International Journal of Knowledge and Language Processing
Volume 3, Number 4, **2012**

www.ijklp.org
KLP International ©2012 ISSN 2191-2734
pp.18–32

The Corpus Construction and Parsing Technology Based on Chinese Semantic Dependency

Shao Yanqiu, Liang Chunxia, Mao Ning
Natural Language Processing Group, Research Department
Beijing City University
Beijing, 100083, China
{yqshao, nancylcx, maoning }@bcu.edu.cn

Received July 2012; revised October 2012

ABSTRACT. *Semantic dependency analysis is one of the deep semantic analysis methods. Its semantic structure is simple and clear. We propose a set of semantic dependency annotation scheme including the definition of a new semantic relation set, the construction of a large-scale semantic dependency corpus, and the semantic dependency analysis method. Our system not only labels the semantic relations between the predicate and its arguments, but also labels the semantic relation of the word pairs in the phrase, such as noun phrase. We define reverse relation and indirect relation so as to treat the situation that a verb acts as a modifier and a verbal noun acts as the head of the noun phrase. According to the relation system, a large scale Chinese semantic dependency relation Treebank is constructed by the combination of automatic and manual methods. A maximum entropy model is built to assign the semantic dependency relations to the head-dependent pairs. The result shows that ME model performs much better than MST parser.*

Keywords: semantic dependency; semantic relation; semantic dependency Treebank; semantic analysis; semantic dependency parsing

1. **Introduction.** Language generally has three important layers- sound, form and meaning[1]. Among these three layers, meaning is the most important layer[2]. For natural language processing, semantic analysis is the key technology to understand the meaning of the sentence, and it could not be replaced by syntactic parsing. Sometimes, maybe the syntactic constituents of the two sentences are same, but the corresponding semantic relations are different. For example, these two Chinese sentences, “他写完了”(*He has finished.*) and “文章写完了”(*The paper has been finished.*), in Chinese, the results of the syntactic parsing of them are same. Both of them could be represented by the same Chinese constituent structure of “*NP-VP*”. But the semantic relations between these two syntactic constituents are quite different. The former NP is “他”(*He*) which is the agent of the action “写完” (*finish*) and the latter NP is “文章”(*paper*) which is the patient of action “写完”

(*finish*) . It is obvious that traditional syntactic analysis is not enough to understand the sentence meaning.

Compared with syntactic analysis, semantic analysis could reach the further meaning essence of a sentence through the variable syntactic expression. Especially for Chinese, which is a kind of meaning-combined language, there is no inflection and the syntax is very flexible. However, the deep semantic expression is stable. For example, all these four Chinese sentences "我把香蕉吃了" (*I ate the banana*),"香蕉我吃了" (*The banana, I ate*), "香蕉被我吃了" (T*he banana was eaten by me*), and "我吃了香蕉" (*I ate the banana*), although they are expressed in different syntactic forms, they have the same meaning of "*I ate the banana*". They could be represented by a unified semantic form: "吃 (我, 香蕉), namely, "*eat* (*I, banana*) ", where "我" (I) is the agent of the action "吃" (*eat*) and "香蕉"(*banana*) is the patient of the action "吃". It could be seen that semantic analysis is the key point to understand the real meaning of the sentence.

Currently, the sentence level semantic analysis mainly focuses on some shallow semantic parsing tasks such as semantic role labeling (SRL). As a transition of semantic analysis, SRL plays an important role. But it has some limitations[3]. This kind of shallow parsing only analyzes the relationship between the main verb and its arguments, and it doesn't analyze the intrinsic semantic relation of the argument, such as the relation between a noun and its modifier. Besides, the definitions of semantic relation of SRL are not rich. For example, only 6 tags, Arg0 to Arg5 are used to represent the relationship between the verb and its arguments. For different verbs or same verbs but belong to different framesets, the same *Arg* maybe means quite different meaning. Thus, there must be many ambiguities on the semantic expression of the identical sign. Semantic dependency analysis (SDA) could help avoid some problems of SRL. SDA is a kind of deep semantic analysis technology.

2. **Related Work.** SDA or Semantic dependency parsing (SDP) is proposed to solve some of the above problems. From the definition of semantic dependency parsing, it could be known why this kind of method could do it.

2.1. **The Concept of Semantic Dependency Parsing.** The theoretical foundation of SDP is dependency grammar[4]. SDP integrates dependency structure and semantic information, and describes the sentence structure and semantic relation clearly and deeply. Different from SRL which only deals with the relations between the predicate and related arguments, SDP considers all the relations among dependent words (or modifiers) and their corresponding head words, and every word has and only has one father node, namely, the head word in the sentence. SDP covers not only the relations around the main predicate, but also more other auxiliary semantic relations, such as quantity, attribute, frequency, etc. For example, in the phrase "*beautiful girl* " there is an *attribute* relation between "*beautiful*" and "*girls*". Another example, the semantic relation of the phrase "木头椅子" (*wooden chair*) is "*material*" . However, these kinds of relations are not tagged in SRL.

