Research on Tibetan Semantic Role Labeling using an Integrated Strategy

Permalink
https://escholarship.org/uc/item/58s1r746

Journal
Himalayan Linguistics, 15(1)

Authors
Long, Congjun
Li, Lin

Publication Date
2016

DOI
10.5070/H915130064

License
CC BY-NC-ND 4.0

Peer reviewed
Research on Tibetan semantic role labeling using an integrated strategy

Congjun Long
Chinese Academy of Social Sciences; Chinese Academy of Sciences

Lin Li
Qinghai Normal University

ABSTRACT
Semantic role labeling is one of the most significant research fields of natural language processing. Researchers have already made many achievements in English and Chinese semantic role labeling. Until now, however, Tibetan semantic role labeling has remained at an early stage due to the absence of a Tibetan corpus with semantic role annotation and relatively outdated research approaches. Tibetan is rich with syntactic markers that naturally divide a sentence into semantic chunks and indicate the semantic relationships between these chunks. Thus, in this paper, we propose a semantic role classification and an integrated strategy for Tibetan semantic role labeling. Transformation-Based Error-driven Learning and Conditional Random Fields have been employed in our study. Additionally, a number of linguistic rules have been introduced into our approach as well. Our integrated strategy achieves 83.91% in precision, 82.78% in recall, and an F-score of 85.71.

KEYWORDS
Tibetan, Semantic Role Labeling, TBL, CRFs
Research on Tibetan semantic role labeling using an integrated strategy

Congjun Long
Chinese Academy of Social Sciences; Chinese Academy of Sciences

Lin Li
Qinghai Normal University

1 Introduction

Semantic role labeling (hereafter referred to as SRL) plays a significant role in information processing, which is one of the main fields of research in natural language processing. SRL enables a computer to approximate a human understanding of language. The process of SRL can be summarized as follow: 1) to design a semantic role classification; 2) to identify all semantic chunks in a sentence.

Gildea and Jurafsky (2002) conducted the earliest research in SRL; they develop a SRL system and tested their system on two testing materials, achieving precisions of 82% and 65% on the two corpora respectively. A further contribution of theirs, in CoNLL2004, emphasizes the classifying of syntactic chunking, using the same training corpus, they achieved an accuracy of 72.43%, a recall rate of 66.77% and an F value of 69.49% (Kadri Hacioglu, etal. 2004). CoNLL2007\(^1\) organizes an independent session for SRL, and CoNLL2008\(^2\) set up SRL as a shared task to observe performances of SRL and syntactic parsing.

Chinese researchers have long dedicated research to SRL. In CoNLL2005, Liu (2005) submits his work on English SRL using a maximum entropy model. His work employs a corpus with syntactic constituent tags and corrects the results with a rule-based approach. His approach reaches 79.65% in precision, 71.34% in recall, and an F-score of 75.27%. After CoNLL2005, researchers have made many achievements in Chinese SRL (Yu et al. 2007; Wang Bukang et al., 2010; Liu et al. 2007). Particularly worth mentioning is Ding’s work (Ding et al. 2009); Ding has successfully introduce chunking results into SRL.

The lack of available Tibetan syntactic tree banks means that it is not possible to apply achievements in syntax parsing and dependency parsing to Tibetan SRL. Fortunately, Tibetan has a large number of syntax markers that divide a sentence into several chunks naturally. Researchers (Jiang 2003, 2005, Li et al. 2013; Long et al. 2004) have studied Tibetan chunking from various perspectives, but they have not yet explored the correspondence between chunks and semantic roles. In this work, we propose a multi-part strategy using rule-based and statistic-based approaches to

\(^1\) http://www.cs.jhu.edu/EMNLP-CoNLL-2007/
\(^2\) http://www.clips.ua.ac.be/conll2008/
Tibetan SRL. First, we adopt Conditional Random Fields (hereafter referred to as CRFs) to identify semantic roles in our corpus; secondly, a linguistic rule bank is applied to correct the results of first step. Our rule bank is built up by TBL (transformation based learning) that is an automatic rule extraction algorithm that starts with a small list of manual rules.

