

Data Homogeneity and Semantic Role Tagging in Chinese

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 tackle the task by identifying the relevant headwords in a sentence as a first step to partially locate the corresponding constituents to be labelled. We also explore the effect of data homogeneity by experimenting with a textbook corpus and a news corpus, representing simple data and complex data respectively. Results suggest that while the headword location method remains to be improved, the homogeneity between the training and testing data is important especially in view of the characteristic syntax-semantics interface in Chinese. We also plan to explore some class-based techniques for the task with reference to existing semantic lexicons, and to modify the method and augment the feature set with more linguistic input.

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 becomes indispensable. The expansion of the Penn Treebank into a Proposition Bank (Kingsbury and Palmer, 2002) is a typical move in this direction. Lexical resources also need to be enhanced with semantic information (e.g. Fellbaum *et al.*, 2001). The ability to identify semantic role relations cor-

rectly 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 calls for automatic labelling to streamline the process. The task is essentially done in two phases, namely *recognising* the constituents bearing some semantic relationship to the target verb in a sentence, and then *labelling* them with the corresponding semantic roles.

In their seminal proposal, Gildea and Jurafsky (2002) approached the task using various features such as headword, phrase type, and parse tree path. While such features have remained the basic and essential features in subsequent research, parsed sentences are nevertheless 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 (Gildea and Palmer, 2002; Gildea and Hockenmaier, 2003).

More precisely, parse information is rather more critical for the identification of boundaries of candidate constituents than for the extraction of training data. Its limited function in training, for instance, is reflected in the low coverage reported (e.g. You and Chen, 2004). As full parses are not always accessible, many thus resort to shallow syntactic information from simple chunking, even 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 con-

siderable variability in its syntax-semantics interface; and when one comes to more nested and complex sentences such as those from news articles, it becomes more difficult to capture the sentence structures by typical examples.

Thus in the current study, we approach the problem in Chinese in the absence of parse information, and attempt to identify the headwords in the relevant constituents in a sentence to be tagged as a first step. In addition, we will explore the effect of training on different datasets, simple or complex, to shed light on the relative importance of parse information for indicating constituent boundaries in semantic role labelling.

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 (2002), for instance, used a set of roles defined according to the FrameNet model (Baker *et al.*, 1998), 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 (2002) 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, as in the CoNLL-2004 shared task (Carreras and Màrquez, 2004) and SENSEVAL-3 (Litkowski, 2004).

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 (You and Chen, 2004). In our present study, we go for a set of more abstract semantic roles similar to the thematic roles for English used in VerbNet (Kipper *et al.*, 2002). These roles are generalisable to most Chinese verbs and are not

dependent on particular predicates. They will be further introduced in Section 3.

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 (2002), 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:

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

Subsequent studies exploited a variety of implementation of the learning component. Transformation-based approaches were also used (e.g. see Carreras and Màrquez (2004) for an overview of systems participating in the CoNLL shared task). Swier and Stevenson (2004) 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 trick often lies in locating the candidate constituents to be labelled. A parser of some kind is needed. Gildea and Palmer (2002) 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. Sun and Jurafsky (2004) also reported a drop in F -score with automatic syntactic parses compared to perfect parses for role labelling in Chinese, despite the comparatively good results of their parser (i.e. the Collins parser ported to Chinese). The necessity of parse information is also reflected from recent evaluation exercises. For instance, 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 use 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 (2004), 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.

In this work, we study the problem in Chinese, treating it as a headword identification and labelling task in the absence of parse information, and examine how the nature of the dataset could affect the role tagging performance.

3 The Data

3.1 Materials

In this study, we used two datasets: 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 and Modern Education Research Society Ltd. Texts for Primary One to Six were digitised, segmented into words, and annotated with parts-of-speech (POS). This results in 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 corpus¹ (Tsou *et al.*, 2000; Kwong and Tsou, 2003). 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, resulting in 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.

¹ <http://www.livac.org>

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 human annotator and the 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 - Lin et al., 1994)*, our semantic roles include the necessary arguments for most 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 Figure 1.

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 constituents².

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 both corpora.

4 Automatic Role Labelling

The automatic labelling was based on the statistical approach in Gildea and Jurafsky (2002). In Section 4.1, we will briefly mention the features used 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.

