Semantic Role Tagging for Chinese at the Lexical Level

Oi Yee Kwong and Benjamin K. Tsou

Language Information Sciences Research Centre, City University of Hong Kong,
Tat Chee Avenue, Kowloon, Hong Kong
{rlolivia, rlbtsou}@cityu.edu.hk

Abstract. This paper reports on a study of semantic role tagging in Chinese, in the absence of a parser. We investigated the effect of using only lexical information in statistical training; and proposed to identify the relevant headwords in a sentence as a first step to partially locate the corresponding constituents to be labelled. Experiments were done on a textbook corpus and a news corpus, representing simple data and complex data respectively. Results suggested that in Chinese, simple lexical features are useful enough when constituent boundaries are known, while parse information might be more important for complicated sentences than simple ones. Several ways to improve the headword identification results were suggested, and we also plan to explore some class-based techniques for the task, with reference to existing semantic lexicons.

1 Introduction

As the development of language resources progresses from POS-tagged corpora to syntactically annotated treebanks, the inclusion of semantic information such as predicate-argument relations is becoming indispensable. The expansion of the Penn Treebank into a Proposition Bank [11] is a typical move in this direction. Lexical resources also need to be enhanced with semantic information (e.g. [5]). In fact the ability to identify semantic role relations correctly is essential to many applications such as information extraction and machine translation; and making available resources with this kind of information would in turn facilitate the development of such applications.

Large-scale production of annotated resources is often labour-intensive, and thus needs automatic labelling to streamline the work. The task can essentially be perceived as a two-phase process, namely to *recognise* the constituents bearing some semantic relationship to the target verb in a sentence, and then to *label* them with the corresponding semantic roles.

In their seminal proposal, Gildea and Jurafsky approached the task using various features such as headword, phrase type, and parse tree path [6]. Such features have remained the basic and essential features in subsequent research, irrespective of the variation in the actual learning components. In addition, parsed sentences are often required, for extracting the path features during training and providing the argument boundaries during testing. The parse information is deemed important for the performance of role labelling [7, 8].

More precisely, in semantic role labelling, parse information is rather more critical for the identification of boundaries for candidate constituents than for the extraction

R. Dale et al. (Eds.): IJCNLP 2005, LNAI 3651, pp. 804 – 814, 2005.

of training data. Its limited function in training, for instance, is reflected in the low coverage reported (e.g. [21]). However, given the imperfection of existing automatic parsers, which are far from producing gold standard parses, many thus resort to shallow syntactic information from simple chunking, though results often turn out to be less satisfactory than with full parses.

This limitation is even more pertinent for the application of semantic role labelling to languages which do not have sophisticated parsing resources. In the case of Chinese, for example, there is considerable variability in its syntax-semantics interface; and when one has more nested and complex sentences such as those from news articles, it becomes more difficult to capture the sentence structures by typical examples.

It is therefore worthwhile to investigate alternatives to the role labelling task for Chinese under the parsing bottleneck, both in terms of the features used and the shortcut or compromise to at least partially pin down the relevant constituents. A series of related questions deserve consideration here:

1. how much could we achieve with only parse-independent features in the role labelling process;
2. with constituent boundaries unknown in the absence of parse information, could we at least identify the headwords in the relevant constituents to be tagged; and
3. whether the unknown boundary problem varies with the nature of the dataset, e.g., will the degradation in performance from known boundaries to unknown boundaries be more serious for complicated sentences than for simple sentences.

So in the current study we experiment on the use of parse-independent features for semantic role labelling in Chinese, for locating the headwords of the constituents corresponding to arguments to be labelled. We will also compare the results on two training and testing datasets.

In Section 2, related work will be reviewed. In Section 3, the data used in the current study will be introduced. Our proposed method will be explained in Section 4, and the experiment reported in Section 5. Results and future work will be discussed in Section 6, followed by conclusions in Section 7.