Fig1. is an example of one Chinese sentence which is parsed by semantic dependency. It can be seen that each word has one and only one father, namely, head word, and the arrow of the arc points from the head word to the dependent word. For example, the head word

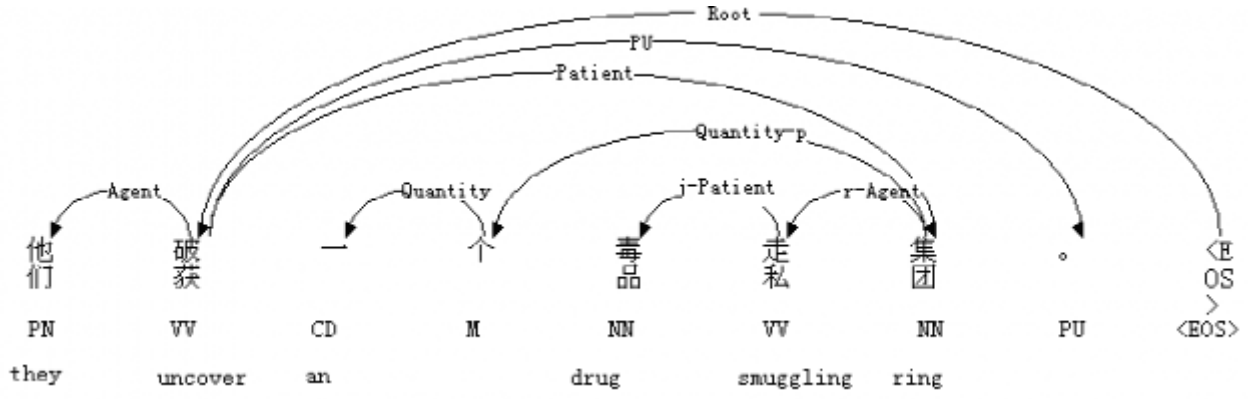

FIGURE 1. An example of Chinese Semantic Dependency Parsing.
("They uncovered a drug smuggling ring.")

"破获" (*uncover*) is the father of the dependent word "他们" (*they*), and the arrow points from "破获" to "他们". It also can be seen that the whole sentence has only one kernel word whose father is not the word in the sentence, e.g., the predicate "破获". Here, a special father tag "*EOS*" is assigned to it and the semantic relation is labeled as "*Root*".

As can be seen from the example, SDP not only analyzes the semantic role of the predicate, but also analyzes the internal structures of noun phrases which are not annotated in SRL. For example, for the noun phrase "一个毒品走私集团" (*a drug smuggling ring*), the semantic relation between word "一个" (a) and "集团" is "*quantity-p*" , and the relation between "走私" and "集团" is "*r-Agent*" . In SRL, the whole phrase "一个毒品走私集团" is labeled as " *Arg1*" which means "*patient*", and the system will not analyzed each word pair in detail.

From above, it can be seen SDP presents complete semantic information of a sentence. It is a real deep semantic analysis. The definition of semantic dependency relation system, the construction of semantic dependency Treebank and the analysis technology are the main contents of this paper.

2.2. **State-of-the-Art Development.** Semantic analysis includes the semantic analysis theory, the definition of semantic relations, the construction of corpus and the parsing model. As for semantic analysis theory, besides semantic parsing, there are some other theories, such as argument structure[5], semantic role labeling[6], case grammar[7] and so on.

About the definition of Chinese semantic relations, different linguists propose different classification standards. Yuan Yulin presents 40 relation tags including thematic role tag set, logic relation tag set and discourse relation tag set[5]. Feng Zhiwei researched on the argument structures of Chinese verbs, adjective, and some nouns from 1970s to early 1980s, and his tag set includes 30 argument relations[8]. Lu Chuan's Paratactic network include 6 classes and 26 relations. Lin Xingguang points out 22 basic cases[9]. Dong Zhendong classified 83 categories of semantic relations from the events in his HowNet, and the categories are divided into main semantic roles and auxiliary semantic roles [10].

About resource construction, there is no large scale corpus for semantic dependency parsing published to the public. There are two kinds of corpus: syntactic dependency corpora and semantic role labeling corpora. Penn Treebank[11]is the more popular English constituent structure syntactic Treebank, which has a high level of consistency and tagging

accuracy, and has become the acknowledged training and testing set for the current research on English syntactic parsing. As regards Chinese, the famous corpus are Sinica Treebank (in traditional Chinese character) developed by Academia Sinica, Penn Chinese Treebank from Pennsylvania University, TCT (Tsinghua Chinese Treebank, Dang Zhengfa and Zhou Qiang transfered TCT to dependency structures by head node mapping list and the rules of the types of dependency relations[12]), and the dependency Treebank built by Research Center of Information Retrieval, Harbin Institute of Technology[13]. PropBank (Proposition Bank) is a semantic role labeling corpora based on Penn Treebank developed by Pennsylvania University[14]. PropBank only tags the predicates and only includes 20 relation roles. There are 6 core roles, and the same core roles may have different meaning for different predicate verbs. They also developed Chinese PropBank.

There are no very practical algorithms designed for semantic dependency analysis and the most relevant algorithms are those methods based on syntactic dependency analysis and corresponding SRL algorithms. The classical methods are graph-based and transformation-based algorithms, and some researchers tried to combine both of the two methods[15],[16],[17]. There are many deep research on SRL, whether feature selection or machine learning algorithm such as SVM, maximum entropy, decision tree, kernel-based methods, etc. Some researchers also regarded some lexical sense as the deep features to label the semantic roles[18]. For labeling model, both local model and global model are studied and applied[19]. All of these studies are helpful to the model building of semantic dependency analysis.