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.

4.1 Training

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

Headword (*head*): The headword from each constituent marked with a semantic role was identified. For example, in the second sentence in Figure 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 Figure 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 $P(r | head, posit, HPos, prep, t)$ 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. To start with, we attempt to partially locate them by identifying their headwords first, as explained in the following sections.

<i>Example: (Students always feel there is nothing to write about for their essays.)</i>						
同學 們	作文 時	常 常	感 到	沒 什 麼	可 寫	
<i>Student (-pl)</i>	<i>write essay time</i>	<i>always</i>	<i>feel</i>	<i>(neg) anything</i>	<i>can</i>	<i>write</i>
Experiencer	Time		Target	Theme		
<i>Example: (Next week, the school will hold a story-telling contest.)</i>						
下 星期	學 校	舉 行	講 故 事	比 賽		
<i>Next week</i>	<i>school</i>	<i>hold</i>	<i>tell story</i>	<i>contest</i>		
Time	Agent	Target	Patient			

Figure 1 Examples of semantic roles with respect to a given predicate

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 identifying 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 with 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 distributions³. Hence in each round we look for

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

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

Figure 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, they would not be considered since Patient was already located in a previous round. So in the end the headwords identified for the test sentence are 我們 for Agent, 問題 for Patient and 時候 for Time.

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.

³ In this experiment, we back off in the following order: $P(r \mid \text{head}, \text{posit}, \text{HPos}, \text{prep } t)$, $P(r \mid \text{head}, \text{posit}, t)$, $P(r \mid \text{head}, t)$, $P(r \mid \text{HPos}, \text{posit}, t)$, $P(r \mid \text{HPos}, t)$. However, the *prep* feature becomes obsolete when constituent boundaries are unknown.

5 The Experiment

5.1 Testing

The system was trained on the textbook corpus and the news corpus separately, and tested on both corpora (the data is *homogeneous* if the system is trained and tested on materials from the same source). The testing was done under the “known constituent” condition and “unknown constituent” condition. 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 Tables 1 and 2, for training on homogeneous dataset and different dataset respectively, and testing under the known constituent condition and the headword location condition.

When trained on homogeneous data, the results were good on both datasets under the known constituent condition, 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. When trained on a different dataset, however, the accuracy was maintained for textbook data, but it decreased for news data, for the known constituent condition.

For the headword location condition, the performance in general was expectedly inferior to that for the known constituent condition. Moreover, this degradation seemed to be quite consistent in most cases, regardless of the nature of the training set. In fact, despite the effect of training set on news data, as mentioned above, the degradation

from known constituent to headword location is nevertheless the least for news data when trained on different materials.

Hence the effect of training data is only obvious in the news corpus. In other words, both sets of training data work similarly well with textbook test data, but the performance on news test data is worse when trained on textbook data. This is understandable as the textbook data contain fewer examples and the sentence structures are usually much simpler than those in newspapers. Hence the system tends to miss many secondary roles like location and time, which are not sufficiently represented in the textbook corpus. The conclusion that training on news data gives better result might be premature at this stage, given the considerable dif-

ference in the corpus size of the two datasets. Nevertheless, the deterioration of results on textbook sentences, even when trained on news data, is simply reinforcing the importance of data homogeneity, if nothing else. More on data homogeneity will be discussed in the next section.

In addition, the surprisingly low precision under the headword location condition is attributable 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.

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.

<i>Candidate Headwords</i>	<i>Round 1</i>	<i>...</i>	<i>Round 4</i>	<i>Final Result</i>
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	

Figure 2 Example illustrating the procedures for locating candidate headwords

	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

Table 1 Results for Training on Homogeneous Datasets

	Textbook Data			News Data		
	Precision	Recall	F ₁	Precision	Recall	F ₁
Known Constituent	91.85	88.02	89.86	80.30	66.80	72.93
Headword Location	38.87	57.29	46.32	37.89	42.01	39.84

Table 2 Results for Training on Different Datasets

6 Discussion

6.1 Role of Parse Information

According to Carreras and Màrquez (2004), 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. With homogeneous training and testing data, our experimental results for the headword location condition, with no syntactic information available at all, give an F₁ score of 52.89 and 44.35 respectively for textbook data and news data. Such results are in line with and comparable to those reported for the unknown boundary condition with automatic parses in Gildea and Palmer (2002), for instance. Moreover, when they used simple chunks instead of full parses, the performance resulted in a drop to below 50% precision and 35% recall with relaxed scoring, hence their conclusion on the necessity of a parser.

