IJCLA VOL. 3, NO. 2, JUL-DEC 2012, PP. 63–76
RECEIVED 30/10/11 ACCEPTED 09/12/11 FINAL 08/06/12

Mapping Synsets in WordNet to Chinese

SHI WANG

Chinese Academy of Sciences, China

ABSTRACT

WordNet is a large lexical database which has important influence on many computational linguistics related applications, but unfortunately cannot be used in other languages except English. This paper presents an automatic method to map WordNet synsets to Chinese, and then generate an homogeneous Chinese WordNet. The proposed approach is grounded on the viewpoint that most cognitive concepts are languages independent, and can be mapped from one language to another unambiguously. Firstly, we utilize offline/online English-Chinese lexicons and term translation system to translate the words in WordNet. One English word is translated to multiple Chinese words, and one synsets is translated to a group of Chinese words. We secondly cluster these Chinese words into synonym-sets according to their senses. And finally, we select the right synonym-set for given synset. We regard the proper word-set choosing process as a classifier problem, and put forward 9 classifying features based on relations in WordNet, Chinese morphologies, and translation intersections. Besides, an lexico-syntactic patterns based heuristic rule is combined for higher recall. Experiment results on WordNet 3.0 show the overall synsets translating coverage of out method is 85.12% with the precision of 81.37%.

KEYWORDS: *WordNet translation, Chinese WordNet, lexical resources, computational linguistics*

1 INTRODUCTION

WordNet is a widely-used large-scale lexical database in which nouns, verbs, adjectives and adverbs are grouped into sets of cognitive concepts

(also called synsets) [1]. Synsets are interlinked by means of conceptual semantic relationships and then construct a net. Up to now, there are totally 155,287 words and 117,659 synsets in WordNet 3.0.

WordNet has been used in a large range of applications including natural language process, information retrieval, word sense disambiguation, text classification, image retrieval, etc. Unfortunately, this valuable resource cannot be directly used in other languages except English.

This paper introduces an automatic method for the construction of Chinese WordNet by mapping WordNet synsets to Chinese. The root of our work is that most synsets are languages independent and can be directly mapped to other languages unambiguously, though words in synsets may not be explicitly one by one translated. Most of synsets in WordNet, which express cognitive concepts in real world, can also be expressed by Chinese. If we map all synsets to Chinese, we obtain Chinese WordNet in which synsets are interlinked by identical semantic relations as in WordNet.

We firstly utilize offline/online English-Chinese lexicons and term translation system to translate the words in WordNet. One English word is translated to multiple Chinese words, and one synsets is translated to a group of Chinese words. We secondly cluster these Chinese words into synonym-sets according to their senses. And finally, we select the right synonym-set for given synset.

Regarding the proper word-set choosing process as a classifier problem, we put forward 9 classifying features based on relations in WordNet, Chinese morphologies, and translation intersections. Besides, an lexico-syntactic patterns based heuristic rule is combined for higher recall. Experiment results on WordNet 3.0 show the overall synsets translating coverage of out method is 85.12% with the precision of 81.37%. Experiment data and final results is available from `http://www.knowology.cn/cicling12/ChWordNet.rar` and `http://www.cicling.org/2012/data/33`.

The remainder of the paper is organized as follows. In section 2 we present related work. Section 3 described the proposed method in detail and section 4 gives its experimental results. Finally, we discuss shortcomings of our work and conclude this paper.

2 Related Work

The Global WordNet Association Association [2] provide a free, public and non-commercial organization that provides a platform for discussing,

sharing and connecting WordNets for all languages in the world. The Association held a conference every two years.

EruoWordNet has been built according to same structure with WordNet[3]. EuroWordNet is a multilingual database with WordNet for several European languages including Dutch, Italian, Spanish, German, French, Czech and Estonian, and are structured in the same way as the WordNet.

In Asia [4] shows an evaluation of the Korean WordNet. The purpose of their work is to study how well the manually created lexical taxonomy is built. Evaluation is done level by level, and the reason for selecting words for each level is that we want to compare each level and to find relations between them.