2 Related Work

The definition of semantic roles falls on a continuum from abstract ones to very specific ones. Gildea and Jurafsky [6], for instance, used a set of roles defined according to the FrameNet model [2], thus corresponding to the frame elements in individual frames under a particular domain to which a given verb belongs. Lexical entries (in fact not limited to verbs, in the case of FrameNet) falling under the same frame will share the same set of roles. Gildea and Palmer [7] defined roles with respect to individual predicates in the PropBank, without explicit naming. To date PropBank and FrameNet are the two main resources in English for training semantic role labelling systems.

The theoretical treatment of semantic roles is also varied in Chinese. In practice, for example, the semantic roles in the Sinica Treebank mark not only verbal arguments but also modifier-head relations within individual constituents, following a head-driven principle [4]. In our present study, we use a set of more abstract semantic roles, which are generalisable to most Chinese verbs and are not dependent on particular predicates. They will be further introduced in Section 3.

The major concerns in automatic semantic role labelling include the handling of alternations (as in "the window broke" and "John broke the window", where in both cases "the window" should be tagged as "patient" despite its appearance in different positions in the sentences), and generalisation to unseen constituents and predicates. For the latter, clustering and semantic lexicons or hierarchies have been used (e.g. [6]), or similar argument structures are assumed for near-synonyms and verbs under the same frame (e.g. [11]).

Approaches in automatic semantic role labelling are mostly statistical, typically making use of a number of features extracted from parsed training sentences. In Gildea and Jurafsky [6], the features studied include phrase type (*pt*), governing category (*gov*), parse tree path (*path*), position of constituent with respect to the target predicate (*position*), voice (*voice*), and headword (*h*). The labelling of a constituent then depends on its likelihood to fill each possible role *r* given the features and the target predicate *t*, as in the following, for example:

$$P(r \mid h, pt, gov, position, voice, t)$$

Subsequent studies exploited a variety of implementation of the learning component, including Maximum Entropy (e.g. [1, 12]), Support Vector Machines (e.g. [9, 16]), etc. Transformation-based approaches were also used (e.g. [10, 19]). Swier and Stevenson [17] innovated with an unsupervised approach to the problem, using a bootstrapping algorithm, and achieved 87% accuracy.

While the estimation of the probabilities could be relatively straightforward, the key often lies in locating the candidate constituents to be labelled. A parser of some kind is needed. Gildea and Hockenmaier [8] compared the effects of Combinatory Categorial Grammar (CCG) derivations and traditional Treebank parsing, and found that the former performed better on core arguments, probably due to its ability to capture long range dependencies, but comparable for all arguments. Gildea and Palmer [7] compared the effects of full parsing and shallow chunking; and found that when constituent boundaries are known, both automatic parses and gold standard parses resulted in about 80% accuracy for subsequent automatic role tagging, but when boundaries are unknown, results with automatic parses dropped to 57% precision and 50% recall. With chunking only, performance further degraded to below 30%. Problems mostly arise from arguments which correspond to more than one chunk, and the misplacement of core arguments.

A couple of evaluation exercises for semantic role labelling were organized recently, such as the shared task in CoNLL-2004 using PropBank data [3], and the one in SENSEVAL-3 using the FrameNet dataset [15]. Most systems in SENSEVAL-3 used a parser to obtain full syntactic parses for the sentences, whereas systems participating in the CoNLL task were restricted to using only shallow syntactic information. Results reported in the former tend to be higher. Although the dataset may be a factor affecting the labelling performance, it nevertheless reinforces the usefulness of full syntactic information.

According to Carreras and Màrquez [3], for English, the state-of-the-art results reach an F_1 measure of slightly over 83 using gold standard parse trees and about 77 with real parsing results. Those based on shallow syntactic information is about 60.

The usefulness of parse information for semantic role labelling would be especially interesting in the case of Chinese, given the flexibility in its syntax-semantics interface (e.g. the object after 吃 'eat' could refer to the *Patient* as in 吃蘋果 'eat apple', *Location* as in 吃食堂 'eat canteen', *Duration* as in 吃三年 'eat three years', etc.). In the absence of sophisticated parsing resources, however, we attempt to investigate how well one could simply use a set of parse-independent features and backward guess the likelihood of headwords to partially locate the candidate constituents to be labelled.

