International Journal of Knowledge and Language Processing
Volume 4, Number 4, **2013**

www.ijklp.org
KLP International ©2013 ISSN 2191-2734
pp.**33–43**

Construction and Application of the Chinese Function Word Usage Knowledge Base

Zhang Kunli, Zan Hongying, Chai Yumei, Han Yingjie and Zhao Dan

College of Information Engineering
Zhengzhou University
Zhengzhou, Henan
China
{ieklzhang, iehyzan, ieymchai, ieyjhan, iedzhao}@zzu.edu.cn

Received June 2013; revised August 2013

Abstract. *The contemporary Chinese function words with distinct usages play complex syntactic roles and have strong individual characteristics. Because of this, the detailed usage description and formalized representation of Chinese function words are of great significance in Chinese syntax analysis and semantic understanding. In this paper, the authors firstly review the current research of Chinese function words and lexical knowledge base and secondly they describe the construction process about the triune usage knowledge base of contemporary Chinese function words and define knowledge base as a combination of the usage dictionary, the usage rule base and the usage corpus. Then they present the problems existing in the knowledge base. On the basis of the finished knowledge base, the authors study the automatic identification of the usage of the Chinese function words and discuss the potential applications of their knowledge base.*

Keywords: Function Word Usage Knowledge Base; Usage Dictionary; Usage Rule Base; Usage Corpus; Usage automatic tagging

1. **Introduction.** In Chinese, function words play an important role in describing the relationship between content words and their meanings are intangible. Chinese is an analytic language and lacks morphological changes in the strict sense [1]. In comparison with other languages, such as English, Chinese function words undertake a more onerous grammatical task and play a key role in text semantic understanding and grammatical analysis. So research on function words is an important part in the study of contemporary Chinese. In the study of Chinese function words, we should not only understand the meaning of each function word but also pay attention to the usage of each function word [2].

Meanwhile language processing systems need knowledge bases for support [3]. Particularly lexical knowledge base plays an important role in the natural language processing.

According to the needs of research on natural language processing and its application, the Chinese function words in this paper include adverbs, prepositions, conjunctions, auxiliaries, modal particles and location words. In order to build the triune knowledge base of contemporary Chinese function words [4], we began with investigating the usage of function words and constructed the Chinese Function words usage Knowledge Base, abbreviated as CFKB, which includes a usage dictionary, a usage rule base and a usage corpus.

The rest of the paper is organized as follows. In Section 2, we reviewed the related works of Chinese function words and the Chinese knowledge base. In section 3, we gave the construction process of CFKB and its recent results in CFKB. In section 4, we gave the results of automatic identification of function words usage on the base of CFKB and furthermore we studied the application of CFKB. At last, we drew a conclusion and listed further works.

2. **Related Work.** In the field of linguistics, many researchers have been focusing on the studies on the meaning and the usage of Chinese function words. There are some function words dictionaries [5-8], in which the function words are differentiated, analyzed and discussed in detail in combination with examples. There are also some research papers and monographs focused on detailed description of function words. The above studies are all human-oriented. Representatives of the contemporary Chinese lexical knowledge bases mainly include: Hownet, CFN (Chinese FrameNet)，TongYiCi CiLin and GKB(The Grammatical Knowledge base of Contemporary Chinese).These above lexical knowledge bases are weak in terms of Chinese function words information included [4].

In conclusion, it is the urgent demand of natural language processing to construct perfect Chinese function words lexical knowledge base. In 2003，the “trinity” design concept of CFKB was prompted by Yu [3]. The attributes of the machine dictionary, the proper-scope of corpus and rule base in CFKB were discussed in Liu [9]. Peng [10] studied grammatical functions of Chinese prepositions, and gave the preliminary machine dictionary and rule base of Chinese prepositions. These are the foundations of the CFKB construction in this paper.