2 Tibetan semantic role classification

2.1 Tibetan markers

Tibetan SRL focuses on completing two tasks: 1) to detect boundaries of semantic chunks; 2) to recognize the semantic type of a chunk. Markers in Tibetan convey information of both chunk boundaries and semantic type of a chunk. Take example sentence 1 as an example to illustrate functions of Tibetan markers.

Example sentence 1: \[\text{ཁོ/} \text{rh}/ \text{ས་/} \text{ka}/ \text{ང/} \text{rh}/ \text{འི/} \text{kg}/ \text{ངས་འཁོར་/} \text{ng}/ \text{གསར་/} \text{a}/ \text{ར/} \text{kd}/ \text{བȪོད་/} \text{vt}/ \text{ཀྱིན་/} \text{t}/ \text{xp}\]

Lat: khos ngavi rlangs vkhor gsar bar bstod pa byed kyin yod.  
Eng: He praises my new car a lot.

In sentence 1, \(\text{s}/\text{ka}\) (Lat: sa; Eng: agentive) is an agentive case marker; \(\text{r}/\text{kd}\) (Lat: -r, Eng: dative) is an dative case marker. These two markers divide the sentence into three chunks and also imply the roles of the semantic chunks which precede them. In Tibetan, a marker may have more than one grammar functions, e.g. \(\text{r}/\text{kd}\) can be a dative case marker or a locative case marker. Thus, the multifunctionality of Tibetan markers yields some difficulties for Tibetan SRL. In this paper, we develop our research based on a corpus with part-of-speech annotation; hence, the function of a marker is already identified. Tibetan markers can be classified into two major categories: 1) case markers such as agentive case, instrumental case, objective case, and so on; 2) particle words such as likening particle, enumerating particle and so on.

2.2 Tibetan semantic role classification

Tibetan semantic role classification is foundational to Tibetan SRL research; a classification scheme serves as guidance for semantic role annotation in a corpus. Envisioning a semantic role classification for Tibetan is a project for linguistic engineering. If the classification system is too complex, it benefits linguistic research, but creates many problems for corpus annotation. If the classification system is too simple, it cannot satisfy the requirements of SRL research. Therefore it is crucial to set up a proper semantic role classification. Yuan has deeply studied semantic role classification from micro-, meso-, and macroscopic perspectives. Microscopic classification contains semantic roles based on specific verbs and specific domains. Mesoscopic classification consists of various semantic cases, whose foundation is verb categories instead of specific verbs. Macroscopic classification consists only of distinguishing a proto-agent and proto-patient (Yuan 2007). Tibetan semantic role classification is similar to mesoscopic classification, because it concentrates on syntactic

---

1 The tagset adopted in this article come from “Pos tagset specification of modern tibetan for information processing (draft)”, Zhao Xiaobing, Sun Yuan, Long Congjun et.al, the commercial press, 2015.6. (信息处理用现代藏语词性标记规范（草案）,赵小兵、孙媛、龙从军等，商务印书馆，2015年6月）

4 The “Lat” means “Latin transliteration” and the “Eng” means “English translation”,

114
markers and auxiliaries. In the meantime, we propose three principles for setting up a semantic role classification for Tibetan.

(1) To learn from successful experiences in English and Chinese semantic role classifications (Zhou et al. 1999, 2001; Yang 2011; Lu 2001; Lin 1999) and adjust them according to the characteristics of Tibetan.

(2) Syntactic markers are helpful for computers to understand human language. Compared to Chinese, Tibetan exhibits a large quantity of syntactic markers that provide an important basis for SRL. Therefore, Tibetan markers need to be taken into account. Table 1 lists syntax markers and possible corresponding semantic roles.