3 The Data

3.1 Materials

As mentioned in the introduction, we attempted to investigate the difference between labelling simple sentences and complex ones. For this purpose, sentences from primary school textbooks were taken as examples for simple data, while sentences from a large corpus of newspaper texts were taken as complex examples.

Two sets of primary school Chinese textbooks popularly used in Hong Kong were taken for reference. The two publishers were Keys Press [22] and Modern Education Research Society Ltd [23]. Texts for Primary One to Six were digitised, segmented into words, and annotated with parts-of-speech (POS). The two sets of textbooks amount to a text collection of about 165K character tokens and upon segmentation about 109K word tokens (about 15K word types). There were about 2,500 transitive verb types, with frequency ranging from 1 to 926.

The complex examples were taken from a subset of the LIVAC synchronous corpus[1] [13, 18]. The subcorpus consists of newspaper texts from Hong Kong, including local news, international news, financial news, sports news, and entertainment news, collected in 1997-98. The texts were segmented into words and POS-tagged, amounting to about 1.8M character tokens and upon segmentation about 1M word tokens (about 47K word types). There were about 7,400 transitive verb types, with frequency ranging from 1 to just over 6,300.

3.2 Training and Testing Data

For the current study, a set of 41 transitive verbs common to the two corpora (hereafter referred to as textbook corpus and news corpus), with frequency over 10 and over 50 respectively, was sampled.

Sentences in the corpora containing the sampled verbs were extracted. Constituents corresponding to semantic roles with respect to the target verbs were annotated by a trained annotator, whose annotation was verified by another. In this study, we worked with a set of 11 predicate-independent abstract semantic roles. According to the *Dictionary of Verbs in Contemporary Chinese* (*Xiandai Hanyu Dongci Dacidian*, 現代漢語動詞大詞典) [14], our semantic roles include the necessary arguments for most

[1] http://www.livac.org

verbs such as *Agent* and *Patient*, or *Goal* and *Location* in some cases; and some optional arguments realised by adjuncts, such as *Quantity*, *Instrument*, and *Source*. Some examples of semantic roles with respect to a given predicate are shown in Fig. 1.

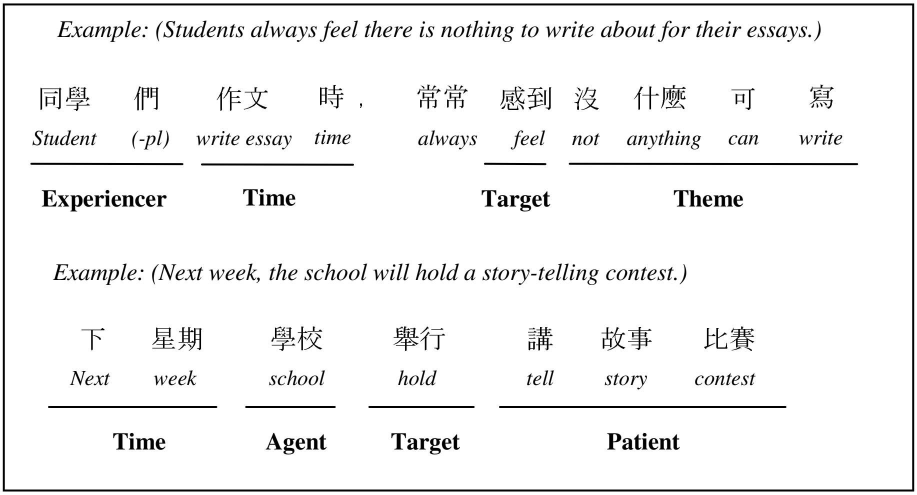

Fig. 1. Examples of semantic roles with respect to a given predicate

Altogether 980 sentences covering 41 verb types in the textbook corpus were annotated, resulting in 1,974 marked semantic roles (constituents); and 2,122 sentences covering 41 verb types in the news corpus were annotated, resulting in 4,933 marked constituents[2].