3. **The Construction of CFKB.**

3.1. **The Foundation of CFKB Construction.** CFKB includes a function words usage dictionary, a function words usage rule base and a function words usage tagged corpus. The dictionary was firstly built. Then on the basis of the usage description in dictionary the rule base was built, which was used in rule-based method to automatically annotate the usages in corpus. The corpus was artificially proofread crossly with more than two people, then the standard corpus of identification of usage was formed. And the usage dictionary and usage rule base were modified and adjusted according to the proofreading feedback. The specific steps of construction process are as follows:

1) The framework was designed according to grammar characteristics of different kinds of POS words.
2) Referring to classical documents and corpus, the entries in function words usage dictionary were determined and the contents of the attributes were fulfilled.
3) According to the usage description in dictionary, the usage rule base which obeyed the specification of usage rule description was artificially built.
4) The usage automatic identification algorithm which was based on usage rules was developed. And using this algorithm the function words in one month's segmented and POS tagged People's Daily were automatically annotated.
5) The corpus was artificially proofread with more than two people. And the proofread specification was formed in this process.
6) According to the experience of artificial proofreading process, the contents in usage dictionary and the usage rule base were modified and adjusted;
7) The next month's corpus would be automatically annotated using new usage rule base. Then turn to step 5.

Following above steps, which lasted 7 years, the usage dictionary, the usage rule base and the corpus including 7 months' People's Daily in which function words usage were tagged have been completed, and have been gradually improved.

3.2. **Contemporary Chinese Function Words Usage Dictionary.** The construction of contemporary Chinese function words usage dictionary involved three steps: design framework, fulfill content and modify dictionary according to feedback.

Designing Framework

The attributes in usage dictionary were designed into four categories: ID, usage description, syntactic function description and category. The coding principle of ID is same for six POS words. Each usage has a unique ID in the usage dictionary, which links the usage dictionary, the rule base and the corpus. In general, the coding frame of ID is "p_z[_tn] [_m] [x] [y]", where "p" represents the word's POS, "z" represents the word's PINYIN(Chinese Pronouncing notation), "t" shows it is a homophone with other word, and "n" represents its sequent number, "m" represents the sequent number of word sense, "x" represents the sequent number(using a, b, c, d…) of usage in one sense, "y" represents the sequent number(using a, b, c, d...) of divided usage in one usage and "[]" represents optional in the frame. The further detail about ID can be seen in [11]. Usage description, syntactic function description and category in dictionary vary with each POS. For example, With regard to category attributes, the usages of conjunction focus on the relationship, the usages of adverb on the classification, and the usages of preposition on the object type. The further detail about dictionary framework can be seen in [11-13]. As a whole, in the framework designs, all word classes in CFKB have both uniform attributes and different ones because of their distinctive properties, which enables CFKB to serve a more important function in natural language processing.

Fulfilling Content

After the framework of usage dictionary was determined, on the basis of Liu and Peng's work, referred to GKB(<y>), Lv [5](<b>), Zhang [8](<x>), "Contemporary Chinese Dictionary"(5th edition<h>), and the segmentation and POS corpus of People's Daily(<r>), Jan 1998, and Jan to Jun 2000, combined with the practical usage of Chinese function, every usage of a function word was decomposed and regarded as a record in usage dictionary. In view of the different opinions of different scholars, on the basis of the demand for language processing, the usage descriptions of every word are decomposed, differentiated, and then fulfilled into the dictionary after the operational features were extracted. In order to record the information source in usage dictionary, the corresponding content markers "<y>", "<b>", "<h>", "<x>" and "<r>" represents the five different sources above. If the source is not in above five, the corresponding content will be marked "<z>". The example of function words usage dictionary is shown in Fig. 1.

