# SemEval-2016 Task 9: Chinese Semantic Dependency Parsing

| Wanxiang Che $^{\dagger}$ | Yanqiu Shao‡                 | Ting Liu $^{\dagger}$              | Yu Ding <sup>†</sup> |  |  |  |  |
|---------------------------|------------------------------|------------------------------------|----------------------|--|--|--|--|
|                           |                              |                                    |                      |  |  |  |  |
| <sup>†</sup> {car, tliu,  | yding}@ir.hit.edu.cn         | ‡yqshao@blcu.                      | .edu.cn              |  |  |  |  |
| School of Com             | puter Science and Technology | logy School of Information Science |                      |  |  |  |  |
| Harbin 1                  | Institute of Technology      | Beijing Language and Cu            | ulture University    |  |  |  |  |
| Har                       | bin, China, 150001           | Beijing, China,                    | 100083               |  |  |  |  |
|                           |                              |                                    |                      |  |  |  |  |

#### Abstract

This paper describes the SemEval-2016 Shared Task 9: Chinese semantic Dependency Parsing. We extend the traditional treestructured representation of Chinese sentence to directed acyclic graphs that can capture richer latent semantics, and the goal of this task is to identify such semantic structures from a corpus of Chinese sentences. We provide two distinguished corpora in the NEWS domain with 10,068 sentences and the TEXT-BOOKS domain with 14,793 sentences respectively. We will first introduce the motivation for this task, and then present the task in detail including data preparation, data format, task evaluation and so on. At last, we briefly describe the submitted systems and analyze these results.

#### **1** Introduction

This task is a rerun of the task 5 at SemEval 2012 (Che et al., 2012), named Chinese semantic dependency parsing (SDP). In the previous task, we aimed at investigating "deep" semantic relations within sentences through tree-structured dependencies. As traditionally defined, syntactic dependency parsing results are connected trees defined over all words of a sentence and language-specific grammatical functions. On the contrary, in semantic dependency parsing, each head-dependent arc instead bears a semantic relation, rather than grammatical relation. In this way, semantic dependency parsing results can be used to answer questions directly, like *who did what to whom when and where*.

However, according to the meaning-text linguistic theory (Mel'čuk and Žolkovskij, 1965), a theoretical framework for the description of natural languages, it is said that trees are not sufficient to express the complete meaning of sentences in some cases, which has been proven undoubted in our practice of corpus annotation. This time, not only do we refine the easy-to-understand meaning representation in Chinese in order to decrease ambiguity or fuzzy boundary of semantic relations on the basis of Chinese linguistic knowledge, we extend the dependency structure to directed acyclic graphs that conform to the characteristics of Chinese, because Chinese is an parataxis language with flexible word orders, and rich latent information is hidden in facial words.

Figure 1 illustrates an example of semantic dependency graph. Here, "她 (she)" is the argument of "脸色 (face)" and at the same time it is an argument of "病 (disease)". Researchers in dependency parsing community realized dependency parsing restricted in a tree structure is still too shallow, so they explored semantic information beyond tree structure in task 8 at SemEval 2014 (Oepen et al., 2014) and task 18 at SemEval 2015 (Oepen et al., 2015). They provided data in similar structure with what we are going to provide, but in distinct semantic representation systems. Once again we propose this task to promote research that will lead to deeper understanding of Chinese sentences, and we believe that freely available and well annotated corpora which can be used as common testbed is necessary to promote research in data-driven statistical dependency parsing.



Figure 1: An example of our proposed DAG-based semantic dependency representation.

The rest of the paper is organized as follows. Section 2 gives an overview of semantic dependency parsing, with specific focus on the proposed DAG semantic representation. Section 3 describes the technical details of the task. Section 4 presents the participating systems, and Section 5 compares and analyzes the results. Finally, Section 6 concludes the paper.

## 2 Semantic Dependency Parsing

Given a complete sentence, semantic dependency parsing (SDP) aims at determining all the word pairs related to each other semantically and assigning specific predefined semantic relations. Semantic dependency analysis represents the meaning of sentences by a collection of dependency word pairs and their corresponding relations. This procedure survives from syntactic variations.