The role labelling system was trained on 90% of the sample sentences from the textbook corpus and the news corpus separately; and tested on the remaining 10% of the respective corpora.

4 Automatic Role Labelling

The automatic labelling was based on the statistical approach in Gildea and Jurafsky [6]. In Section 4.1, we will briefly mention the features employed in the training process. Then in Sections 4.2 and 4.3, we will explain our approach for locating headwords in candidate constituents associated with semantic roles, in the absence of parse information.

4.1 Training

In this study, our probability model was based mostly on parse-independent features extracted from the training sentences, namely:

---

[2] These figures only refer to the samples used in the current study. In fact over 35,000 sentences in the LIVAC corpus have been semantically annotated, covering about 1,500 verb types and about 80,000 constituents were marked.

***Headword (head)*:** The headword from each constituent marked with a semantic role was identified. For example, in the second sentence in Fig. 1, 學校 (school) is the headword in the constituent corresponding to the *Agent* of the verb 舉行 (hold), and 比賽 (contest) is the headword of the noun phrase corresponding to the *Patient*.

***Position (posit)*:** This feature shows whether the constituent being labelled appears before or after the target verb. In the first example in Fig. 1, the *Experiencer* and *Time* appear on the left of the target, while the *Theme* is on its right.

***POS of headword (HPos)*:** Without features provided by the parse, such as phrase type or parse tree path, the POS of the headword of the labelled constituent could provide limited syntactic information.

***Preposition (prep)*:** Certain semantic roles like *Time* and *Location* are often realised by prepositional phrases, so the preposition introducing the relevant constituents would be an informative feature.

Hence for automatic labelling, given the target verb *t*, the candidate constituent, and the above features, the role *r* which has the highest probability for $P(r \mid head, posit, HPos, prep, t)$ will be assigned to that constituent. In this study, however, we are also testing with the unknown boundary condition where candidate constituents are not available in advance, hence we attempt to partially locate them by identifying their headwords to start with. Our approach is explained in the following sections.

4.2 Locating Candidate Headwords

In the absence of parse information, and with constituent boundaries unknown, we attempt to partially locate the candidate constituents by trying to identify their corresponding headwords first. Sentences in our test data were segmented into words and POS-tagged. We thus divide the recognition process into two steps, locating the headword of a candidate constituent first, and then expanding from the headword to determine its boundaries.

Basically, if we consider every word in the same sentence as the target verb (both to its left and to its right) a potential headword for a candidate constituent, what we need to do is to find out the most probable words in the sentence to match against individual semantic roles. We start with a feature set with more specific distributions, and back off to feature sets with less specific distributions. Hence in each round we look for

$$\arg\max_{r} P(r \mid feature\ set)$$

for every candidate word. Ties are resolved by giving priority to the word nearest to the target verb in the sentence.

Fig. 2 shows an example illustrating the procedures for locating candidate headwords. The target verb is 發現 (discover). In the first round, using features *head*, *posit*, *HPos*, and *t*, 時候 (time) and 問題 (problem) were identified as *Time* and *Patient* respectively. In the fourth subsequent round, backing off with features *posit* and *HPos*, 我們 (we) was identified as a possible *Agent*. In this round a few other words were identified as potential *Patients*. However, since *Patient* was already located in the previous round, those come up in this round are not considered. So in the end the headwords identified for the test sentence are 我們 (we) for *Agent*, 問題 (problem) for *Patient* and 時候 (time) for *Time*.

Sentence:
溫習的時候，我們發現了許多平時沒有想到，或是未能解決的問題，於是就去問爸爸。
During revision, we discover a lot of problems which we have not thought of or cannot be solved, then we go and ask father.