For Chinese, CiLin [5] and HowNet [6] are analogous but very different resources. CiLin has a four-layer semantic structure but does not provide clear relations between words. HowNet is an extra-linguistic knowledge base which unveils inter-concept relations and inter-attribute relations of the concepts. It uses sememes to explain all the concepts and relations in it, which is different from the relational analysis methodology adopted by WordNet. [7] and [8] integrated CiLin and HowNet with WordNet.

Because built manually requires great efforts, much work focused on automatical WordNet translation these years. [9] proposes a method to map Chinese words into WordNet by integrating five linguistic resources including English/Chinese sense-tagged corpora, English/Chinese thesauruses, and a bilingual dictionary. A Chinese WordNet and a Chinese-English WordNet are derived from the structures of WordNet.

[10] uses a statistics-based method that looks for the intersection of word sense to translate of synset of WordNet. [11] describes automatic techniques for mapping entries to WordNet senses.

[12] examines the validity of cross-lingual lexical semantic relations inferences by bootstrapping a Chinese WordNet. They claim that such correspondences must be based on lexical semantic relations, rather than top ontology or word translations.

3 METHOD

In brief, we firstly translate synsets into a group of Chinese synonym-sets based on word translations, and then select right one for given synset. Taking synset “tiger, Panthera tigris – (large feline of forests in most of Asia having a tawny coat with black stripes; endangered)” for instance, there are four steps for mapping it to Chinese:

1. Translating each word in synset to Chinese
   - tiger → 虎/tiger, 公虎/male tiger, 暴徒/mob, 凶徒/villain
   - Panthera tigris → 老虎/tiger, 虎/tiger
2. Clustering translations into synonym-sets according to their senses
   - tiger → {虎/tiger, 公虎/male tiger}, {暴徒/mob, 凶徒/villain}
   - Panthera tigris → {老虎/tiger, 虎/tiger}
3. Choosing right synonym-sets for synset
   - tiger → {虎/tiger, 公虎/male tiger}✓, {暴徒/mob, 凶徒/villain}×
   - Panthera tigris → {老虎/tiger, 虎/tiger}✓
4. Merging the right synonym-sets for synset as result
   - result = {虎/tiger, 公虎/male tiger, 老虎/tiger},

In step 3, symbols ✓/× represent whether the word-set was chosen or not. As a result, synset {tiger, Panthera tigris} is mapped to {虎/tiger, 公虎/male tiger, 老虎/tiger}. We note that semantic relationships which linked with it in WordNet are still unchanged. So if we can map all synsets to Chinese, we obtain an Chinese WordNet.

3.1 *Definitions*

Definition 1. *For a particular sense* ***ss*** *of an English word, its* ***sense translation*** $T_{ss}(ss) = \{cw_1, ..., cw_n\}$ *is a set of Chinese synonyms which express its meaning.*

Example. {虎/tiger, 公虎/male tiger} is a one sense translation for "tiger".

Definition 2. *For an English word* ***ew*** *with* $\mathtt{m}$ *senses* $\{ss_1, ..., ss_m\}$*, its* ***clustered word translations*** $T_{wd}(ew) = \{T_{ss}(ss_1), ..., T_{ss}(ss_m)\}$ *is the set of its senses translations.*

Example. T_{wd}(tiger) $= \left\{ \begin{matrix} \{\text{老虎/tiger, 虎/tiger}\} \\ \{\text{暴徒/mob, 凶徒/villain}\} \end{matrix} \right\}$

Definition 3. *Given an English synset* ***esy*** *of* $\mathtt{m}$ *words* $\{ew_1, ..., ew_m\}$*, its candidate translations* $CT_{esy}(esy) = T_{wd}(ew_1) \cup, ..., \cup T_{wd}(ew_m)$*, are called* ***synset candidate translations****. This the union of its words' translations.*