<table>
<thead>
<tr>
<th>Syntax markers</th>
<th>Semantic Role</th>
<th>Syntax markers</th>
<th>Semantic Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>-s/gis/gyis/kyis/yis</td>
<td>Agent (AG)</td>
<td>-s/gis/gyis/kyis/yis</td>
<td>Comitative (CE)</td>
</tr>
<tr>
<td>-s/gis/gyis/kyis/yis</td>
<td>Genitive (GE)</td>
<td>-s/gis/gyis/kyis/yis</td>
<td>Outcome (OE)</td>
</tr>
<tr>
<td>-s/gis/gyis/kyis/yis</td>
<td>Causative (CA)</td>
<td>-s/gis/gyis/kyis/yis</td>
<td>Manner (MR)</td>
</tr>
<tr>
<td>-s/gis/gyis/kyis/yis</td>
<td>Patient (PT)</td>
<td>-s/gis/gyis/kyis/yis</td>
<td>Instrument (IT)</td>
</tr>
<tr>
<td>-s/gis/gyis/kyis/yis</td>
<td>Source (SE)</td>
<td>-s/gis/gyis/kyis/yis</td>
<td>Time (TM)</td>
</tr>
<tr>
<td>-s/gis/gyis/kyis/yis</td>
<td>Target (TT)</td>
<td>-s/gis/gyis/kyis/yis</td>
<td>Location (LC)</td>
</tr>
<tr>
<td>-s/gis/gyis/kyis/yis</td>
<td>Basis (BS)</td>
<td>-s/gis/gyis/kyis/yis</td>
<td>Direction (DN)</td>
</tr>
</tbody>
</table>

Table 1. Syntax markers and corresponding semantic role

(3) There is no one-to-one match between syntax markers and semantic roles, and there are a number of Tibetan semantic chunks without any syntactic marker. For example, ཏེས་འབུག་བསྒྲུབ་སྒྲུབ་ཙོལ་ཏུ (Lat: nga tsho phyi drovi dus tshod gsum steng ngo thug pa yin. Eng: we met at three o’clock in the afternoon”), there is no syntactic marker among three chunks. But there exists a correspondence relationship of semantic roles, if ཏེས་འབུག་ (Lat: ngo thug, Eng: meet) has two arguments, one of argument is human and the other is always time or place. According to the type, we construct semantic frames for some verbs. The correspondence relationship between special sentences and semantic roles also merits consideration, for instance, causee and beneficiary are semantic roles related to causative sentences and factitive verb. Based on the linguistic phenomena discussed above, the classification for Tibetan SRL proposed in this paper is shown in table 2.
3 SRL rule bank construction

Compared to statistics-based approaches, rule-based approaches do not have obvious advantages. But statistic approaches cannot be fully used because of the lack of resources. In this regard, a rule-based SRL approach is still useful at this stage. Thus, TBL (Transformation Based Learning) is applied in this paper to build up a rule bank. Firstly, a small-scale basic rule collection is manually set up. Secondly, we build up an expanded rule collection through TBL automatically extracting from a corpus.

3.1 Basic rule collection

Basic rule collection consists of semantic chunk boundary rules and correspondence relationship rules of case markers and auxiliaries. The basic rule collection is mainly gathered by language experts; it is composed of four types of rules. They are left boundary rules, right boundary rules, left-right boundary rules, and left-right boundary exception rules. The basic rule collection has 271 rule entries in total, which includes 114 right boundary features, 119 double boundaries features, 15 left boundary features and 35 double boundaries exception features. Amount of correspondence rule between semantic roles and case markers is 63. A part of rules are listed in appendix 1.

3.2 Expanded rule collection

Based on our basic rule collection, TBL is used to learn expanded rules from the corpus, which form an expanded rule collection. TBL employs a learning algorithm to acquire transition rules from the corpus; therefore a high efficient learning algorithm is an essential part of TBL. A learning algorithm needs three types of language materials: (1) Tibetan corpus with semantic role annotation, (2) Tibetan corpus labeled by basic rule collection, (3) a basic rule template collection. Through comparing (1) and (3), an expanded rule collection is formed.

---

5 The “Referent” and “Category” are used to describe the subject and complement of copula, for instance, [པེ་ཅིང་/ns] [RE] [ནི་/up] [ནི་/up] [ོང་/ns] [CT] [ཡིན/vl] [xp](Lat: be cing ni krung govi rgyas sa yin. Eng: Beijing is the capital of China.)

6 The “Object” is used to describe the property and of statue of “Experiencer”, for instance, [ནམ་/nt] [འȣ་/ve] [དེ་/ks] [འི་/kg] [མཁའ་དིངས་/ng](Lat: lha savi mkhav dbyings nam g-yav dag se vdag. Eng: The weather is very sunny in Lhasa.)