Candidate Headwords	Round 1	…	Round 4	Final Result
溫習 (revision)			~~Patient~~	
時候 (time)	Time		----	**Time**
我們 (we)			Agent	**Agent**
平時 (normally)				
想到 (think)			~~Patient~~	
能 (can)				
解決 (solve)			~~Patient~~	
問題 (problem)	Patient		----	**Patient**
去 (go)			~~Patient~~	
問 (ask)			~~Patient~~	
爸爸 (father)			~~Patient~~	

Fig. 2. Example illustrating the procedures for locating candidate headwords

4.3 Constituent Boundary

Upon the identification of headwords for potential constituents, the next step is to expand from these headwords for constituent boundaries. Although we are not doing this step in the current study, it can potentially be done via some finite state techniques, or better still, with shallow syntactic processing like simple chunking if available.

5 The Experiment

5.1 Testing

The system was trained and tested on the textbook corpus and the news corpus respectively. The testing was done under the "known constituent" and "unknown constituent" conditions. The former essentially corresponds to the known-boundary condition in related studies; whereas in the unknown-constituent condition, which we will call "headword location" condition hereafter, we tested our method of locating candidate headwords as explained above in Section 4.2. In this study, every noun, verb, adjective, pronoun, classifier, and number within the test sentence containing the target verb was considered a potential headword for a candidate constituent corresponding to some semantic role. The performance was measured in terms of the precision (defined as the percentage of correct outputs among all outputs), recall (defined as the percentage of correct outputs among expected outputs), and F_1 score which is the harmonic mean of precision and recall.

5.2 Results

The results are shown in Table 1, for testing on both the textbook corpus and the news corpus under the known constituent condition and the headword location condition.

Table 1. Results on two datasets for known constituents and headword location

	Textbook Data			**News Data**		
	Precision	*Recall*	F_1	*Precision*	*Recall*	F_1
Known Constituent	93.85	87.50	90.56	90.49	87.70	89.07
Headword Location	46.12	61.98	52.89	38.52	52.25	44.35

Under the known constituent condition, the results were good on both datasets, with an F_1 score of about 90. This is comparable or even better to the results reported in related studies for known boundary condition. The difference is that we did not use any parse information in the training, not even phrase type. Our results thus suggest that for Chinese, even without more complicated syntactic information, simple lexical information might already be useful in semantic role tagging.

Comparison of the known constituent condition with the headword location condition shows that performance for the latter has expectedly dropped. However, the degradation was less serious with simple sentences than with complex ones, as is seen from the higher precision and recall for textbook data than for news data under the headword location condition. What is noteworthy here is that recall apparently deteriorated less seriously than precision. In the case of news data, for instance, we were able to maintain over 50% recall but only obtained about 39% precision. The surprisingly low precision is attributed to a technical inadequacy in the way we break ties. In this study we only make an effort to eliminate multiple tagging of the same role to the same target verb in a sentence on either side of the target verb, but not if they appear on both sides of the target verb. This should certainly be dealt with in future experiments. The differential degradation of performance between textbook data and news data also suggests the varied importance of constituent boundaries to simple sentences and complex ones, and hence possibly their varied requirements for full parse information for the semantic labelling task.

6 Discussion

According to Carreras and Màrquez [3], the state-of-the-art results for semantic role labelling systems based on shallow syntactic information is about 15 lower than those with access to gold standard parse trees, i.e., around 60. Our experimental results for the headword location condition, with no syntactic information available at all, give an F_1 score of 52.89 and 44.35 respectively for textbook data and news data. This further degradation in performance is nevertheless within expectation, but whether this is also a result of the difference between English and Chinese remains to be seen.

In response to the questions raised in the introduction, firstly, the results for the known constituent condition (F_1 of 90.56 and 89.07 for textbook data and news data respectively) have shown that even if we do not use parse-dependent features such as governing category and parse tree path, results are not particularly affected. In other words, lexical features are already very useful as long as the constituent boundaries are given. Secondly, in the absence of parse information, the results of identifying the relevant headwords in order to partially locate candidate constituents were not as satisfactory as one would like to see. One possible way to improve the results, as suggested above, would be to improve the handling of ties. Other possibilities including a class-based method could also be used, as will be discussed below. Thirdly, results for news data degraded more seriously than textbook data from the known constituent condition to the headword location condition. This suggests that complex sentences in Chinese are more affected by the availability of full parse information. To a certain extent, this might be related to the relative flexibility in the syntax-semantics interface of Chinese; hence when a sentence gets more complicated, there might be more intervening constituents and the parse information would be useful to help identify the relevant ones in semantic role labelling.