Example. $CT_{sy}\left(\left\{ \begin{matrix} \text{tiger,} \\ \text{Panthera} \\ \text{tigris} \end{matrix} \right\}\right) = \left\{ \begin{matrix} \{\text{老虎/tiger, 虎/tiger}\} \\ \{\text{暴徒/mob, 凶徒/villain}\} \\ \{\text{老虎/tiger, 虎/tiger}\} \end{matrix} \right\}$

3.2 *Getting Synsets Candidate Translations*

Words translating is the base of our whole approach. Besides common words, there are also lots of multi-word expressions in WordNet, including technical terms (“hydroflumethiazide”), fixed expressions (“by and large”), compound phrases (“car park”), verb-particle constructions (“look up”), and light verbs (“make a face”), etc., which are all difficult to translate using traditional dictionaries.

In order to translate as many words as possible, we utilize 8 resources which are complementary with each other as listed in Table 1.

Table 1. Word translating resources

ID	Resource	Translations clustered according to senses?
1	American Heritage	yes
2	Modern E-C	yes
3	Modern Comprehensive E-C	yes
4	Concise E-C	no
5	Landau E-C	common words: no; terms: yes
6	HaiCi Online[1]	no
7	Google Online[2]	yes
8	TermTrans [13]	yes

When translating words using these resources, we want to cluster translations into synonym-sets which will be used to form Chinese synsets as last. Table 1 also shows whether the resources’ translations have already been clustered or not. Accordingly, we devise words translating procedure.

- **Translating common words.** Given an English word, translating it using dictionaries which have already clustered their translations according to word senses, that is, resources 1, 2, 3, and 7. Clustering translations into word-sets as these dictionaries provide.
- **Translating rarely used words offline.** If not translated, translating using Concise E-C dictionary. Concise E-C has the largest size among all lexicons, and most rarely used words which are not disposed in step 1, such as “harpsichordist”, appear in it. According to Zipf law [14], these rarely used words often have unique sense. So

[1] http://dict.cn

[2] http://translate.google.cn

although translations of Concise E-C are not organized well, the one word translation for rarely used words are adoptable.

- **Translating multi-word expressions offline.** If not translated, translating only using the term translations of Landau E-C dictionary.
- **Translating rarely used words online.** If not translated, translating using HaiCi online dictionary which will automatically transform morphology of word. HaiCi can automatically transform morphologies of words and return related translations. For example, if we look up "antlered" which is not embodied in HaiCi, it will return the translation of "antler" and illuminate that "antlered" is the adjective morphology of "antler" meanwhile. This feature can highly improve the word translation coverage.
- **Translating multi-word expressions online.** If not translated, translating using TermTrans. TermTrans can dispose multi-word expressions. And because most multi-word expressions have unique translation, we only accept the best result TermTrans gives.

We ensure translations are separated according to senses by taking the one-word translation for Concise E-C dictionary, HaiCi online dictionary, and discarding translations for common words in Landau E-C dictionary.

Although resources we adopted are carefully selected, it is inevitable that there are still some words cannot be translated. In experiment section, we will give translation coverage rate in detail.

As shown in definition 3, synsets candidate translations is the union of their containing clustered words translations.

3.3 *Selecting Sense Translations for Synsets*

As presented above, each synset is translated to a group of synonym-sets in which some are right for the synset and others are not. In a special case that there is only one candidates synonym-set, there is no other choice besides accepting it. We call such synsets *clear synsets*. In our experiment, 26.06% synsets in WordNet are clear synsets. For the other synsets, we managed to select right sense translations.

We regarded the selecting procedure as a classifying problem. For a candidate synonym-set, we concluded a group of features to judge whether it is the proper one or not. The features are designed based on relations in WordNet, Chinese morphologies, and translation intersections. A binary classifier was trained using the features introduced below.

INNER-INTERSECTION FEATURE Words in a same synset are synonyms, so their proper translations should share common words. Taking synset "tiger, Panthera tigris" for example, the right sense translations for the two words have a common word "虎/tiger". So if two candidate sense translations have intersections, they are both likely to be the right ones.