In this paper, we define a Chinese semantic dependency scheme based on Chinese-specific linguistic knowledge, which represents the meaning of sentences in graphic formats (Figure 1).

## 2.1 Structure of Chinese Semantic Dependency Graph

We used semantic dependency graphs to represent the meanings of sentences, which contain dependency relations between all the word pairs with direct semantic relations. Predicates includes most predicative constituents (i.e. most verbs and a small number of nouns and adjectives), and arguments are defined as all the possible participants in the real scene corresponding to a certain predicate (e.g. the eater, food, tool, location, time in the scene related to "eat"). One principle of building dependency arcs is to find arguments for predicates in content words preferentially because they are the ones that related



2. Difference between syntactic and semantic

**Figure 2:** Difference between syntactic and semantic dependency on preposition

to predicates directly. Unlike syntactic dependency, which inserts non-content words between predicate and its "real arguments" (Figure 2). Due to the completeness of the representation of relations between words, some words have relations with more than one other word (some words have more than one child, and some have more than one father), which forms direct acyclic graphs finally. We define a set of labels to describe dependency relations between words.

#### 2.2 Semantic Dependency Relations

On the basis of SemEval 2012 task 5, we refined the semantic relation set in terms of more solid Chinese linguistic theories, except for the reference of HowNet (Dong and Dong, 2006), a popular Chinese semantic thesaurus. We mainly referred to the idea of semantic network of Chinese grammar defined by Lu (2001). He adapted semantic network to Chinese, which is the formal network for "semantic composition systems" by distinguishing the hierarchies of "semantic relations", "semantic alignment" and "semantic orientation". We borrowed his ideas of semantic unit classification and semantic combination, and integrated them with the dependency grammar to re-divided boundary for each semantic relation and re-organized the total label set for clarity and definiteness.

Semantic units are divided from high to low into event chains, events, arguments, concepts and markers. Arguments refer to noun phrases related to certain predicates. Concepts are simple elements in basic human thought or content words in syntax. Markers represent the meaning attached to entity information conveyed by speakers (e.g., speakers' tones or moods). These semantic units correspond to compound sentences, simple sentences, chunks, content words and function words. The meaning of sentences is expressed by event chains. Event chains consist of multiple simple sentences. The meaning of simple sentences is expressed by arguments, while arguments are reflected by predicate, referential or defining concepts. Markers are attached to concepts.

The meaning of sentences consists of the meaning of semantic units and their combinations, including semantic relations and attachments. Semantic attachments refer to markers on semantic units. Semantic relations are classified into symmetric and asymmetric types. Symmetric relations include coordination, selection, and equivalence relations, while asymmetric relations include:

Collocational relations occur between core and non-core roles. For example, in "工人 (worker) 修 理 (repair) 地下 (pipeline) 管道 (pipeline)" serves as a non-core role, and is the patient of "修理 (repair)," which is a verb and core role. Relations between predicates and nouns belong to collocational relations. Semantic roles usually refer to collocational relations, Table 1 presents the 32 semantic roles we defined, divided into eight small categories. Additional relations refer to the modifying relations among concepts within an argument; all semantic roles are available, e.g. in "地下 (underground) 的 (de) 管道 (pipeline)", "地下 (underground)" is the modifier of "管道 (pipeline)", which refers to location relation. Connectional relations are bridging relations between two events that are neither symmetric nornested relation. For example, for "如果 (If) 天气 (weather)好 (good), (,) 我 (I) 就 (will) 去 (go) 故宫 (the Summer Palace)," the former event

is the hypothesis of the latter. Events in Chinese semantic dependency have 15 relations. According to the above classification of sentence hierarchies, we can get to know how each sentence component constitutes the entire meaning of sentences. We design semantic dependency relations in terms of this theoretical basis.

## 2.3 Special Situations

On the analysis of the nature of Chinese language, two special situations need special handling. we list them here and describe their annotation strategies.