3. **Semantic Dependency Relationship System.** Different semantics theory defines different semantic relation set. According to different reference theory, a new semantic relation set is built in this paper.

3.1. **Semantic Relation Set.** By comparing different semantic relation systems, the relation tag set of HowNet, a famous Chinese semantic knowledge thesaurus, is selected to be one of the main reference set. Considering that the granularity of some relation in HowNet is too small, the combination job is done. At the same time, the modifier relation and syntactic tag are not very rich in HowNet, some enlargements are made. Besides HowNet, both of the semantic relation system of LuChuan and YuanYulin are considered in this paper. The relations of HowNet are extended and combined and a new semantic relationship system is constructed.

There are two main kinds of newly-built semantic relations in this paper, reverse relation and indirect relation, which aim at the situations of verbs acted as modifier and verbal noun acted as the phrase head word respectively. In a noun phrase, when a verb modifies a noun, the relation between the head and the dependent word is assigned a reverse relation. For instance, these two Chinese noun phrases “走私集团” (*smuggling ring*) and “开车的女士” (*a woman who is driving a car*), here, both “走私”(*smuggle*) and “开车” (*drive*) are verbs and modifiers of a noun. However, in this situation , if the relation is only labeled as “modifier relation”, the real semantic relation of “*agent*” of head-dependent word pair “集团-走私” (*ring-smuggle*) and ”女士-开车” (*woman-drive*) will be lost. Because if the phrase is converted to a sentence, e.g., “集团走私” (*The ring is smuggling.*) and “女士在开

车” (*The woman is driving the car.*), the verbs “*smuggle*” and “*drive*” will become the head words of the sentence. But in phrase level, the verbs are modifier. Therefore, the relation of verb modifiers and head word is labeled as reverse relation r-XXX, e.g, “*r-Agent*”. Fig.2 gives an example of reverse relation and non-reverse relation.

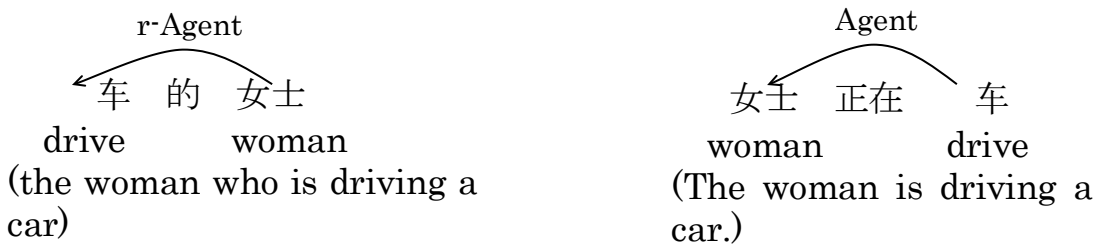

FIGURE 2. The comparison of reverse relation with non-reverse relation.

As for indirect relation, it means when a verbal noun is the head word of a phrase, the semantic relation is labeled as an indirect relation j-XXX, e.g, j-Patient. For instance, the phrase “对他的支持” (*the support for him*) , here, “支持” (*support*) is a verbal noun and it is the head word in this phrase, then, indirect relation “*j-patient*” is labeled between head word “支持” and dependent word “他” (*him*) [20].

Besides the above two newly-defined relations, some HowNet relations are modified, combined and deleted because of the low occurrence frequency. For example, “*DurationBeforeEvent*” and “*DurationAfterEvent*” are combined to “*Duration*” for the less occurrence, and some new tags which have the syntactic function, such as syntactic subjunctive relation, e.g., “如果…那么”(*if ...then*), are defined, “*s-condition*”.

On the whole, in our semantic dependency relation system, there are 20 sentence-level main roles including subject roles like “*agent*”, “*experiencer*”, etc., and object roles like “*patient*”, “*product*”, etc. There are 43 auxiliary roles such as “*space*”, “*time*”, “*maner*”, etc. Phrase-level roles include 18 direct modifier roles, reverse relations, and indirect relations. Besides these roles, 16 syntactic roles are contained in the system such as “*concession*”, “*condition*”, “*purpose*”, etc. Totally, 101 roles are defined and all of them are listed in Table 1.

The definition of each semantic relation in this paper is described as follows:

(a) TAG：agent|施事
DEF： The agent of self-acting.
Dep_S：她ß婉拒（她婉拒了他的约请）|Sheßrefuse (She refused his invitation.)

(b) TAG：possessor|领有者
DEF：The subject of possession relationship or the possessor of some entity
Dep_S：他ß有（他在北京有两处住房）| Heßhas (He has two apartments in Beijing.)