We give the explicit measuring function for this feature as follows, which quantifies the shared words number of candidate sense translations in a same synset.

$$F_{II}(T_{ss}(ss_i)) = \\ |\{T_{ss}(ss_j) \in CT_{sy}(esy) | T_{ss}(ss_i) \cap T_{ss}(ss_j) \neq \emptyset|$$

OUTER-INTERSECTION FEATURES In WordNet, SIMILAR-TO (SIM for simplicity) is conceptual relationship which reflects two adjective synsets are similar. For example, "{absorbing, engrossing, fascinating, gripping, riveting}" is similar to "{interesting}".

Being similar is close to being synonymous. So, enlightened by the inner-intersection feature, we proposed outer-intersection feature based on the hypothesis that similar synsets would share common translations. To be specified, for a pair of synsets which satisfied SIM relations, if two candidate translation share some words, the two candidates are both likely to be right ones.

For other two relations SEE-ALSO (SEE) and VERB-GROUP (GRP), we can get analogical features. The three outer-intersections features are calculated as follows:

$$F_{\text{SIM}}(T_{ss}(ss_i)) = \\ |\{T_{ss}(ss_j) \in CT_{sy}(\text{SIM}(esy)) | T_{ss}(ss_i) \cap T_{ss}(ss_j) \neq \emptyset|$$

$$F_{\text{SEE}}(T_{ss}(ss_i)) = \\ |\{T_{ss}(ss_j) \in CT_{sy}(\text{SEE}(esy)) | T_{ss}(ss_i) \cap T_{ss}(ss_j) \neq \emptyset|$$

$$F_{\text{GRP}}(T_{ss}(ss_i)) = \\ |\{T_{ss}(ss_j) \in CT_{sy}(\text{GRP}(esy)) | T_{ss}(ss_i) \cap T_{ss}(ss_j) \neq \emptyset|$$

where {SIM|SEE|GRP}(*esy*) are the {SIM|SEE|GRP} linked synsets of *esy* in WordNet.

LEXICAL CONSTRUCTION FEATURES ATTRIBUTE is a relation between noun synsets and adjective synsets which express that the adjective synsets are attributes of noun synsets. For example, {"able"} is an attribute of {"ability"}.

In Chinese, the nouns plus auxiliary "的/of" is likely to be form its attribute adjectives. We use this word formation rule to judge synsets which are linked by ATTRIBUTE relations. Taking {"able"} and {"ability"} for example,

- CT_{esy}({able}) = {
  {能/able,可/able,会/able},
  {有能力的/capable,能干的/capable,有才能的/able} }
- CT_{esy}({ability}) = {
  {能力/ability,能耐/ability,才能/talent,本领/ability} }

We can easily determine that {有能力的/capable,能干的/capable,有才能的/able} is right for synset {able} because in Chinese, a noun added suffix "的/of" often constructs the corresponding attribute adjective. In the same manner, we can also propose four other lexical features based on HYPERNYM, SISTER, PART-OF and ANTONYM relations.

In Chinese, hypernyms are ofter suffixes of hyponyms (for example, "动物/animal" is hypernym and also suffix of "哺乳动物/mammal"), and then sisters are often share common suffixes ("哺乳动物/mammal" and "爬行动物/reptiles" are in sister synsets, and also share same suffix literally). Parts and wholes sometimes contain same prefixes ("屋顶/roof" is a part of "屋子/house", and they have same prefix), and antonyms can be obtained by simply adding special prefix like "反, 非, 不/aiti-, un-, no-" to words.

$$F_{\text{ATTR}}(T_{ss}(ss_i)) = |\{T_{ss}(ss_j) \in CT_{sy}(\text{ATTR}(esy))|f_a(T_{ss}(ss_i), T_{ss}(ss_j))|$$

$$F_{\text{HYP}}(T_{ss}(ss_i)) = |\{T_{ss}(ss_j) \in CT_{sy}(\text{HYP}(esy))|f_h(T_{ss}(ss_i), T_{ss}(ss_j))|$$

$$F_{\text{SIST}}(T_{ss}(ss_i)) = |\{T_{ss}(ss_j) \in CT_{sy}(\text{SIST}(esy))|f_s(T_{ss}(ss_i), T_{ss}(ss_j))|$$