- Reverse relations. When a verb modifies a noun, a reverse relation is assigned with the label r-XX (XX refers to a single-level semantic relation). Reverse relation is designed because a word pair with the same semantic relation appears in different sentences with different modifying orders. Reverse relation is used to distinguish them. For example, the verb "打 (play)" modifies the kernel word "男孩 (boy)" in (a) of Figure 3, so r-agent is assigned; while in (b) "打 (play)" is a predicate and "男孩 (boy)" is the agent role with the label agent.
- Nested events. We define another kind of special relation-nested relation to mark that one sentence is degraded as a constituent for another sentence. Two events have a nested relation, i.e., one event is degraded as a grammatical item of the other, which belong to two semantic hierarchies. For example, in the sentence in Figure 4, the event "小 (little) 孙女 (granddaughter) 在 (be) 玩 (play) 电脑 (computer)" is degraded as a content of the action of "看见 (see)". A prefix "d" is added to single-level semantic relations as distinctive label.

Finally, we got 45 labels to describe relations between main semantic roles and relations within arguments, 19 labels for relations between different predicates. We also defined 17 labels to mark the auxiliary information of predicates. The total semantic relation set is shown in Table 1.

## 3 Task Description

This task contains two tracks, which are closed track and open track. People participating in closed track

| Semantic roles          |   |  |  |  |  |  |  |  |
|-------------------------|---|--|--|--|--|--|--|--|
| Subject roles           | Agt(Agent), Exp(Experiencer), Aft(Affection), Poss(Possessor)   |  |  |  |  |  |  |  |
| Object roles            | Pat(Patient), Cont(Content), Prod(Product), Orig(Origin), Comt(Comitative)<br>Comp(Comparison)  |  |  |  |  |  |  |  |
| Copula roles            | Belg(Belongings), Clas(Classification), Accd(According)   |  |  |  |  |  |  |  |
| Cause roles             | Reas(Reason), Int(Intention), Cons(Consequence)   |  |  |  |  |  |  |  |
| Condition roles         | Mann(Manner), Tool, Matl(material)  |  |  |  |  |  |  |  |
| Space-time roles        | Time, Loc(Location), Dir(Direction), Proc(Process), Sco(Scope)  |  |  |  |  |  |  |  |
| Measurement roles       | Quan(Quantity), Qp(Quantity-phrase), Freq(Frequency), Seq(Sequence)   |  |  |  |  |  |  |  |
| Special attribute roles | Desc(Description), Host, Nmod(Name-modifier), Tmod(Time-modifier)   |  |  |  |  |  |  |  |
| Reverse relations       | r + semantic roles  |  |  |  |  |  |  |  |
| Nested relations        | d + semantic roles  |  |  |  |  |  |  |  |
| Event relations         |   |  |  |  |  |  |  |  |
| Symmetric relations     | eCoo(Coordination), eSelt(Selection), eEqu(Equivalent)  |  |  |  |  |  |  |  |
| Consecutive relations   | ePrec(Precedent), eSucc(Successor), eProg(Progression), eCau(Cause)<br>eAdvt(adversative), eResu(Result), eInf(Inference), eCond(Condition), eS<br>upp(Supposition), eConc(Concession), eAban(Abandonment), eMetd(Method)<br>ePurp(Purpose), ePref(Preference), eSum(Summary), eRect(Recount) |  |  |  |  |  |  |  |
| Semantic markers        |   |  |  |  |  |  |  |  |
| Relation markers        | mConj(Conjection), mAux(Auxiliary), mPrep(Preposition)  |  |  |  |  |  |  |  |
| Attachment markers      | mTone, mTime, mRang(Range), mDegr(Degree), mMod(Modal)<br>mFreq(Frequency), mDir(Direction), mPars(Parenthesis), mNeg(Negation)   |  |  |  |  |  |  |  |
| Auxiliary markers       | mMaj(Majority), mSepa(Separation), mRept(Repetition), mVain<br>mPunc(Punctuation)   |  |  |  |  |  |  |  |

Table 1: Label set of the semantic relation of BH-SDP-v2



could use only the given training data to train their systems and any other additional resource is forbidden. While any knowledge beyond training data is allowed in open track. The closed track encourages participates to focus on building dependency



Figure 4: Nested relations.

parsing systems on graphs and the open track stimulates researchers to try how to integrate linguistic resource and world knowledge into semantic dependency parsing. The two tracks will be ranked separately. We provide two training files containing sentences in each domain. There is no rules for the use of these two training data.