	A	B	C	D	E
1	ID	词语	释义	例句	用法
2	d_dou1_1	都	表示总括全部。<b>	大伙儿~同意<b>|一天功夫把这些事~办完了<b>|只要除问话以外，所总括的对象必须放在“都”前。也可以说“全都”，总括的意思更明显。<b>
3	d_dou1_1a	都	表示总括全部。<b>	给谁~行<b>|怎么办~可以<b>	所总括的对象可以用表示任指的疑问指代词。<b>
4	d_dou1_1b	都	表示总括全部。<b>	不论大小工作，我们~要把它做好<b>|无论干什么事情	所总括的对象前可以用连词“不论、不管、无论、凡是、只要”。<b><z>
5	d_dou1_1c	都	表示总括全部。<b>	你~去过哪儿？<b>|老王刚才~说了些什么？<b>	问话时总括的对象（疑问代词）放在“都”后。<b>
6	d_dou1_1d	都	表示总括全部。与“是”搭配	~是你一句话把他惹翻了<b>	与“是”搭配。<b>
7	d_dou1_2	都	甚至。“都”轻读。<b>	我~不知道你会来<b>|真抱歉，我~忘了你的名字了<b	修饰动词或动词短语。<z>
8	d_dou1_2a	都	甚至。“都”轻读。与“连”	连这么重的病~给治好了<b>|连书包里的东西~淋湿了	与“连”字同用。<b>
9	d_dou1_2b	都	甚至。“都”轻读。<b>	你怎么问~不问我一声？<b>|拉~拉不住他<b>|看~没1	“都”前后用同一个动词（前一肯定，后一否定）。<b>A~A(不|没|没有|未|dé)或A~(不|没|没有|未|dé)#A 。
10	d_dou1_2c	都	甚至。“都”轻读。<b>	一口~没喝<b>|一个人~不见<b>|我们那时没有，一本一+量词+…~+动词（否定式）。<b>	
11	d_dou1_2d	都	甚至。“都”轻读。用于表示	你~搬不动，我更不行了<b>|为了营造绿化林带，家~	用于表示让步的小句，引出表示主要意思的小句。<b>
12	d_dou1_3	都	已经。<b>	饭~凉了，快吃吧！<b>|~十二点了，还不睡！<b>	句末常用“了”。<b>

FIG.1. EXAMPLE OF FUNCTION WORDS USAGE DICTIONARY

Modifying Dictionary according to Feedback

Firstly, by means of the feedback form proofread corpus process, in the usage dictionary some words are added and other words are deleted. For example, in the corpus if the word "又[*you*](*and*)" and "既[*ji*](*both*)" occur together, the POS of "又[*you*](*and*)" is labeled "/c"(conjunction). Such as:

这样/rz 既/c 方便/v 广大/b 市民/n 参加/v 活动/vn ，/wd 又/c 能/vu 更/d 好/a 地/ui 维持/v 秩序/n ，/wd 确保/v 安全/an 。/wj（20000101-10-014-006/m）(*So this can both offer convenience for the masses to take part in activities and better keep order to ensure safety.*)

According to the labeled POS in corpus, the word "又[*you*](*and*)" was adopted as conjunction in 2009 version [12] dictionary. But through analysis, the semantic meaning of "又[*you*](*and*)" in the above sentence is "several action, state, condition accumulative together", which still belongs to the category of adverbs. So this word was deleted from conjunction usage dictionary. In the original usage dictionary, the word "沿途[*yantu*](*along the path*)" was adopted and the word "沿街[*yanjie*](*along the street*)" was not adopted. Through the study of corpus, these two words are same in terms of usage or attribute, So "沿街[*yanjie*](*along the street*)" was added into adverb usage dictionary.

Secondly, some usages are added and other usages are deleted. For example, when the

adverb "也[*ye*](*also*)" represented the association relationship, besides progressive, alternative, adversative, suppositional, concessions, conditional and causal relationships, the transition relationship was also found in corpus. Therefore a new usage of the word "也[*ye*](*also*)" was added in usage dictionary.

The construction of usage dictionary is a continuous improvement process. The 2007 version [11] adverb usage dictionary contains 1153 adverbs and 1946 usages. The words and usages distribution of 2009 version [12] and present version (2013) dictionary are shown in Table 1. As shown in Table 1. , except auxiliary the number of each kind of words and usage is changed significantly.

TABLE 1. THE DISTRIBUTION OF CHINESE FUNCTION WORD'S USAGES IN IN CFKB

Usages No / Words No / POS	1	2	3	4	5	6	7	8	9	10	above 10	words in total 2013	usages in total 2013	words in total 2009	usages in total 2009
Adverb	1214	179	84	38	21	12	4	3	2	1	8	1566	2356	1566	2356
Proposition	66	30	23	7	4	5	7	0	1	1	2	141	331	141	331
Conjunctio	156	50	55	24	16	7	4	0	1	2	0	315	696	315	696
Auxiliary	30	4	3	1	0	1	2	0	0	0	1	45	144	45	144
Modality	30	7	7	4	2	0	1	4	0	0	2	58	169	58	169
Locality	164	34	11	24	19	6	6	4	6	1	1	276	641	276	641

Up to now, the function words usage dictionary is starting to take shape. But there are still some problems. Such as the common words (with high frequency in corpus), on which more linguists have researched, their senses and usages are divided into smaller particle sizes than those of uncommon words. The uniform particle division standard of sense and usage should be taken into account in the future.