TABLE 1. The List of Semantic Dependency Relations

Sentence-Level Main Roles	
Subject Roles	agent，experiencer，possessor，existent，whole，relevant
Object Roles	isa, product, content, possession，target, patient，OfPart，contrast, partner，basis，cause，cost，scope，concerning
Sentence-Level Auxiliary Roles	
Time Roles	duration，TimeFin，TimeIni，time，TimeAdv
Location and State Roles	LocationFin，LocationIni，LocationThru，StateFin，StateIni，state，SourceWhole，direction，distance，location
Manner and Result Roles	accompaniment，succeeding，degree，frequency，instrument，material，means，method，angle，times，sequence-p，sequence，negation，all，modal，emphasis，manner，aspect，comment，ResultCom，StateCom，DirectionCom，CanCom，ResultContent，ResultEvent，ResultIsa，ResultWhole，result
Phrase-Level Roles	
Direct Modifier	d-agent, d-material, d-category，d-member，d-content，d-domain，d-quantity，d-quantity-p，d-deno，d-deno-p，d-sequence，d-sequence-p，d-host，d-TimePhrase，d-LocPhrase，d-InstPhrase,d-attribute，d-restrictive
Verb as Modifier	r-{Sentence-Level Main Semantic Roles} e.g. r-agent，r-patient， r-possessor.....
Verbal Noun as Head Word	j-{ Sentence-Level Main Semantic Roles} e.g. j-agent，j-patient，j-target……
Syntactic Roles and Others	
Syntactic Roles	s-cause，s-concession，s-condition，s-coordinate，s-or，s-progression， s-besides，s-succession，s-purpose，s-measure，s-abandonment，s-preference，s-summary，s-recount，s-concerning，s-result
Others	aux-depend，prep-depend，PU，ROOT

In the examples above, "*TAG*" represents the expression of Chinese and English semantic relation. "*DEF*" means the definition of semantic relation. "*Dep_S*" shows an example of this semantic relation. There is an arrow in every example. The arrow starts from a head node and ends at a dependent node and it could be expressed as "*dependent word ß head word*" or "*head word -> dependent word*".

3.2. **The Difference of Semantic Structure Between HowNet and Our System.** From the above definition, it could be seen that the structure of semantic dependency relation system is very simple. Every node has only one father and one sentence only has one kernel node. This kind of structure could be expressed by a two dimension table like Table 2. In fact, this kind of table could be easily converted into a tree. The advantages of this conversion are that many algorithms relevant to trees could be used.

TABLE 2. An example of the expression of semantic dependency relation of a Chinese sentence (See the sentence in Figure 1)

Word Numbr	Word	Corresponding English word	Father	Role
01	他们	they	02	Agent
02	破获	uncover	09	Root
03	一	an	04	Quantity
04	个		07	Quantity-p
05	毒品	drug	06	j-Patient
06	走私	smuggle	07	r-Agent
07	集团	ring	02	Patient
08	。		02	PU
09	<EOS>			

The main idea of semantic structure in HowNet is that Chinese sentence is not a tree, but graph. HowNet combines the tree with graph to express semantic information. It embodies this idea in that maybe there are multi kernel nodes in one sentence, but every labeling unit in one identical level has only one father and every labeling unit in different level has other fathers. That means the same unit could have several roles because of the different identity in different levels in the same sentence. This is shown in Table 3. From Table 3, it can be seen that the sentence has two kernel nodes, “破获” (*uncover*) and ”走私” (*smuggle*).

TABLE 3. An example of the HowNet semantic expression

Word Number	Word	Corresponding English word	Father	Child	Role 1	Role2	Description
01	他们	they	02		Agent		
02	破获	uncover		01,07			kernel
03	一	an	07				
04	个						
05	毒品	drug	06		Patient		
06	走私	smuggle	07-2	05			secondary kernel
07	集团	ring	02		Patient		
07-2	集团	ring		06		Agent	

The graph form of HowNet enriches the semantic structure expression. However, this

form has high complexity and the tree form is simpler. The problem of the secondary kernel in HowNet could be solved by the reverse relation which is defined in 3.1. For example, “走私” (*smuggle*) is labeled as a secondary kernel, the relation between “走私” and “集团” (*ring*) is “*Agent*” . This kind of tagging takes “走私” as the kernel in the second level. But in fact, the head word of the phrase “走私集团” is “集团” not “走私”. Different from HowNet, semantic dependency expression take “集团” as the head word, and this kind of expression is accord with the real conditions. In order to differentiate the agent relation in the situation of noun head from that of verb head, reverse relation is introduced. The introduction of reverse relation makes the semantic structure become more concise, clear and structural.

4. **Building Semantic Dependency TreeBank.** Two methods are used to construct the Tree Bank. One is to transform the existing syntax or semantic role labeling corpus, and the other is to tag the new corpus manually. In the process of manual annotation, in order to improve the tagging efficiency, active learning method is applied to help labeling corpus.

4.1. **Data Selection.** 10,400 sentences were selected from the Chinese Prop Bank as the raw corpus from which to create the Chinese Semantic Dependency Parsing corpus. These sentences were chosen for the annotation for three reasons. First, gold syntactic dependency structures can be of great help in semantic dependency annotation, as syntactic dependency arcs are often consistent with semantic ones. Second, the semantic role labels in PropBank can be very useful in the present annotation work. Third, the gold word segmentation and Part-Of-Speech can be used as the annotation input in this work.

4.2. **The Automatic Transformation of Original Corpus.** Before the manual annotation, some automatic transformation is done to reduce the workload.