7 http://www.cs.jhu.edu/~rflorian/fntbl/download.noform.html
4 Statistic-based model and feature selection

4.1 CRFs model

CRF, a discriminant probability model, is widely used in sequential annotation tasks. As a statistic-based model, CRFs originates from maximum entropy (ME) model and performs very well in annotation tasks. Furthermore, CRFs does not have the data sparseness problem that exists in ME and is not based on a conditional independence assumption. Normally, CRFs adopt a first order chain structure shown in Figure 1.

![Figure 1. First order chain structure](image)

4.2 Semantic role annotation system

A semantic role annotation system is crucial for SRL, because the amount of annotation tag directly influences the performance of a SRL recognition model especially when the training corpus is limited. Our training corpus is annotated manually, and our corpus offers information of part-of-speech, semantic chunk boundary, semantic role type, etc. Our system adopts BIO annotation approach, “S” means a word is outside of a chunk, “B” means a word is at the beginning of a chunk, and “I” means a word located inside of a chunk. Therefore, we design an integrated tag system to express semantic role information. Our integrated tags contain separate tiers for word form, part-of-speech, chunk boundary tags, and semantic role information as shown in Table 3.

<table>
<thead>
<tr>
<th>Word Form</th>
<th>བས</th>
<th>ཟོ</th>
<th>གྲ</th>
<th>འཛོ</th>
<th>ཆེ</th>
<th>འཛོ</th>
<th>ཆེ</th>
<th>འཛོ</th>
<th>ཆེ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part-of-speech</td>
<td>བྷ</td>
<td>བྷ</td>
<td>བྷ</td>
<td>བྷ</td>
<td>བྷ</td>
<td>བྷ</td>
<td>བྷ</td>
<td>བྷ</td>
<td>བྷ</td>
</tr>
<tr>
<td>Boundary Tag</td>
<td>S</td>
<td>M</td>
<td>B</td>
<td>I</td>
<td>I</td>
<td>E</td>
<td>M</td>
<td>B</td>
<td>E</td>
</tr>
<tr>
<td>Semantic Role Tag</td>
<td>AG</td>
<td>M</td>
<td>TT</td>
<td>TT</td>
<td>TT</td>
<td>TT</td>
<td>M</td>
<td>P</td>
<td>P</td>
</tr>
</tbody>
</table>

Table 3. Integrated Semantic Role Tag System

---

8 The corpus, with five thousand sentences and about forty thousands of words, was built by Institute of Ethnology & Anthropology Chinese Academy of Social Sciences. The material of corpus come from primary and secondary textbooks and other grammatical textbooks. All sentences are written texts. The POS tagset come from “Pos tagset specification of modern Tibetan for information processing (draft)”, Zhao Xiaobing, Sun Yuan, Long Congjun et.al , the commercial press, 2015.6. The boundary and semantic role makers are tagged by Congjun Long. Some results about the Tibetan information processing can be found in web page: [http://103.247.176.245:8081/](http://103.247.176.245:8081/)
4.3 Feature selection

CRFs provide feature function definitions by feature templates, which simplifies feature selection and feature function definition.

In this paper, we apply basic features that are word-form and part-of-speech; and expanded features that are syllable amount, predicate verb category, and the distance between a predicate verb and a semantic chunk. Details are described as follows.

Word form refers to formal attribute of a word, for instance, ཆོང (khong, he) and དཔལ་རང་ (khyed rang, you) have two different word form attributions.

Part-of-speech presents category information of a word. For instance, ཆོང (Lat: khong; Eng: he) and དཔལ་རང་ (Lat: khyed; Eng: you) are pronoun, གནས་ ولو (Lat: gnas tshul; Eng: situation) is a noun.

Syllable amount means the number of syllables inside of a semantic chunk; we take this feature as a reference for boundary recognition.

This sentence contains four semantic chunks: chunk1 [ sở/ve] (Lat: yod; Eng: have) is a predicate chunk, chunk 2 [/xp] is a punctuation mark, the syllable amount of chunk 3 and chunk 4 are seven and four respectively.