3.3. **Contemporary Chinese Function Words Usage Rule Base.** On the base of Chinese function word usage dictionary, three steps were taken in the construction of contemporary Chinese function words usage rule base: determine the rule description and specification, construct usage rule and modify rules according to feedback.

Determining the Rule Description and Specification

Based on usage description in Chinese function word usage dictionary, we have distilled feasible criteria including the features of the first word in a sentence (short for F), words to the left of the function word in a sentence (short for M), words close to the left of the function word in a sentence (short for L), words close to the right of the function word in a sentence (short for R), words to the right of the function word in a sentence (short for N) and end word or punctuation in a sentence (short for E), and determine the rule description in Backus Normal Form (BNF). The principles of detailed description are as follows. Here is the general form of usage rule.

@<ID>→[F][M][L][R][N][E] ^F→<word1> | <word2> | … | a | v | n | … ^M→ <word1> | <word2> | … | a | v | n | … ^L→<word1> | <word2> | … | a | v | n | …

^R→<word 1>｜<word 2>｜…｜a｜v｜n｜… ^N→< word 1>｜<word 2>｜…｜a｜v｜n｜… ^E→<word1>｜<word 2>｜…｜a｜v｜n｜…

In the usage rule, "@" represents the rule start symbol, "^" represents the connector between feature definitions which indicates the conjunction relationship of features, "ID" is the usage ID, "<word 1>" and "a, v, n " represents the words and POS which appear in the feature position respectively, the symbol "~" represents the observed function word itself , other meta symbols, such as "→", "|", "*", "( )" and "[ ]", are in the general means of BNF.

In addition to the above six features, framework and semantic field are employed in the rule, which has three description forms. The first one is: the meta symbols, "A" or "B", will be adopted in the rule if the framework is formed by same words or same POS. For example：

$不

@<d_bu4_2a>→A~A ^A→a // "A" represents the same words around "不[*bu*](*not*)", such as "干净不干净(*clean or not clean*)"

@<d_bu4_2e>→~B~B ^B→f // "A" represents the same words around "不[bu](not)", such as "不上不下(be in a dilemma)"

The second one is: the meta symbols, "T" and "S", will be adopted in the rule, with "%" as a marker if two words before and after the observed function word.

$不

@<d_bu4_2a>→%S%~%T% //Such as in phrase "吃饭不吃(*eat or not eat*)？", word "吃[*chi*](*eat*)"(T) is the subset of word "吃饭[*chifan*](*eat*)"(S).

The last one is semantic field. One semantic field is saved in a file whose name is referred in usage rule with a pair of single style quotes as marker. For example:

$十分

@<d_shi2fen1_1b>→R ^R→'xinli_v.txt' //Semantic field of psychological verbs is saved in the file named "xinli_v.txt".

Constructing Usage Rules

The usage rules are manually constructed according to the description in the dictionary and ordered for higher precision of automatic identification. We can use multiple rules to describe one usage if the usage is complex. As shown in Fig. 1, "DOU" is a Chinese adverb and has 3 senses with 11 usages. The rules of this word are shown in Fig. 2. The description of usage <d_dou1_2b> in usage dictionary is defined as "there will be the two same verbs in the right context and the left context, of which one is affirmative and another is negative". And in the different observed sentences we found that there were zero words or more than one words between "DOU" and the verb in the right context. If the verb is in the close right context, this usage will be easy to automatically identify. So there are two rules to describe the usage <d_dou1_2b>.