#### **3.1** Corpus Statistics

Since texts in rich categories have different linguistic properties with different communication purpose. This task provides two distinguished corpora in appreciable quantity respectively in the domain of NEWS and TEXTBOOKS (from primary school textbooks). Each corpus contains particular linguistic phenomena. We provide 10,068 sentences of NEWS and 14,793 sentence of TEXTBOOKS. The sentences of news keep the same with the data in task 5 at SemEval 2012, which come from the Chinese PropBank 6.01 (Xue and Palmer, 2003) as the raw corpus to create the Chinese semantic dependency corpus. Sentences were selected by index: 1-121, 1001-1078, 1100-1151. TEXTBOOKS refer to shorter sentences with various ways of expressions, i.e., colloquial sentences (3,000), primary school texts (11,793). Detailed statics are described in Table 2.

|      |       | Train   | Dev    | Test   |
|------|-------|---------|--------|--------|
| NEWS | #sent | 8,301   | 534    | 1,233  |
|      | #word | 250,249 | 15,325 | 34,305 |
| TEXT | #sent | 10,817  | 1,546  | 3,096  |
|      | #word | 128,095 | 18,257 | 36,097 |
|      |       | G       |        |        |

Table 2: Statics of the corpus.

## 3.2 Data Format

All data provided for the task uses a columnbased file format, similar to the one of the 2006 CoNLL Shared Task (Table 3). Each training/developing/testing set is a text file, containing sentences separated by a blank line. Each sentence consists of more than one tokens, and each token is represented on one line consisting of 10 fields. Buchholz and Marsi (2006) provide more detailed information on the format. It's worth noting that if one word has more than one heads, it will appear in more than one lines in the training/developing/testing files continuously. Fields are separated from each other by a tab. Only five of the 10 fields are used: token id, form, pos tagger, head, and deprel. Head denotes the semantic dependency of each word, and deprel denotes the corresponding semantic relations of the dependency. In the data, the lemma column is filled with the form and the cpostag column with the postag.

## 3.3 Evaluation

During the phase of evaluation, each system should propose parsing results on the previously unseen testing data. Similar with training phase, testing files containing sentences in two domains will be released separately. The final rankings will refer to the average results of the two testing files (taking training data size into consideration). We compare predicted dependencies (predicate-roleargument triples, and some of them contain roots of the whole sentences) with our human-annotated ones, which are regarded as gold dependencies.

Our evaluate measures are on two granularity, dependency arc and the complete sentence. Labeled and unlabeled precision and recall with respect to predicted dependencies will be used as evaluation measures. Since non-local dependencies (following Sun et al. (2014), we call these dependency arcs making dependency trees collapsed non-local ones) discovery is extremely difficult, we will evaluate non-local dependencies separately. For sentences level, we will use labeled and unlabeled exact match to measure sentence parsing accuracy. Following Task 8 at SemEval 2014, below and in other taskrelated contexts, we abbreviate these metrics as:

- Labeled precision (LP), recall (LR), F1 (LR) and recall for non-local dependencies (NLR);
- Unlabeled precision (UP), recall (UR), F1 (UF) and recall for non-local dependencies (NUR);
- Labeled and unlabeled exact match (LM, UM).

When ranking systems participating in this task, we mainly refer to the average F1 (LF) on the two testing sets.

## 4 Participating Systems

Fifteen organizations were registered to participate in this task. Finally, five systems were received from three organizations. These systems are as follows:

1. IHS-RD-Belarus. This system applied transition-based dependency parsing with online reordering, in order to deal with non-projective dependency arcs. The model is trained with both gold training instances and auto-parsed training instances, referred to as *bootstrapping*. Additional semantic features extracted from the IHS Goldfile Question-Answering system are utilized and demonstrated to be effective. It also used graph

| ID | FORM | LEMMA | CPOS | PPOS | FEAT | HEAD | REL   | PHEAD | PREL |
|----|------|-------|------|------|------|------|-------|-------|------|
| 1  | 现在   | 现在    | NT   | NT   | -    | 4    | Time  | -     | _    |
| 2  | 她    | 她     | PN   | PN   | _    | 3    | Poss  | _     | -    |
| 2  | 她    | 她     | PN   | PN   | _    | 7    | Exp   | _     | _    |
| 3  | 脸色   | 脸色    | NN   | NN   | _    | 4    | Exp   | _     | _    |
| 4  | 难看   | 难看    | VA   | VA   | _    | 7    | eCau  | -     | _    |
| 5  | ,    | ,     | PU   | PU   | _    | 4    | mPunc | -     | -    |
| 6  | 好像   | 好像    | AD   | AD   | -    | 7    | mMod  | -     | -    |
| 7  | 病    | 病     | VV   | VV   | -    | 0    | ROOT  | -     | -    |
| 8  | 了    | 了     | SP   | SP   | _    | 7    | mTone | -     | -    |