Predicate verb category refers to the semantic type of a predicate verb. In this study, we classify verbs into one valence verbs, two valence verbs and three valence verbs according to the number of their arguments, for instance, if the verb འགྲོ་ (Lat: vgro; Eng: go) have two semantic roles, we constructed the rules such as (EX, LC, འགྲོ་). Predicate verb category influences the number of semantic roles.

Distance between a predicate verb and a semantic chunk refers to the syllable amount between a semantic chunk and a predicate verb. Normally, a patient is closer to a predicate than an agent. This pattern is helpful in semantic role recognition.

5 Experiments and results

5.1 Tools and corpus

CRF++ package developed by Dr. Taku Kudo is employed in this study. Firstly, we build up a baseline model that only adopts the basic features mentioned above. Based on the baseline model, we conduct three groups of experiments that adopt different expanded features. By experiment results, we can tell whether the expanded features improve the Tibetan SRL model. 5000 sentences are used for training our Tibetan SRL model, and 500 sentences are used for testing our model.

5.2 Results and analysis

Precision (P), recall (R), and F-score are applied to evaluate our Tibetan SRL model in this paper. And their formulas are listed as follow.

P is the proportion of arguments predicted by a model which are correct.

R is the proportion of correct arguments which are predicted by a system. The formula of F-score is $F\text{-score}= \frac{2PR}{P + R}$.
The precision of the baseline model is 68.88%, the recall rate is 63.10%, and the F-score is 64.85%. Experimental results suggest that the performance of our model is obviously improved when expanded features are introduced. The feature of syllable number contributes the most, specifically precision, recall, and F-score reach 78.60%, 71.34%, and 74.80% respectively. These results are shown in Figure 2.

Figure 2. Experimental results based on different features.

(F1: Syllable amount means the syllable quantity of a semantic chunk; F2: semantic type of a predicate verb; F3: the syllable amount between a semantic chunk and a predicate verb)

The errors of the statistic-based semantic role recognition model can be categorized as follow:

(1) Boundary detection errors, for instance, sample sentence 2.

Sample sentence 2:

[k̃l̃ng][ng/BS][k̃lt̃ar][ua][c/rd][RE][k̃t̃aw][AT][k̃ṽl][ṽy][xp] (Lat: kung k̃ri lt̃ar na de ga t̃shod red dam. Eng: How many meters are these?)

Our result:

[k̃l̃ng][ng/BS][k̃lt̃ar][ua][c/rd][RE][k̃t̃aw][AT][k̃ṽl][ṽy][xp]

The word k̃lt̃ar (Lat: lt̃ar; Eng: according to) is a boundary marker, but our model cannot detect the right boundary, which leads to fail in BS chunk recognition.

Sample sentence 3:

[k̃lt̃c]/rh̃gi/kt̃ng[PT][c/rh][TT][k̃ṽd][k̃ṽt][ṽy][xp] (Lat: khyed rang gi ming nga la shod dang. Eng: Tell me your name.)

Our result:

Semantic role chunk [k̃lt̃c]/rh̃gi/kt̃ng (Lat: khyed rang gi ming; Eng: your name) and [c/rh] (Lat: nga; Eng: me) cannot be detected correctly, hence {PT} cannot be recognized correctly either.

(2) The boundary is recognized correctly, but the semantic role is incorrectly tagged. Mistakes in semantic role recognition include two types. One is that one or more semantic chunk is missing; the other is incorrect semantic annotation.

Sample sentence 4:
5.3 Experiment improvement

Based on practical experience, a rule-based approach is useful for a NLP system, especially when the scale of corpus is limited. Therefore, in this work, both basic rule collection and expanded rule collection are applied into our SRL model10. To discover the effects of our features, we build up three SRL models. Model1 is the baseline model, Model2 is statistic-based model with expended features, and Model3 is integrated model employing both rule-based and statistic-based approaches.

According to the results, we find that Model3 performs the best in Tibetan SRL. Its precision reaches 82.78%, recall reaches 85.71%, and the F-score reaches 83.91%. These results suggest that the rule-based approach does contribute a great deal to our SRL task.

---

10 There are two stages to use the rules, one is to adjust the boundary errors by using the boundary rules, the other is to adjust the semantic role labeling errors by using the semantic markers. However, not all of adjustment are right. But the proportion of correct adjustments is higher than the erroneous ones.
However, it is clear that SRL of Tibetan still has many problems; we only experimented with the simple sentences. Chunks embedded boundary and the SRL of clause embedded still require research. Moreover, the sentence boundary of Tibetan language is indistinct in continuous text.