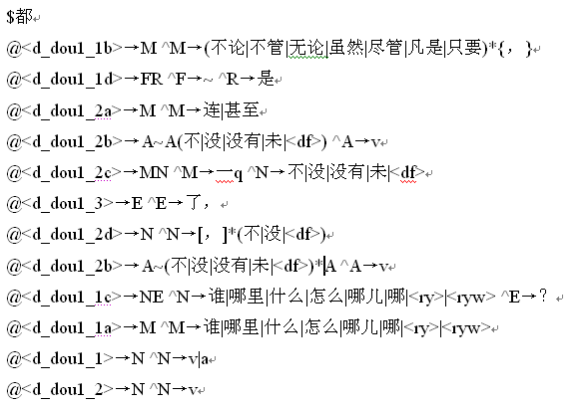

FIG.2. EXAMPLE OF USAGE RULES

Modifying Rules according to Feedback

Based on the proofread corpus, there are two ways to modify the usage rules. The first one is manual modification. As for the content of rules, we analyzed these sentences which had been tagged with error usage with rule-based method， distilled the feasible feature and then modified the usage rule. As for the order of rules, rules with high occurrence frequency or rules which have good effect in automatic identification would be put to a prior position of rules queue to ensure that each preposition is annotated by the best usage rule in automatic identification based on rules.

The second one is automatic modification. As for these function words usage that couldn't be automatically recognized with rule-based method and were tagged <FAIL> in coups, an error-driven learning approach is adopted to generate new usage rules [14]. This approach has three steps: Firstly, they set up conversion template and generate candidate rules. Secondly, they scored of candidate rules using the objective function. Then they chose the highest score rule as the update rule.

Up to now, we have finished the description of all usage rules and constructed the Chinese function word usage rule base which includes 2456 adverb rules, 385 preposition rules, 747 conjunction rules, 165 auxiliary rules, 182 modality rules and 761 locality rules.

3.4. **Contemporary Chinese Function Words Corpus.** Contemporary Chinese function words corpus is formed by usage annotating on the segmentation and POS corpus for 7 months of People's Daily, Jan, 1998, and Jan to Jun, 2000 which includes 8.7 million words. We first used the rule-based method to automatically annotate the preposition usage in corpus. Secondly the machine-tagged corpus was artificially proofread by two researchers respectively from linguistics and computer science field. Then a third person joined them to discuss inconsistent usages to both sides and determined the final usage tag. And the tagging criterion of one certain function word was formed in the discussion.

Usage annotation is to annotate the corresponding label of usage ID next to the function words. Here is an example of the Chinese function word usage corpus.

20000401-01-001-006/m　中国/ns　**和/c<c_he2_1>**　印度/ns　**都/d<d_dou1_1>**

是/vl 世界/n 文明/a 古国/n ，/wd 两/m 国/n **之间/f<f_zhi1jian1_1c>** **的/ud<u_de5_t2_1a>** 友好/a 交往/vn 源远流长/iv 。/wj (*China and India are both ancient civilizations in the world and the two peoples enjoy time-honored friendly exchanges.*)

During the process of manual proofreading, we also analysed the original word segmentation and POS tagging. If the original corpus word segmentation or POS tagging is not right, we tagged these words with "@", then we used different treatments in various situations. In addition, word segmentation or POS tagging errors can be automatically found by analysing these words which were marked with "<FAIL>" by rule-based method [15].

Up to now, the total number of function words we have tagged and proofread in seven months of People's Daily corpus is about 1.21 million, and the standard contemporary Chinese function words corpus has been formed. Although a variety of measures have been taken in the proofreading process, it is difficult to ensure that all annotations are consistent. Therefore consistency check of usage annotations is the ongoing work.

4. **Application of CFKB.**

4.1. **Automatic Identification of Function Words Usage.** Automatic identification of function word usage is an important part of the construction and application of CFKB. The study on automatic identification involves three ways: rule-base method, statistics-based method and combination of rule-base and statistics-based method.

The rule-based method is able to start reading corpus and usage rules into memory and then use six types of matcher and Special framework matcher to match and parse usage rules and determine the annotating result. Yuan [16] introduced design requirements of all kinds of verifier and realization of automatic annotating functional words usage based on rules. Zan [17], Zhou [18] and Han [19] described the detailed process of automatic usage identification based on rules for conjunction, modality and auxiliary respectively. So far the precisions to automatically recognize function words in the seven-months' People's Daily, by using rule-based method, are respectively 84.36% for adverbs, 71.71% for prepositions, 83.68% for conjunctions, 40.71% for auxiliaries, 78.85% for modalities, and 88.14% for localities.