The more degradation in performance observed in the news data is nevertheless within expectation, and it suggests that simple and complex data seem to have varied dependence on parse information. We will further discuss this below in relation to data homogeneity.

6.2 Data Homogeneity

The usefulness of parse information for semantic role labelling is 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.).

As reflected from the results, the nature of training data is obviously more important for the news data than the textbook data; and the main reason might be the failure of the simple training

data to capture the many complex structures of the news sentences, as we suggested earlier. The relative flexibility in the syntax-semantics interface of Chinese is particularly salient; 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.

With respect to the data used in the experiment, we tried to explore the complexity in terms of the average sentence length and number of semantic role patterns exhibited. For the news data, the average sentence length is around 59.7 characters (syllables), and the number of semantic role patterns varies from 4 (e.g. 打算 ‘to plan’) to as many as 25 (e.g. 進行 ‘to proceed with some action’), with an average of 9.5 patterns per verb. On the other hand, the textbook data give an average sentence length of around 39.7 characters, and the number of semantic role patterns only varies from 1 (e.g. 決定 ‘to decide’) to 11 (e.g. 舉行 ‘to hold some event’), with an average of 5.1 patterns per verb. Interestingly, the verb 進行, being very polymorphous in news texts, only shows 5 different patterns in textbooks.

Thus the nature of the dataset for semantic role labelling is worth further investigation. The design of the method and the feature set should benefit from more linguistic analysis and input.

6.3 Future Work

In terms of future development, apart from improving the handling of ties in our method, as mentioned above, we plan to expand our work in several respects. The major part would be 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. Xue and Palmer (2004), 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. In particular, given the importance of data homogeneity as observed from the experimental results, and the challenges posed by the characteristic nature of Chinese, we intend to improve our method and feature set with more linguistic consideration.

7 Conclusion

The study reported in this paper has thus tackled semantic role labelling in Chinese in the absence of parse information, by attempting to locate the corresponding headwords first. We experimented with both simple and complex data, and have explored the effect of training on different datasets. Using only parse-independent features, our results under the known boundary condition are comparable to those reported in related studies. The headword location method can be further improved. More importantly, we have observed the importance of data homogeneity, which is especially salient given the relative flexibility of Chinese in its syntax-semantics interface. As a next step, we plan to explore some class-based techniques for the task with reference to existing semantic lexicons, and to modify the method and augment the feature set with more linguistic input.

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

- 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.
- 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.
- 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.
- Gildea, D. and Jurafsky, D. (2002) Automatic Labeling of Semantic Roles. *Computational Linguistics*, 28(3): 245-288.
- 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.
- Gildea, D. and Hockenmaier, J. (2003) Identifying Semantic Roles Using Combinatory Categorical Grammar. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, Sapporo, Japan.
- 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.
- Kipper, K., Palmer, M. and Rambow, O. (2002) Extending PropBank with VerbNet Semantic Predicates. In *Proceedings of the AMTA-2002 Workshop on Applied Interlinguas*, Tiburon, CA.
- Kwong, O.Y. and Tsou, B.K. (2003) Categorical 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, pages 115-118.
- Lin, X., Wang, L. and Sun, D. (1994) *Dictionary of Verbs in Contemporary Chinese*. Beijing Language and Culture University Press.
- 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.
- Sun, H. and Jurafsky, D. (2004) Shallow Semantic Parsing of Chinese. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2004)*, Boston, pp.249-256.
- 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.
- 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.
- 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.
- 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.
- 啓思中國語文 *Qisi Zhongguo Yuwen*. Primary 1-6, 24 volumes, 2004. Hong Kong: Keys Press.
- 現代中國語文 *Xiandai Zhongguo Yuwen*. Primary 1-6, 24 volumes, 2004. Hong Kong: Modern Education Research Society Ltd.