In terms of future development, apart from improving the handling of ties in our method, as mentioned in the previous section, we plan to expand our work in several respects, the major part of which is on the generalization to unseen headwords and unseen predicates. As is with other related studies, the examples available for training for each target verb are very limited; and the availability of training data is also insufficient in the sense that we cannot expect them to cover all target verb types. Hence it is very important to be able to generalize the process to unseen words and predicates. To this end, we will experiment with a semantic lexicon like *Tongyici Cilin* (同義詞詞林, a Chinese thesaurus) in both training and testing, which we expect to improve the overall performance.

Another area of interest is to look at the behaviour of near-synonymous predicates in the tagging process. Many predicates may be unseen in the training data, but while the probability estimation could be generalized from near-synonyms as suggested by a semantic lexicon, whether the similarity and subtle differences between near-synonyms with respect to the argument structure and the corresponding syntactic realisation could be distinguished would also be worth studying. Related to this is the possibility of augmenting the feature set with semantic features. Xue and Palmer [20], for instance, looked into new features such as syntactic frame, lexicalized constituent type, etc., and found that enriching the feature set improved the labelling performance.

Another direction of future work is on the location of constituent boundaries upon the identification of the headword. As mentioned earlier on, this could probably be tackled by some finite state techniques or with the help of simple chunkers.

7 Conclusion

The study reported in this paper has thus tackled the unknown constituent boundary condition in semantic role labelling for Chinese, by attempting to locate the corresponding headwords first. We experimented with both simple and complex data. Using only parse-independent features, our results on known boundary condition are comparable to those reported in related studies. Although the results for headword location condition were not as good as state-of-the-art performance with shallow syntactic information, we have nevertheless suggested some possible ways to improve the results. We have further observed that the influence of full syntactic information is more serious for complex data than simple data, which might be a consequence of the characteristic syntax-semantics interface of Chinese. As a next step, we plan to explore some class-based techniques for the task, with reference to existing semantic lexicons.

Acknowledgements

This work is supported by Competitive Earmarked Research Grants (CERG) of the Research Grants Council of Hong Kong under grant Nos. CityU1233/01H and CityU1317/03H.

References

1. Baldewein, U., Erk, K., Padó, S. and Prescher, D. (2004) Semantic Role Labelling With Chunk Sequences. In *Proceedings of the Eighth Conference on Computational Natural Language Learning (CoNLL-2004)*, Boston, Massachusetts, pp.98-101.
2. Baker, C.F., Fillmore, C.J. and Lowe, J.B. (1998) The Berkeley FrameNet Project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics (COLING-ACL '98)*, Montreal, Quebec, Canada, pp.86-90.
3. Carreras, X. and Màrquez, L. (2004) Introduction to the CoNLL-2004 Shared Task: Semantic Role Labeling. In *Proceedings of the Eighth Conference on Computational Natural Language Learning (CoNLL-2004)*, Boston, Massachusetts, pp.89-97.
4. Chen, F-Y., Tsai, P-F., Chen, K-J. and Huang, C-R. (1999) Sinica Treebank (中文句結構樹資料庫的構建). *Computational Linguistics and Chinese Language Processing, 4(2)*: 87-104.
5. Fellbaum, C., Palmer, M., Dang, H.T., Delfs, L. and Wolf, S. (2001) Manual and Automatic Semantic Annotation with WordNet. In *Proceedings of the NAACL-01 SIGLEX Workshop on WordNet and Other Lexical Resources*, Invited Talk, Pittsburg, PA.
6. Gildea, D. and Jurafsky, D. (2002) Automatic Labeling of Semantic Roles. *Computational Linguistics, 28(3)*: 245-288.
7. Gildea, D. and Palmer, M. (2002) The Necessity of Parsing for Predicate Argument Recognition. In *Proceedings of the 40th Meeting of the Association for Computational Linguistics (ACL-02)*, Philadelphia, PA.
8. Gildea, D. and Hockenmaier, J. (2003) Identifying Semantic Roles Using Combinatory Categorial Grammar. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, Sapporo, Japan.