It is well known, rule-based methods has its limitations. Using standard function word usage corpus mentioned in 3.4 as training data, statistical models, such as SVM (Support Vector Machine), ME (Maximum Entropy) and CRF (Conditional Random Fields), were adopted in the automatic identification of Chinese function word usages. Zan [17,20-21] and Zhang [22-23] studied the usages' automatic identification of adverb "就[*jiu*](*as soon as*)", adverb "才[*cai*](*just*)", common conjunctions, common adverbs and preposition "才[*zai*](*in*)" using statistics-based method and the precisions of usage identification based on statistical model were about 25% more than these of based on rule respectively.

Although the overall precision of statistics-based method is high, it could be found that some usages' results of rule-based method are better than these of statistics-based method.

Zhang [24] adopted usage distribution in corpus and the usage precisions of above two methods as parameters and studied the automatic identification on adverb "都[*dou*](*all*)". The precision reached 98.54% and are 16.54% and 8.92% higher than that of rule-based method and statistics-based method respectively. This method combined rule-based method with statistics-based method.

4.2. **Preliminary Application Research on CFKB.** The achievements of CFKB can be directly used in nature language processing. In terms of syntactic analysis, Zan [12] developed a method that used usages identification to modify the dependency syntactic analysis results that were got from LTP of Harbin institute of technology, which can improve the precision of the dependency syntactic analysis. In [25], the conjunction usage identification results were introduced to conjunction phrase structure analysis and the precision based on rules reached 48.67%. Then a statistics-based method was adopted in conjunction phrase structure analysis and conjunction usage as a feature was introduced to CRF model. Compared to not adding usage feature, the highest precision improved 4%.

Using Chinese function word usage automatic identification algorithm, the function words in text will be tagged with usage label. If we associate the tagging results with Chinese function words usage dictionary, it will not only leave certain effect on machine translation, information extraction, question answering system, and other areas of the natural language processing, but also play an assisted role among foreign Chinese teaching in the semantic understanding of Chinese function words, synonym and near righteousness function analysis, prepositions and conjunctions structure of fixed collocation and automatic analysis of function word errors, etc.

5. **Conclusion and Future Works.** The research on contemporary Chinese function words usage plays an important role in semantic and syntactic analysis. In this paper, we introduced the construction process, current situation and existing problems of CFKB which includes the Chinese function word usage dictionary, the Chinese function word usage rule base, and the Chinese function word usage corpus. In addition, automatic identification of the usages of function words was studied according to the rule base and corpus. And application of syntactic analysis using the automatic identification of function words usage was preliminarily discussed.

Next we will continue to improve the quality of CFKB, making sure the three parts of CFKB are in concordance. And we will also attempt the applications based on CFKB.

Acknowledgments. This work was supported by a grant from the Natural Science Foundation of China (No.60970083, No.61272221), 863 Projects of National High Technology Research and Development (No.2012AA011101), the Open Projects Program of National Laboratory of Pattern Identification and the Science and Technology Project of the Education Department of Henan Province (No. 12B520055, No. 13B520381).

REFERENCES

[1] Lv, S.X..: The Problem on Grammatical Analysis of the Contemporary Chinese. Commercial Press, China(1979)(In Chinese)

[2] Lu, J.M., and Ma Z.: Some Comments on the Modern Chinese Function Word. Chinese Press, China(1999)(In Chinese)

[3] Dong, Z.D., and Dong, Q., http://www.keenage.com/

[4] Yu, S.W., Zhu, X.F., Liu, Y.: Knowledge-base of Generalized Functional Words of Contemporary Chinese. Journal of Chinese Language and Computing. Vol.13 No.1: 89-98(2003)(In Chinese)

[5] Lv, S.X.: Contemporary Chinese 800 words. Commercial Press, China(1980)(in Chinese)

[6] Wu, K.Z.: The Contemporary Common Chinese Function Word Dictionary. Zhejiang Education Publishing House, China(1998)(In Chinese)

[7] Hou, X.C.: The Contemporary Chinese Function Word Dictionary. Press of Beijing University, China(1999) (In Chinese)

[8] Zhang, B.: The Contemporary Chinese Function Word Dictionary. Commercial Press, China(2001) (In Chinese)