6 Conclusion

SRL is a main research field of shallow syntax parsing, which plays an important role in natural language understanding because there are still many irresolvable difficulties in full syntax parsing. The research achievements of SRL are widely applied in many fields such as Machine Translation, Question Answering Systems, and Information Retrieval Systems. In this paper, we focus on Tibetan SRL. By adopting an integrated SRL strategy, precision of our model reaches 82.78%. Our model has not yet successfully recognized nested semantic chunks or long distance semantic chunks. In future research, we plan to improve our model by expanding the corpus and refining the rule-bank.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (61132009, 31271337) and the National Social Science Foundation of China (12&ZD174).

Note

The labels of POS in this article are listed here.

```
ng noun-general   rw pronoun-interrogative   t Aspect maker
nh noun-human     rd pronoun-demonstrative h Nominalization maker
ns noun-space     ri pronoun-indefinite   p plural
ni noun-institution ua particle-analogy   y Mood particle
nt noun-time      up particle-pause       e exclamation
nd noun-direction ue particle-enumeration o onomatopoeia
nz noun-others   uf particle-manner      i idom
m numeral        ur particle-result      in idom-noun
q quantity       um particle-purpose     iv idom-verb
d adverb         kg case maker-genitive   ia idom-adjective
c conjunction    ka case maker-agentive   ic idom-conjunction
vl verb-linking   ki case maker-instrumental id idom-adverb
ve verb-exist    kl case maker-locative   j abbreviations
vd verb-direction kd case maker-dative   s syllable
va verb-auxiliary kc case maker-source   w Other symbols
vt verb-transitive kb case maker-compare xp punctuation
vi verb-intransitive kp case maker-possess
a adjective      kx case maker-allative
rh pronoun-human ks case maker-concomitant
```
REFERENCES


Long, Congjun; and Jiang, Di. 2004. “Recognition method with auxiliary verb predicate chunking of Modern Tibetan”. In: Di Er et al. (Eds.): The 2nd Youth Computational Linguistics Conference Papers. (龙从军、江荻. 现代藏语助动词谓语组块的识别方法[A], 第2届青年计算语言学会议论文[C], 2004.)


Yu, Jiangde; Fan, Xiaozhong; Wenbo, Pang; and Zhengtao, Yu. 2007. “Semantic role labeling based on conditional random fields”. Journal of Southeast University 23.3: 361- 364.


Congjun Long
longcj@cass.org.cn
Appendix 1.

<table>
<thead>
<tr>
<th>/kh</th>
<th>/dn̥/</th>
<th>/l</th>
<th>/ub</th>
<th>/kh</th>
<th>/ki</th>
<th>/u/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/u/</td>
<td>/d̥/</td>
<td>/l̥/</td>
<td>/d̥/</td>
<td>/a/</td>
<td>/kh</td>
<td>/o/</td>
</tr>
<tr>
<td>/k̥/</td>
<td>/u/</td>
<td>/n̥/</td>
<td>/d̥/</td>
<td>/k̥/</td>
<td>/k̥/</td>
<td>/n̥/</td>
</tr>
<tr>
<td>/u/</td>
<td>/n̥/</td>
<td>/k̥/</td>
<td>/k̥/</td>
<td>/k̥/</td>
<td>/k̥/</td>
<td>/k̥/</td>
</tr>
<tr>
<td>/u/</td>
<td>/n̥/</td>
<td>/k̥/</td>
<td>/k̥/</td>
<td>/k̥/</td>
<td>/k̥/</td>
<td>/k̥/</td>
</tr>
</tbody>
</table>

Table 1. Rules of double boundary (part)