4.2.1. **To Use the Function Tag of Penn Chinese Treebank** (PCT). Penn Chinese Tree Bank is a constituent structure syntactic TreeBank. PCT is one of our source corpora. Head node finding rules are applied to transform constituent structure to dependency syntactic structure. To reduce the workload, the functional tags of constituent structures are used as references. By writing rules, some parts of semantic relations are tagged automatically. For example, functional tags “*SBJ, OBJ, TMP*” in PCT represent “*Subject, Object, Time*” respectively, and many functional tags suffixed to prepositional phrase “*PP*” such as “*LOC, DIR, MNR*”, represent “*Location, Direction, manner*”. All these tags are useful to help label the semantic relations.

4.2.2. **To Label Semantic Dependency Relation According to Chinese PropBank (CPB).** CPB is constructed by adding a layer of semantic role information to PCT syntactic components. *Arg0-5* are used to represent the core roles. The real meanings of these tags are given in frame work files in CPB. So, the semantic dependency relations of those predicates and their arguments could be built according to semantic roles frameworks in PropBank. The semantic dependency relations are unified and concrete. For example, the roles *Arg0-4* of verb “缩短”(*shorten*) represent “*agent, theme, range, starting point, ending point*” respectively, and they would be transformed to semantic relations of “*Agent，patient，result，StateIni，StateFin*” by our rules.

4.3. **Manual Annotation.** Manual annotation is one of the most important processes to build the corpus.

4.3.1. **Manual Labeling By Using Tagging Tool.** A visual tagging tool is designed to help label corpus conveniently. There are several functions of the tool such as tagging and correcting dependency arc, dependency relation, word segmentation and tagging part-of-speech, finding the same or similar arc relation of current arc, showing semantic dependency framework of verb, and so on. Figure 3 is the interface of the tagging tool.

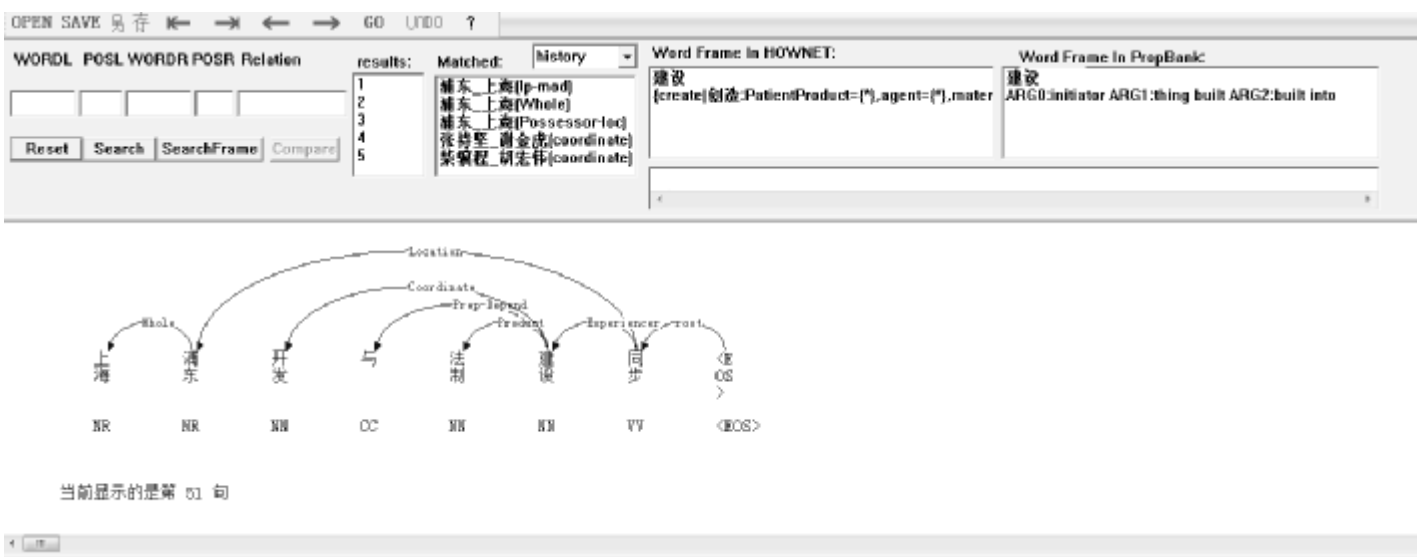

FIGURE 3. Semantic Dependency Relation Tagging Tool.

4.3.2. **Consistency Check.** Facing the same word pair, different annotators may have different tagging results. So, consistency check is necessary. The check includes:

(1) The Check of Complete Match. If two word pairs have the same words, the same arc and the same arc direction, they may have the same semantic relationships. If the relations are different, maybe one of their tagging is wrong.

(2) The Check of Semantic Relation Mapping Set. All of those head-dependent pair words which have the same semantic relation would be checked. Those pair words that do not belong to the semantic relation should be corrected. For example, for semantic relation " *ContentProduct*", all of the parent nodes of the pair words are collected and the corresponding semantic relation mapping set is {制定，题写，发表，建立……}(*establish, write, publish, build, etc.*). Because each verb has one or several semantic frameworks, it could be judged that if the framework of the verb in the set belongs to the semantic relation "*ContentProduct*".