[9] Liu, Y.: The Building of Knowledge Database of Contemporary Chinese Functional Words. Postdoctoral Report. Peking University, Beijing(2004)(In Chinese)

[10] Peng, S.: The Building of Knowledge Base of Contemporary Chinese Prepositions and Related Research. Postdoctoral Report. Peking University, Beijing(2006)(In Chinese)

[11] Zan, H.Y., Zhang K.L., Chai, Y. M., Yu, S. W.: Studies on the Functional Word Knowledge Base of Modern Chinese. Journal of Chinese Information Processing. Vol.21 No.5:107-111(2007)(In Chinese)

[12] Zan, H.Y., Zhu, X.F.: NLP oriented studies on Chinese Functional Words and the Construction of Their Generalized Knowledge Base. Contemporary Linguistics, Vol.11 No.4:124-135(2009)(In Chinese)

[13] Zan, H.Y., Zhang，K.L.,Zhu, X.F., Yu, S.W: Research on the Chinese Function Word Usage Knowledge Base. International Journal on Asian Language Processing. Vol.21 No.4: 185-198(2011)(In Chinese)

[14] Wu, Y.P., Zan, H.Y.. Automatic Updates of Position Words Usage Rules in Modern Chinese Based on Error-driver. In: The 5th Youth Computational Linguistics Workshop, pp. 43-49. Wuhan (2010)(in Chinese)

[15] Han, Y.J., Zhang, K.L., Zan, H.Y, Chai, Y.M. :Automatic Discovery on Auxiliary Word Usage-based POS and Segmentation Errors for Chinese Language. Application Research of Computers. Vol.28 No.4:1318-1321(2011) (In Chinese)

[16] Yuan, Y.C., Zan, H.Y., Zhang, K.L., Zhou, Y.H.: The Automatic Annotation Algorithm Design and System Implementation Rule-base Function Word Usage. In: The 11th Chinese Lexical Semantics Workshop, pp. 163-169. Suzhou (2010) (in Chinese)

[17] Zan, H.Y., Zhou, L.J., Zhang, K.L.: Studies on the Automatic Recognition of Modern Chinese Conjunction Usages. Lecture Notes in Computer Science, Vol. 6838:472-479(2011)

[18] Zhou, Y.H., Mu, L.L., Zan, H.Y, Yuan Y C.: Research on Automatic Recognition of Chinese Modality Usage. Computer Engineering. Vol.36 No.23:155-157(2010)(in Chinese)

[19] Han, Y.J., Zan, H.Y, Zhang, K.L., Chai, Y.M. :Automatic Annotation of Auxiliary Words Usage in Rule-based Chinese Language. Computer Application. Vol.31 No.12:3271-3274 (2011)(in Chinese)

[20] Zan, H.Y., Zhang, J.H., Zhu, X.F., Yu, S.W.: The Studies on the Usages and Their Automatic Identification of Chinese Adverb JIU. Recent Advances of Chinese Lexical Semantics, pp.37- 43 (2010)
[21] Zan, H.Y., Zhang, J.H.: Studies on Automatic Recognition of Chinese Adverb CAI’s usages Based on Statistics. In: Proceedings of the 5th International Conference on Natural Language Processing and Knowledge Engineering: pp393-397(2009)
[22] Zhang, K.L., Zhao, D., Zan, H.Y., Cha,i Y.M.: Studies on Automatic Recognition of Modern Chinese Common Adverbs’ Usages. Journal of Chinese Information Processing. Vol.26 No.6:65-71(2012)(In Chinese)
[23] Zhang, K.L., Zan, H.Y., Han, Y.J., Zhang, T.F.: Studies on Automatic Recognition of Contemporary Chinese Common Preposition Usage. In: Proceedings of CLSW2012: pp219-229(2012)
[24] Zhang, J.J., Zan, H.Y.: Automatic Recognition Research on Chinese Adverb DOU’s Usages. Journal of Peking University(Natural Science). Vol.49 No.1:165-169(2013)(In Chinese)
[25] Zan, H.Y., Zhou, L.J., Zhang, K.L.: Modern Chinese Conjunction Phrase Recognition Based on Usage. Journal of Chinese Information Processing. Vol.26 No.6:72-78(2012)(In Chinese)