<table>
<thead>
<tr>
<th>/t/</th>
<th>/rh</th>
<th>/n̥/</th>
<th>/a/</th>
<th>/u/</th>
<th>/n̥/</th>
<th>/n̥/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/u/</td>
<td>/a/</td>
<td>/k̥/</td>
<td>/d̥/</td>
<td>/a/</td>
<td>/k̥/</td>
<td>/d̥/</td>
</tr>
<tr>
<td>/p̥/</td>
<td>/r̥/</td>
<td>/p̥/</td>
<td>/n̥/</td>
<td>/a/</td>
<td>/k̥/</td>
<td>/d̥/</td>
</tr>
<tr>
<td>/t̥/</td>
<td>/n̥/</td>
<td>/k̥/</td>
<td>/d̥/</td>
<td>/a/</td>
<td>/k̥/</td>
<td>/d̥/</td>
</tr>
<tr>
<td>/r̥/</td>
<td>/r̥/</td>
<td>/m̥/</td>
<td>/m̥/</td>
<td>/a/</td>
<td>/k̥/</td>
<td>/d̥/</td>
</tr>
<tr>
<td>/n̥/</td>
<td>/n̥/</td>
<td>/k̥/</td>
<td>/d̥/</td>
<td>/a/</td>
<td>/k̥/</td>
<td>/d̥/</td>
</tr>
<tr>
<td>/r̥/</td>
<td>/m̥/</td>
<td>/n̥/</td>
<td>/a/</td>
<td>/k̥/</td>
<td>/d̥/</td>
<td>/a/</td>
</tr>
<tr>
<td>/t̥/</td>
<td>/n̥/</td>
<td>/k̥/</td>
<td>/d̥/</td>
<td>/a/</td>
<td>/k̥/</td>
<td>/d̥/</td>
</tr>
</tbody>
</table>

Table 2. Rules of right boundary (part)

<table>
<thead>
<tr>
<th>/v̥/</th>
<th>/dn̥/</th>
<th>/l̥/</th>
<th>/a/</th>
<th>/m̥/</th>
<th>/p̥/</th>
<th>/a/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/v̥/</td>
<td>/dn̥/</td>
<td>/l̥/</td>
<td>/a/</td>
<td>/m̥/</td>
<td>/p̥/</td>
<td>/a/</td>
</tr>
<tr>
<td>/v̥/</td>
<td>/dn̥/</td>
<td>/l̥/</td>
<td>/a/</td>
<td>/m̥/</td>
<td>/p̥/</td>
<td>/a/</td>
</tr>
<tr>
<td>/v̥/</td>
<td>/dn̥/</td>
<td>/l̥/</td>
<td>/a/</td>
<td>/m̥/</td>
<td>/p̥/</td>
<td>/a/</td>
</tr>
</tbody>
</table>

Table 3. Rules of erasing right boundary (part)
Long and Li: Research on Tibetan semantic role labeling using an integrated strategy

<table>
<thead>
<tr>
<th>Role</th>
<th>Correspondence Rule</th>
<th>Role</th>
<th>Correspondence Rule</th>
<th>Role</th>
<th>Correspondence Rule</th>
<th>Role</th>
<th>Correspondence Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{kp} )</td>
<td>{GE}</td>
<td>( \text{l} )</td>
<td>{MR}</td>
<td>( \text{kl} )</td>
<td>{LC}</td>
<td>( \text{up} )</td>
<td>{RE}</td>
</tr>
<tr>
<td>( \text{kx} )</td>
<td>{DN}</td>
<td>( \text{kl} )</td>
<td>{LC}</td>
<td>( \text{ka} )</td>
<td>{AG}</td>
<td>( \text{kx} )</td>
<td>{DN}</td>
</tr>
<tr>
<td>( \text{kd} )</td>
<td>{TT}</td>
<td>( \text{ka} )</td>
<td>{AG}</td>
<td>( \text{ki} )</td>
<td>{IT}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{ub} )</td>
<td>{OE}</td>
<td>( \text{lc} )</td>
<td>{IT}</td>
<td>( \text{ka} )</td>
<td>{AG}</td>
<td>( \text{ki} )</td>
<td>{IT}</td>
</tr>
<tr>
<td>( \text{ua} )</td>
<td>{BS}</td>
<td>( \text{kd} )</td>
<td>{TT}</td>
<td>( \text{ka} )</td>
<td>{AG}</td>
<td>( \text{ki} )</td>
<td>{IT}</td>
</tr>
</tbody>
</table>

Table 4. Correspondence rule between semantic roles and case makers (part)