., Langé, J.-M., Oakes, M., Véronis, J.: The exploitation of multilingual annotated corpora for term extraction. In: Garside, R., Leech, G., McEnery, A. (eds.) Corpus Annotation – Linguistic Information from Computer Text Corpora, pp. 220–230. Longman, London (1997)
5. Michiels, A., Dufour, N.: DEFI, a tool for automatic multi-word unit recognition, meaning assignment and translation selection. In: Proceedings of 1st International Conference on Language Resources & Evaluation, Granada, Spain, pp. 1179–1186 (1998)
6. Diana, M., Sophia, A.: Trucks: a model for automatic multiword term recognition. J. Nat. Lang. Process. **8**(1), 101–126 (2000)
7. Merkel, M., Andersson, M.: Knowledge-lite extraction of multi-word units with language filters and entropy thresholds. In: Proceedings of 2000 Conference User-Oriented Content-Based Text and Image Handling (RIAO 2000), Paris, France, pp. 737–746 (2000)
8. Piao, S.S., McEnery, T.: Multi-word unit alignment in English-Chinese parallel corpora. In: Proceedings of Corpus Linguistics 2001, Lancaster, UK, pp. 466–475 (2001)
9. Sag, I.A., Baldwin, T., Bond, F., Flickinger, D.: Multiword expressions: a pain in the neck for NLP. In: LinGO Working Paper No. 2001-03, Stanford University, CA (2001)
10. Baldwin, T., Bannard, C., Tanaka, T., Widdows, D.: An empirical model of multiword expression decomposability. In: Proceedings of ACL-2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment, Sapporo, Japan, pp. 89–96 (2003)
11. Dias, G.: Multiword unit hybrid extraction. In: Proceedings of Workshop on Multiword Expressions: Analysis, Acquisition and Treatment, at ACL 2003, Sapporo, Japan, pp. 41–48 (2003)

12. Nivre, J., Nilsson, J.: Multiword units in syntactic parsing. In: Proceedings of LREC-2004 Workshop on Methodologies & Evaluation of Multiword Units in Real-world Applications, Lisbon, Portugal, pp. 37–46 (2004)
13. Pereira, R., Crocker, P., Dias, G.: A parallel multikey quicksort algorithm for mining multiword units. In: Proceedings of LREC-2004 Workshop on Methodologies & Evaluation of Multiword Units in Real-world Applications, Lisbon, Portugal, pp. 17–23 (2004)
14. Piao, S.S., Rayson, P., Archer, D., Wilson, A., McEnery, T.: Extracting multiword expressions with a semantic tagger. In: Proceedings of Workshop on Multiword Expressions: Analysis, Acquisition and Treatment, at ACL 2003, Sapporo, Japan, pp. 49–56 (2003)
15. Piao, S.S., Rayson, P., Archer, D., McEnery, T.: Comparing and combining a semantic tagger and a statistical tool for MWE extraction. Comput. Speech Lang. **19**(4), 378–397 (2005)
16. Rayson, P., Archer, D., Piao, S.S., McEnery, T.: The UCREL semantic analysis system. In: Proceedings of Workshop on Beyond Named Entity Recognition Semantic Labelling for NLP Tasks in Association with LREC 2004, Lisbon, Portugal, pp. 7–12 (2004)
17. Piao, S.S., Sun, G., Rayson, P., Yuan, Q.: Automatic extraction of Chinese multiword expressions with a statistical tool. In: Proceedings of 44th Annual Meeting of the Association for Computational Linguistics (2006)
18. Jiang, D.: On syntactic chunks and formal markers of Tibetan. Minor. Lang. China (3), 30–39 (2003a)
19. Jiang, D., Long, C.: The markers of non-finite VP of Tibetan and its automatic recognizing strategies. In: Proceedings of 20th International Conference on Computer Processing of Oriental Languages (ICCPOL 2003) (2003b)
20. Huang, X., Sun, H., Jiang, D., Zhang, J., Tang, L.: The types and formal markers of nominal chunks in contemporary Tibetan. In: proceedings of 8th Joint Conference on Computational Linguistics (JSCL 2005) (2005)
21. Nuo, M., Liu, H., Ma, L., Wu, J., Ding, Z.: Construction of Chinese-Tibetan multi-word equivalence pair dictionary. J. Chin. Inf. Process. **26**(3), 98–103 (2012)
22. Nuo, M., Liu, H., Zhao, W., Ma, L., Wu, J., Ding, Z.: Tibetan base noun phrase identification framework based on Chinese-Tibetan sentence aligned corpus. In: Proceedings of 26th International Conference on Computational Linguistics Conference, pp. 2141–2157 (2012)
23. Lü, X., Zhang, L., Hu, J.: Statistical substring reduction in linear time. In: Su, K.-Y., Tsujii, J., Lee, J.-H., Kwong, O.Y. (eds.) IJCNLP 2004. LNCS (LNAI), vol. 3248, pp. 320–327. Springer, Heidelberg (2005). doi:10.1007/978-3-540-30211-7_34
24. Nagao, M., Mori, S.: A new method of N-gram statistics for large number of n and automatic extraction of words and phrases from large text data of Japanese. In: COLING-1994 (1994)

http://www.springer.com/978-3-319-50495-7

Natural Language Understanding and Intelligent Applications
5th CCF Conference on Natural Language Processing and Chinese Computing, NLPCC 2016, and 24th International Conference on Computer Processing of Oriental Languages, ICCPOL 2016, Kunming, China, December 2–6, 2016, Proceedings
Lin, C.-Y.; Xue, N.; Zhao, D.; Huang, X.; Feng, Y. (Eds.)
2016, XXII, 952 p. 377 illus., Softcover
ISBN: 978-3-319-50495-7