(3) The Check of Pattern Matching. For those words that have the same pattern, e.g. "稳定性" (*stability*) and "综合性" (*comprehensiveness*), which should have the same relation when they act as a modifier. This check could find some same kind of errors.

4.3.3. **Automatic Assistant Tagging.** Tagging the relation of the arc is one of the main work in building the corpus. The maximum entropy model is used to help automatically label the relation. The features which are selected to train the model include word and POS of child node and parent node, the direction of the arc, the distance between child and parent, POSs of left and right word of parent, semantic dependency framework of parent

node and so on.

In the process of building the corpus, 1000 sentences are labeled manually first, and then based on these 1000 sentences, a maximum entropy model is trained on the basis of the above features. The later labeling work is done based on the model and at the same time, the manual correction work is still needed. The model is improved with the increasing of the training data. Because there is an initiative labeling at first, the process of annotation is simplified. The efficiency of labeling is improved.

5. **Maximum Entropy Based Semantic Dependency Relation Assignment.** Based on the above corpus, a maximum entropy based semantic dependency analysis model is built to label the semantic relation automatically.

5.1. **The Framework of Semantic Dependency Analysis System.** The main task of semantic dependency analysis is to assign a semantic relation tag to a head-dependent pair. Thus, SDA could be processed as a problem of classification. Of course, the classification should be done on the basis of a series of preparation including word segmentation, POS tagging, syntactic analysis. Here, one of the most important tasks of syntactic analysis is dependency structure tagging. The purpose of dependency structure tagging is to assign an only father of each word, namely, to decide head-dependent pairs or to draw the arcs of the pairs. The corresponding model is head word assignment model (HWA) in Figure 4. The next important model is semantic dependency relation assignment model (SDRA).

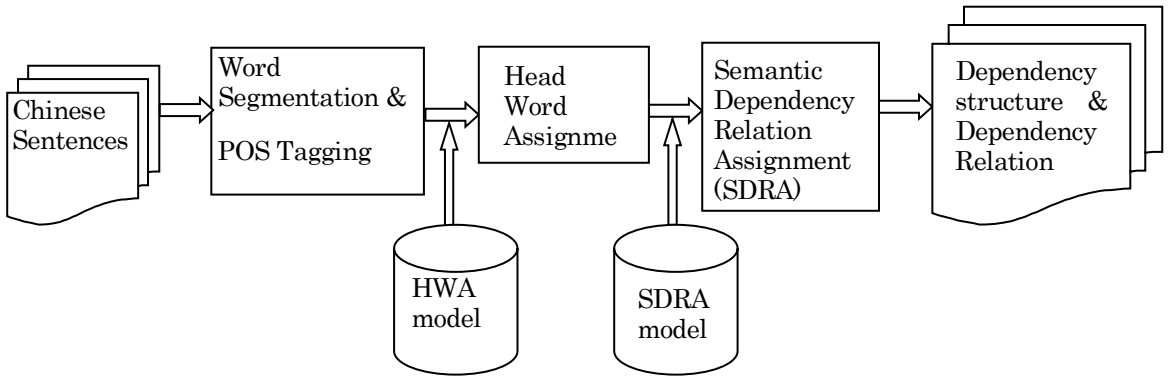

FIGURE 4. Semantic Dependency Analysis System Frame

In the process of corpus construction, the head word assignment is implemented on the basis of a rule set, which is similar to those used by the syntactic constituent to find the head of a phrase. While in the process of system test, the head word assignment model is based on the syntactic dependency model of HIT-SCIR. The UAS of this syntactic system is about 80.6%. According to the result of the head word assignment, the model of maximum entropy based semantic dependency relation assignment (SDRA) is built.

5.2. **Feature Extraction Based on Maximum Entropy Model.** The main idea of Maximum Entropy (ME) model is to build model for all of the known factors and exclude all of the unknown factors. That means the model should be such kind of probability distribution that could satisfy all the known facts and will not be influenced by unknown

facts. The most significant advantage of it is that it does not be limited to the condition of feature independence. Therefore, any features that could bring benefits to the classification could be added to the system and the interaction of each other is not considered. Compared with other classifier, ME model is easier to be applied to the problem of multi-classification and it could output a probability value which could be used to the succeeding inference. Considering that the semantic relation assignment is a typical multi-classification problem, ME classifier is selected to be used in this paper.

Many factors could be considered in the process of the semantic relation prediction. Suppose X is a vector of these factors, and the variable y is a semantic relation. p(y|X) refers to the probability of the prediction of whether a head-dependent pair is one of the semantic relations or not. ME model requires that the entropy H(p) in equation1 must be the maximum when p(y|X) is under certain constrains.

$$H(p) = -\sum_{X,y} p(y \mid X) \log p(y \mid X) \tag{1}$$

In the process of the training and testing ME model, every head-dependent pair is extracted 15 features which are listed in Table 3. Here, take the sentence of Figure1 for example. Suppose the current head-dependent pair is 破获-集团 (uncover-ring). The feature values are shown as follows.