Table 3: Data format.

pre- and post-processing to handle ambiguities of some specific semantic relations (i.e., *eCoo*).

- 2. OCLSP (lbpg, lbpgs, lbpg75). This system proposed a bottom-up neural parser using long short-term memory (LSTM) networks. Since the basic neural parser (lbpg) has no guarantee to produce a dependency graph, they applied Chu-Liu-Edmond's algorithm (Chu and Liu, 1965) to generate the minimal spanning directed graph (lbpgs). To further address the multi-head annotation in this task, a threshold of  $\delta$  is set on the probabilities to decide whether an extra arc exists (lbpg75).
- OSU\_CHGCG. This system proposed to use parsers trained with Chinese Generalized Categorial Grammar (GCG) (Bach, 1981) annotations to obtain the syntactic structures of a sentence. Then the GCG features, along with traditional features (e.g., word, POS, etc.) are fed into a multinomial classifier for semantic dependency classification.

## 5 Results & Analysis

We use LF, UF and NLF, NUF as the main evaluation metrics here. Table 4 shows the results of all participating systems. Note that all of the submitted systems used additional resource beyond training data provided in the task. IHS-RD-Belarus used semantic features extracted from the output of IHS Goldfire Question-Answering system, and both OCLSP and OSU\_CHGCG used GCG features. Therefore, the results reported in Table 4 are all in the open track. Overall, the IHS-RD-Belarus system achieves the best results in both NEWS and TEXT domain. However, it didn't perform well on the prediction of nonlocal labeled dependencies. OSU\_CHGCG instead behaves more promising in the prediction of nonlocal dependencies. OCLSP (lbpg75) achieves remarkable results in the non-local labeled dependencies of the TEXT domain (57.51 of NLF).

From the perspective of methodology, IHS-RD-Belarus is a tree-structure prediction system, lacking the ability of revealing multi-head structures; while both OCLSP and OSU\_CHGCG deal with graph structure, with either post-processing or classification-based models. From the perspective of resource, all the systems demonstrated that features extracted from a syntactic or semantic source are helpful for the SDP task, which is expected.

In general, some novel methods and ideas were proposed for this task, providing evidences for future research on both model design and feature selection of semantic dependency parsing.

#### 6 Conclusion

We described the Chinese Semantic Dependency Parsing task for SemEval-2016, which is designed to analyze the graph-structured semantic representation of Chinese sentences. Five systems were submitted by three organizations. The systems explored the semantic dependency parsing problem along different directions, which will significantly push forward the research on SDP. We also note that the performance of SDP is still far from promising, especially for labeled dependencies and non-local dependencies. Challenges still remain in designing more effective and efficient parsing algorithms for graph-