$$F_{\text{PART}}(T_{ss}(ss_i)) = |\{T_{ss}(ss_j) \in CT_{sy}(\text{PART}(esy))|f_p(T_{ss}(ss_i), T_{ss}(ss_j))|$$

$$F_{\text{ANTI}}(T_{ss}(ss_i)) = |\{T_{ss}(ss_j) \in CT_{wd}(\text{ANTI}(ew))|f_t(T_{ss}(ss_i), T_{ss}(ss_j))|$$

where $f_{\{a,h,s,p,t\}}$ are boolean function described above. Detailed calculating formulas are omitted the sake of brevity.

The above 9 features can be calculated efficiently when classifying synsets candidate translations. In our experiments, we firstly use these features to train a classifier. For the candidates which can not classified, we turn around the following more time-consuming lexico-syntactic patterns rule.

LEXICO-SYNTACTIC PATTERNS FEATURES Lexico-syntactic patterns [15] have the ability to express semantic relationships between concepts, such as "X is a kind of Y" or "X such as Y". In WordNet, all the conceptual relations can be expressed by lexico-syntactic patterns. Then for the ambiguous synsets candidate translations, we can testify them by using such patterns.

For instance, for synset "tiger", $T_{ss}(ss_1)$={虎/tiger, 公虎/male tiger} and $T_{ss}(ss_2)${暴徒/mob, 凶徒/villain}, if we can obtain its hypernym synsets $\{bigcat, cat\}$ whose synset candidate translation is {猫/cat,猫科动物/$felid$}, then we can tell ss_1 is the required one by indexing sentences like "虎是一种猫科动物/tiger is a kind of felid" from corpus.

Using web search engines, we can quickly get the number of snippets which contain certain sentences. In our experiments, we use Google and then restrict our patterns to abide by Google query term expressions. Table 2 displays some of typical patterns we conclude, where c_1 stands for the words in the source synsets and c_2 represents the target synsets' words for a certain relation in WordNet. The double quotation marks that bracket the patterns can make Google search them as whole units, and the wildcards '*' can represent any single word.

For an synset, we firstly find its relative synsets. After filling each word in initial synset and related synsets to corresponding patterns according to their relationship, we feed the query string to Google and judge synset translation by return web pages number.

We did not use the hitting page numbers as features to train a classifier because it is very time costing to get all the numbers for all patterns. A

Table 2. Some lexico-syntactic patterns for synset disambiguation

ID	Relations	Patterns	Patterns in English
01	SYNSET HYPERNYM	c_1是一*c_2	c_1 is a * c_2
02		c_2等c_1	c_2 such as c_1
03	INSTANCE HYPERNYM	c_1属于c_2	c_1 belongs to c_2
04		c_2源自c_1	c_2 is derived from c_1
05	MEMBER-OF	c_1是c_2之一	c_1 is member of c_2
06		c_2中的c_1	c_1 in c_2
07	SUBSTANCE-OF	c_1是c_2的成分	c_1 is substance of c_2
08		c_2由c_1构成	c_2 is made of c_1
09	PART-OF	c_1是c_2的一部分	c_1 is a part of c_2
10		c_2由c_1组成	c_2 is composed of c_1
11	ATTRIBUTE	c_1是c_2的	c_1 is c_2
12		c_2的c_1	c_1 of c_2
13	CAUSE	c_1导致c_2	c_1 cause c_2
14		c_2是因为c_1	c_2 is caused by c_1

empirical method is adopted. That is, if the hitting page number exceeds an experiential threshold for a particular pattern, we accept the candidate translation. If we can query Google or some other huge corpus quickly, we can further use the hits number as features to train the classifier.

3.4 *Merging Selected Sense Translations*

Different dictionaries generate different translations for a same word. For example, for word "tiger", Concise E-C dictionary translates its one sense to {虎/tiger, 公虎/male tiger}, while Modern Comprehensive E-C dictionary gives {老虎/tiger, 虎/tiger}.