TABLE 4. Features List

Feature	Feature description	Feature value
dw	dependent word	集团 (ring)
dw-POS	the POS of dependent word	NN
hw	head word	破获 (uncover)
hw-POS	the POS of head word	VV
dir	the direction of the arc, if the arc is from right to left, then dir is L, otherwise, dir is R	R
dis	the distance between the head word and dependent word, the value is the number of words between hw and dw	5
dw-lw	the left neighbour of dependent word, if there is no word on the left of dw, the value is Null. Analogously, the following features are the same.	走私 (smuggle)
dw-lp	the POS of dw-lw	VV
dw-rw	the right neighbour of dependent word	。
dw-rp	the POS of dw-rw	PU
hw-lw	the left neighbour of head word	他们 (they)
hw-lp	the POS of hw-lw	PN
hw-rw	the right neighbour of head word	一 (an)
hw-rp	the POS of hw-rw	CD
POS-path	the Pos series between head word and dependent word	VV/CD/M/NN/VV/MM

6. **Experiments and Discussion.** Based on maximum entropy, this section gives the results of semantic dependency relation labeling, and then analyzes the results in detail.

6.1. **Evaluation Methods.** The evaluation methods of syntactic dependency parsing are applied to the semantic dependency parsing. There are two main evaluating standards. One is *UAS* (*Unlabeled Attachment Score*) which could evaluate the dependency structure, and the other is *LAS* (*Labeled Attachment Score*) which could evaluate the dependency relationship.

$$UAS = \frac{\text{The number of correctly attached tokens}}{\text{The number of all the tokens}} \times 100\% \tag{2}$$

$$LAS = \frac{\text{The number of correctly attached and labeled tokens}}{\text{The number of all the tokens}} \times 100\% \tag{3}$$

6.2. **Experiment Results.** 9000 sentences are chosen as training data, 700 are developing data, 700 are testing data. Zhangle's maximum entropy toolkit is used to build the SDA labeling model. The iteration time is 100. In order to compare the results of ME model with other model, MST parser is also tested in our experiments. Because of the limitation of MST, only few features are used such as word, POS, and parent node, etc. The total UAS and LAS of MST model are 80.18% and 65.03% respectively. The total UAS and LAS of ME model 80.60% and 78.14%.

The results of some individual semantic relation are listed in Table 5. The table also listed the most easily confused semantic relation with the current SR, e.g, "*Agent*" and "*Experiencer*" is a couple of the most easily confused semantic relations. Besides, in order to analyze the results, the table also lists the number of occurrence time of semantic relation in the testing set.

TABLE 5. Results of semantic dependency relation assignment

Semantic Relation class	Semantic Relations (SR)	The most easily confused SR	Occurrence time of SR	LAS(%)	
				MST	EM
Subject Roles	Agent	Experiencer	572	61.71	83.04
	Existent	Possession	76	28.95	46.05
Object Roles	Patient	Content	350	65.99	79.49
	Beneficiary	Agent	1	0	0
Auxiliary Roles	TimeFin	Time	18	22.22	11.11
	Accompaniment	Succession	12	0	0
Direct Modifiers	d-Attribute	Restrictive	292	79.11	80.89
	d-Restrictive	Attribute	1317	58.16	68.68
Verb Modifiers	r-Agent	Restrictive	46	41.30	41.30
	r-Patient	Attribute	30	46.67	46.67
Verbal noun as phrase head	j-Agent	Possessor	38	34.21	25.64
	j-Patient	Restrictive	58	60.34	46.55
Syntactic Roles	s-Abandonment	Succession	4	0	0
	s-Coordinate	Succession	965	45.39	68.26

6.3. **Error Analysis.** According to the results, it could be seen that the classifier of ME is better than MST. From the point of feature addition, ME is more convenient than MST. Also, the running efficiency of ME is much better than MST.

The distribution of the semantic relations is not very evenly. Some relations such as "*Beneficiary*", "*Abandonment* ", and "*Accompaniment*" rarely occur. The data sparsity is the main reason of the low LAS.

Another important reason of the errors is that the differentiation degree among some of the relations is not very clear. "*Agent*" and "*Experiencer*", "*existent*" and "*possession*", and "*Attribute*" and "*Restrictive*" are prone to labeling wrong. The difference between "*Agent*" and "*Experiencer*" is the action independence which is very difficult to judge automatically. "*Agent*" means independent action, and "*Experiencer*" means not independent action. For example, the two sentences "我们开始吧" (*Let's begin.*) and "会议开始了" ("*The meeting began.*"), the former relation is "*Agent*" and the latter is "*Experiencer*". As for "*Attribute*" and "*d-Restrictive*", "*d-Restrictive*" has the function of distinguishing different things, e.g. "男性" (*male*) and "女性" (*female*), "大型" (*large-scale*) and "小型" (*small-scale*), while "*d-Attribute*" always expresses the attribute of the object, e.g., "漂亮" (*beautiful*), "坚强" (*strong*). Generally, in front of "*Restrictive*" word , such words as "很" and "不" could not be added while "*d-Attibute*" word could be done.

Moreover, some relations could contain other relations. For example, the extension of "*d-Restrictive*" relation is big and it's just like a dustbin of direct modifier relation. Some reverse relation and indirect relation are easy to be taken as "*d-Restrictive*" relation.