|      | System         | LP    | LR    | LF    | UP    | UR    | UF    | NLF   | NUF   | LM    | UM    |
|------|----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|      | IHS-RD-Belarus | 58.78 | 59.33 | 59.06 | 77.28 | 78.01 | 77.64 | 40.84 | 60.20 | 12.73 | 20.60 |
|      | OCLSP (lbpg)   | 55.64 | 58.89 | 57.22 | 72.87 | 77.11 | 74.93 | 45.57 | 58.03 | 12.25 | 18.73 |
| NEWS | OCLSP (lbpgs)  | 58.38 | 57.25 | 57.81 | 76.28 | 74.81 | 75.54 | 41.56 | 54.34 | 12.57 | 20.11 |
|      | OCLSP (lbpg75) | 57.88 | 57.67 | 57.78 | 75.55 | 75.26 | 75.40 | 48.89 | 58.28 | 12.57 | 19.79 |
|      | OSU_CHGCG      | 55.52 | 55.85 | 55.69 | 73.51 | 73.94 | 73.72 | 49.23 | 60.71 | 5.03  | 11.35 |
|      | AVG            | 57.24 | 57.80 | 57.51 | 75.10 | 75.83 | 75.45 | 45.22 | 58.31 | 11.03 | 18.12 |
|      | IHS-RD-Belarus | 68.71 | 68.46 | 68.59 | 82.56 | 82.26 | 82.41 | 50.57 | 64.58 | 16.82 | 40.12 |
|      | OCLSP (lbpg)   | 63.34 | 67.89 | 65.54 | 76.73 | 82.24 | 79.39 | 51.75 | 63.21 | 11.49 | 27.60 |
| TEXT | OCLSP (lbpgs)  | 67.35 | 65.11 | 66.21 | 81.22 | 78.52 | 79.85 | 47.79 | 55.51 | 12.82 | 33.29 |
|      | OCLSP (lbpg75) | 66.43 | 66.43 | 66.38 | 79.97 | 79.85 | 79.91 | 57.51 | 63.87 | 12.56 | 32.09 |
|      | OSU_CHGCG      | 65.36 | 64.98 | 65.17 | 79.06 | 78.60 | 78.83 | 54.70 | 65.71 | 11.36 | 32.02 |
|      | AVG            | 66.24 | 66.57 | 66.38 | 79.91 | 80.29 | 80.08 | 52.46 | 62.58 | 13.01 | 33.02 |

Table 4: Results of the submitted systems.

structured semantics. The annotation standard of semantic dependencies and the quality of our proposed corpus may also be further improved, which we leave to future work.

#### Acknowledgments

This work was supported by the National Key Basic Research Program of China via grant 2014CB340503 and the National Natural Science Foundation of China (NSFC) via grant 61133012 and 61370164.

#### References

- Emmon W Bach. 1981. Discontinous constituents in generalized categorial grammar.
- Sabine Buchholz and Erwin Marsi. 2006. Conll-x shared task on multilingual dependency parsing. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)*, pages 149–164, New York City, June. Association for Computational Linguistics.
- Wanxiang Che, Meishan Zhang, Yanqiu Shao, and Ting Liu. 2012. Semeval-2012 task 5: Chinese semantic dependency parsing. In \*SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012), pages 378–384, Montréal, Canada, 7-8 June. Association for Computational Linguistics.
- Yoeng-Jin Chu and Tseng-Hong Liu. 1965. On shortest arborescence of a directed graph. *Scientia Sinica*, 14(10):1396.

- Zhendong Dong and Qiang Dong. 2006. *HowNet and the Computation of Meaning*. World Scientific.
- Chuan Lu. 2001. *The Semantic Network of Chinese Grammar*. The Commercial Printing house.
- Igor Mel'čuk and AK Žolkovskij. 1965. O vozmožnom metode i instrumentax semantičeskogo sinteza. *Naučno-texničeskaja informacija*, (5):23–28.
- Stephan Oepen, Marco Kuhlmann, Yusuke Miyao, Daniel Zeman, Dan Flickinger, Jan Hajic, Angelina Ivanova, and Yi Zhang. 2014. Semeval 2014 task 8: Broad-coverage semantic dependency parsing. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pages 63–72, Dublin, Ireland, August. Association for Computational Linguistics and Dublin City University.
- Stephan Oepen, Marco Kuhlmann, Yusuke Miyao, Daniel Zeman, Silvie Cinkova, Dan Flickinger, Jan Hajic, and Zdenka Uresova. 2015. Semeval 2015 task 18: Broad-coverage semantic dependency parsing. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), pages 915– 926, Denver, Colorado, June. Association for Computational Linguistics.
- Weiwei Sun, Yantao Du, Xin Kou, Shuoyang Ding, and Xiaojun Wan. 2014. Grammatical relations in chinese: Gb-ground extraction and data-driven parsing. In ACL (1), pages 436–456.
- Nianwen Xue and Martha Palmer. 2003. Annotating the propositions in the penn chinese treebank. In *Proceedings of the Second SIGHAN Workshop on Chinese Language Processing*.