So, multi sense translations will be accepted in the candidates choosing procedure. We merged these synonym-sets to generate a compact and integrative translation because they are actually represents same meaning. After merging, we get the right translations for synsets.

In word translating procedure, we have ensured each word-set are synonyms. Being synonymous is transitive for words. So if we merge the word-sets which share common words, the new formed word-set is also a synonym-set.

Our merging strategy is very strict. Another common used method is based on edit distance, that is, merging word-set which have short edit-distances. In experiment, such a relax strategy performs bad. Most Chinese words are very short and might be very different in sense even

they are very similar in morphology. For example, "老虎/tiger" and "老师/teacher" have short edit distance 1, but are completely different in meaning.

4 EXPERIMENT

4.1 *Word Translation Results*

Table 3 shows the word translation percentage for all resources listed in Table 1.

Table 3. Word translation coverage of all the resources

ID	Resource	Coverage
1	American Heritage Dictionary	35.52%
2	Modern E-C dictionary	32.40%
3	Modern comprehensive E-C dictionary	25.81%
4	Concise E-C dictionary	19.75%
5	Landau E-C dictionary	20.14%
6	HaiCi online dictionary	9.55%
7	Google online dictionary	38.72%
8	TermTrans Tool	6.10%
	Average	84.33%

From Tabel 3, we can see that although every distinct resource's coverage is low, the total coverage can reach 84.33%. That means our resources are complementary with each other. And excluding TermTrans, all the other dictionaries are manually compiled and with very high precision.

Errors are mainly caused by the mixing of translations with different senses. For example, in Modern E-C dictionary, word "forefront" are translated to be "最前面/the part in the front or nearest the viewer,最前线/the position of greatest importance or advancement", but these two words are distinguished in WordNet. Table 3 also demonstrates that merging sense translations does not generate too much errors.

4.2 *Synset Candidate Translations Classifying Results*

For different kinds of synsets (noun, verb, adjective and adverb ones), they can utilize different features. Inter-intersection features for VERB-ALSO relations are not available for Noun synsets, for example. So when

constructing trainset, in order to make sure that each feature can be used, we randomly select 200 positive and 200 negative samples which have valid feature value for each feature. There are 1,500 positive and 1,500 negative samples at all, making up about 0.18% for all sense translations.

We adopted NativeBayes, J48, and AdaboostM1 to train the classifier. The labels are 1/0 and results are verified with 10 cross-validation. The performance for all kinds of synsets are shown in Table 4.

Table 4. Result of classifier

Synset	Label	NaiveBayes			J48			AdaboostM1		
		p	r	F1	p	r	F1	p	r	F1
Noun	1	0.921	0.729	0.814	0.85	0.904	**0.876**	0.863	0.866	0.8
	0	0.657	0.893	0.757	0.816	0.726	**0.768**	0.768	0.764	0.766
Verb	1	0.854	0.818	**0.836**	0.873	0.758	0.812	0.854	0.78	0.816
	0	0.861	0.889	**0.875**	0.827	0.912	0.867	0.837	0.894	0.865
Adj	1	0.883	0.852	0.867	0.858	0.887	**0.872**	0.878	0.856	0.867
	0	0.811	0.849	**0.829**	0.837	0.798	0.817	0.813	0.84	0.827
Adv	1	0.904	0.853	**0.878**	0.824	0.891	0.856	0.892	0.853	0.872
	0	0.801	0.868	**0.833**	0.819	0.721	0.767	0.798	0.85	0.823

From Table 4, we can see performances of the three classifier are similar. This demonstrates the features are well selected. NaiveBayes performance better in verb, adjective, and adverb synsets, while J48 work well for noun synsets. Accordingly, we use J48 to disambiguate noun synsets, and take NaiveBayese for the other ones. Table 4 give their results.