By the analysis of the results, it could be known that the boundary of semantic relation should be defined clearly. The situation such as indistinct, overlapping, inclusion should be tried to avoid. The granularity of the relation system could be coarser. Those relations which rarely occur should be combined to other relations.

7. **Conclusions and Future Work.** Deep semantic parsing is the key to understand sentence meaning. Semantic dependency parsing is one of the deep semantic parsing methods. This paper presents a set of semantic dependency parsing scheme including the tag set design, the corpus construction and semantic relation assignment. This paper integrates some Chinese semantic relation systems given by different semantic relation sets, especially HowNet, and forms a new semantic relationship system. In order to distinguish the situation of verb as a modifier with verb as the head, reverse relation is introduced. In order to express the verbal noun as a head, indirect relation is introduced. A large scale semantic dependency corpus is built based on the method of combination of automatic and manual labeling. In the process of labeling, the consistency check tool is applied to guarantee the consistency of labeling. As for semantic dependency relation assignment, maximum entropy model is applied to classify the relations. The experiment result shows that ME model is better than MST model. Error analysis is done in the paper. Many errors are caused by the unclear boundary of different relations and some are caused by data sparsity.

In order to improve the results of semantic dependency labeling, the definition of the relation tags should be considered more. In the phase of automatic labeling, other machine learning methods such as NB, SVM, DT, etc. should be used. The post process should also be considered in the future work. How to apply the semantic dependency analysis to the fields of NLP is also another important task that should be studied deeply.

Acknowledgements. We want to thank the helpful comments and suggestions from the anonymous reviewers. This work was supported by National Natural Science Foundation of China (No. 61170144).

REFERENCES

[1] Carl Mills. American Grammar: Sound, Form and Meaning. New York, Peter Lang. 1990.

[2] Fromkin, V., Rodman, R. and Hyams, N. (2002) An Introduction to Language. 7th edition. Fort Worth: Harcourt Brace Jovanovich.

[3] Nianwen Xue, Fei Xia, Fu-Dong Chiou, and Martha Palmer. 2005. The penn chinese treebank: Phrase structure annotation of a large corpus. Natural Language Engineering, 11(2):207–238.

[4] Robinson, J. J. Dependency structures and transformation rules. Language,46(2):259–285. 1970.

[5] Yuan Yulin. Matching Even-template with Argument Structure of Verbs:Towards a Verb-driven Approach of Information Extraction. Journal of Chinese Information Processing, 2005,5: 37-43.

[6] Xue, N. and Palmer, M. . Automatic semantic role labeling for Chinese verbs. In Proceedings of 19th International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, 2005.

[7] Fillmore, Charles J. (1968): The case for case. In Bach and Harms (Ed.): Universals in Linguistic Theory. New York: Holt, Rinehart, and Winston, 1-88.

[8] Feng Zhiwei. From Case Grammar to Framenet. Journal of PLA University of Foreign Language. 2006, 3:1-13.

[9] Lu Chuan, Lin Xingguang. Case Relation of Mordern Chinese. Chinese Language Learning. 1989, 5.

[10] Qiang Dong and Zhendong Dong. HowNet and Computation of Meaning. World Scientific Publishing Company. 2006.

[11] M. Marcus, G. Kim, M. Marcinkiewicz, R. MacIntyre, A. Bies, M. Ferguson, K. Katz, and B. Schasberger. 1994. The Penn treebank: Annotating predicate argument structure. In ARPA Human Language Technology Workshop.

[12] Dang Zhengfa, Zhou Qiang. Automatically Convert TreeBank from Phrase structure to Dependency Structure. Journal of Chinese Information Processing, 2004, 19(3):21~27.

[13] Zhenghua Li, Wanxiang Che, Ting Liu. A Study on Constituent-to-Dependency Conversion. Journal of Chinese Information Processing. 2008, 22(6): 14-19.

[14] M. Palmer, D. Gildea, P. Kingsbury. The Proposition Bank: An Annotated Corpus of Semantic Roles. Comput. Linguist. 2005, 31(1):71–106

[15] Covington Michael A. A fundamental algorithm for dependency parsing. Proceedings of 39th Annual ACM Southeast Conference, 2001: 95-102.

[16] Nivre, Joakim, JensNilsson. Pseudo-Projective Dependency Parsing. In Proceedings of the 43rd Annual Meeting of the Association for Computationa lLinguistics (ACL), 2005.

[17] Yamada, Hiroyasu, YujiMatsumoto. Statistical dependency analysis with support vector machines .In

Proceedings of the 8th International Workshop on Parsing Technologies (IWPT), ed. Gertjan Van Noord, 195–206. 2003.

[18] Yanqiu Shao, Zhifang Sui, Ning Mao. Chinese Semantic Role Labeling Based on Semantic Knowledge. The 6th International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE 2010), 2010: 259-265.

[19] Wanxiang Che. Kernel-based Semantic Role Labeling. PhD thesis. Harbin Institute of Technology. 2008.

[20] Wanxiang Che, Meishan Zhang, Yanqiu Shao, Ting Liu. SemEval-2012 Task 5: Chinese Semantic Dependency Parsing. First Joint Conference on Lexical and Computational Semantics (*SEM) 2012, pages 378–384, Montr´eal, Canada. 2012.