9. Hacioglu, K., Pradhan, S., Ward, W., Martin, J.H. and Jurafsky, D. (2004) Semantic Role Labeling by Tagging Syntactic Chunks. In *Proceedings of the Eighth Conference on Computational Natural Language Learning (CoNLL-2004)*, Boston, Massachusetts, pp.110-113.
10. Higgins, D. (2004) A transformation-based approach to argument labeling. In *Proceedings of the Eighth Conference on Computational Natural Language Learning (CoNLL-2004)*, Boston, Massachusetts, pp.114-117.
11. Kingsbury, P. and Palmer, M. (2002) From TreeBank to PropBank. In *Proceedings of the Third Conference on Language Resources and Evaluation (LREC-02)*, Las Palmas, Canary Islands, Spain.
12. Kwon, N., Fleischman, M. and Hovy, E. (2004) SENSEVAL Automatic Labeling of Semantic Roles using Maximum Entropy Models. In *Proceedings of the Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (SENSEVAL-3)*, Barcelona, Spain, pp.129-132.
13. Kwong, O.Y. and Tsou, B.K. (2003) Categorial Fluidity in Chinese and its Implications for Part-of-speech Tagging. In *Proceedings of the Research Note Session of the 10th Conference of the European Chapter of the Association for Computational Linguistics*, Budapest, Hungary, pp.115-118.
14. Lin, X., Wang, L. and Sun, D. (1994) *Dictionary of Verbs in Contemporary Chinese*. Beijing Language and Culture University Press.
15. Litkowski, K.C. (2004) SENSEVAL-3 Task: Automatic Labeling of Semantic Roles. In *Proceedings of the Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (SENSEVAL-3)*, Barcelona, Spain, pp.9-12.
16. Moldovan, D., Girju, R., Olteanu, M. and Fortu, O. (2004) SVM Classification of FrameNet Semantic Roles. In *Proceedings of the Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (SENSEVAL-3)*, Barcelona, Spain, pp.167-170.
17. Swier, R.S. and Stevenson, S. (2004) Unsupervised Semantic Role Labelling. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, Barcelona, Spain, pp.95-102.
18. Tsou, B.K., Tsoi, W.F., Lai, T.B.Y., Hu, J. and Chan, S.W.K. (2000) LIVAC, A Chinese Synchronous Corpus, and Some Applications. In *Proceedings of the ICCLC International Conference on Chinese Language Computing*, Chicago, pp. 233-238.
19. Williams, K., Dozier, C. and McCulloh, A. (2004) Learning Transformation Rules for Semantic Role Labeling. In *Proceedings of the Eighth Conference on Computational Natural Language Learning (CoNLL-2004)*, Boston, Massachusetts, pp.134-137.
20. Xue, N. and Palmer, M. (2004) Calibrating Features for Semantic Role Labeling. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, Barcelona, Spain, pp.88-94.
21. You, J-M. and Chen, K-J. (2004) Automatic Semantic Role Assignment for a Tree Structure. In *Proceedings of the 3rd SigHAN Workshop on Chinese Language Processing*, ACL-04, Barcelona, pp.109-115.
22. 啓思中國語文 *Qisi Zhongguo Yuwen*. Primary 1-6, 24 volumes, 2004. Hong Kong: Keys Press.
23. 現代中國語文 *Xiandai Zhongguo Yuwen*. Primary 1-6, 24 volumes, 2004. Hong Kong: Modern Education Research Society Ltd.