Table 5. Performance of classifier

	Noun	Verb	Adj	Adv	Average
Precision	82.14%	78.35%	81.22%	81.49%	81.37%
Coverage	86.71%	80.16%	83.91%	82.35%	85.21%
Average Words Number	4.13	6.25	6.00	3.01	4.62

4.3 *Lexicon-Syntactic Patterns Results*

Lexicon-syntactic patterns based disambiguation is time consuming. We did not take it as a classifier feature, but used as an heuristic rule. If one

pattern hits enough web pages, the candidate are accepted. Performance of this way is given in Table 6 with the former two ways.

Table 6. Performance of lexical patterns

	Clear synsetsx	Classifier	Lexical patterns
Precision	99.10%	81.37%	91.34%
Coverage	26.06%	47.21%	18.68%

5 Conclusion and Future Work

WordNet is an important resource for many applications but restricted to English, so translating it to Chinese is valuable. Our work is ground on the argument that concepts can be translated from one language to another expressed by synsets. The two major problems for the work are to translate English words and to choose the right translation for synsets. We firstly translate all the words in WordNet using three kinds of complementary resources, and then disambiguate the translation of synsets using a classifying combined with heuristic rules. Experiments show that our method can translate 85.12% of the synset in WordNet 3.0 with a precision of 81.37%.

Our future work will concentrate on how to improve the translate coverage of words, especially the multi-word expressions, in WordNet.

Acknowledgements This work was supported by the National Natural Science Foundation of China, under grants No. 61203284, 60573063, 60573064, 60773059, 61035004, the National High Technology Research and Development Program (863 Program) of China under No. 2007AA01 Z325, and National Social Science Foundation of China under grant No. 10AYY003.

References

1. Miller, G.A.: WordNet: A lexical database for English. Commun. ACM **38** (1995) 39–41
2. Global WordNet Association. `http://www.globalwordnet.org` (2000)

3. Piek, V.: EuroWordNet: A multilingual database with lexical semantic networks. Dordrecht: Kluwer Academic Publishers (1998)
4. Altangere, C., Ho-Seop, C., Cheol-Young, O., Hwa-Mook, Y.: On the evaluation of Korean WordNet. In: TSD 2007. (2007) 123–130
5. Mei, J., Zhu, Y., Gao, Y., , Yin, H.: TongYiCiCiLin. Shanghai Dictionary Press (1982)
6. Dong, Z., Dong, Q. `http://http://www.keenage.com` (2000)
7. Chen, H.H., Lin, C.C., Lin, W.C.: Construction of a Chinese-English WordNet and its application to CLIR. In: Proceedings of the Fifth International Workshop on Infor-mation Retrieval with Asian Languages. (2000)
8. Dorr, B.J., Levow, G.A., Lin, D.: Building a Chinese-English mapping between verb concepts for multilingual applications. In: Proceedings of 4th Conference of the Association for Machine Translation. (2000)
9. Chen, H.H., Lin, C.C., Wen, C.L.: Building a Chinese-English WordNet for translingual applications. ACM Transactions on Asian Language Information Processing **1**(2) (2002) 103–122
10. Liu, M.: A research on translating WordNet nodes to Chinese. Master's thesis, DongBei University (2003)
11. Green, R., Pearl, L., Dorr, B.J., Resnik, P.: Mapping lexical entries in a verbs database to WordNet senses. In: ACL 2001. (2001) 244–251
12. Huang, C.R., Tseng, I.J.E., Tsai, D.B.S.: Translating lexical semantic relations: The first step towards multilingual WordNets. In: COLONG 2002. (2002)
13. Fang, G., Yu, H., Nishino, F.: Chinese-english term translation mining based on semantic prediction. In: ACL 2006. (2006)
14. Manning, C.D., Schütze, H.: Foundations of Statistical Natural Language Processing. MIT Press (1999)
15. Hearst, M.A.: Automatic acquisition of hyponyms from large text corpora. In: COLING 1992. (1992) 539–545

SHI WANG
KEY LABORATORY OF INTELLIGENT INFORMATION PROCESSING,
INSTITUTE OF COMPUTING TECHNOLOGY,
CHINESE ACADEMY OF SCIENCES,
BEIJING, 100190, CHINA
E-MAIL: <WANGSHI@ICT.AC.CNX>