

A Model of Semantic Representations Analysis For Chinese Sentences

Li Tang , Donghong Ji , Lingpeng Yang, Yu Nie
Institute for Infocomm Research
21 Heng Mui Keng Terrace, Singapore 119613
{tangli, dhji, niey, lpying}@i2r.a-star.edu.sg

Abstract

Analyzing the semantic Representations of 5000 Chinese sentences and describing a new sentence analysis method that evaluates semantic preference knowledge, we create a model of semantic representation analysis based on the correspondence between lexical meanings and conceptual structures, and relations that underlie those lexical meanings. We also propose a semantical argument-head relation that combines ‘basic conceptual structure’ and ‘Head-Driven principle’.

1 Introduction

This paper presents the method and design model for labeling semantic dependency structures and feature structures based on words. We have been concentrating on analyzing the semantic relations among different constituents in a single sentence. To enable computer-based analysis of Chinese sentences in natural language texts we have developed a semantic framework, using the English language framework created by Fillmore et al at UC Berkeley as a starting point. In our case, the key task is to determine the relationship between the two direct constituents in terms of the semantic relationship. The grammar functions are also considered for primary identifying the relation.

2 Sentence Analysis methods within Our Semantic Model

We have selected 5000 sentences from linguists’ works and website. These sentences cover not only large scale of domains, but also different kinds of sentential patterns that capture the morph syntactic diversity. It is propitious to give a comprehensive analysis of the corpus and to create a standard semantic model, so we first summarize the first fully instantiated semantic models labeling of semantic representations by human beings in a machine-readable format. Also we have extracted

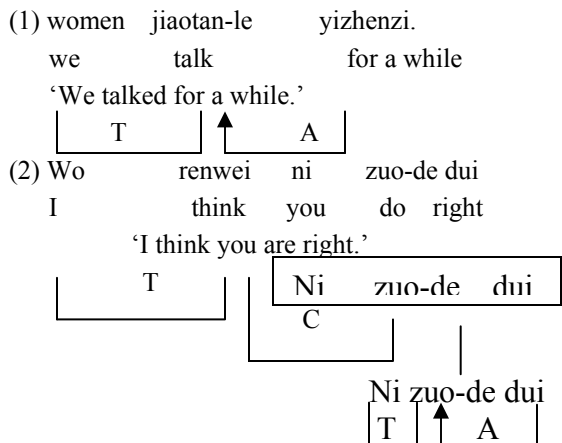
diversity. It is propitious to give a comprehensive analysis of the corpus and to create a standard semantic model, so we first summarize the first fully instantiated semantic models labeling of semantic representations by human beings in a machine-readable format. Also we have extracted several other thousands of sentences from large text corpora and have done a systematic analysis of the semantic relations on the basis of our semantic model. The present paper indicate how situation type are represented, how these representations are compose from the semantic representations of linguistic constituents, and how these type differences affect the expression of sentence.

Semantic Relation Labeling

This work flow include linking and hand-tagged labeling of each relation between direct semantic unit in single sentence, which reflects different semantic representation of the potential realization patterns identified in the formula, and descriptions of the relations of each frame’s basic conceptual structure in terms of semantic actions. A semantic representation is a feature that allows one word in the sentence to point at some other word to which it is related. A word in a sentence may have much direct representation; these are differentiated by the

semantic action. By analyzing the direct semantic representation, we can capture semantic relationships between words. In order to reducing the numerous and complex phrases in the elementary dependency trees, we more readily exhibit the fundamental binary relations between content words by limiting maximal depth of the subtrees involved in the parsing process was varied from 1 to 3, than using patterns, such limitations subtrees involved in the parsing process was varied from 1 to 3, than using patterns, such limitations improve probability estimations, while retaining the full power of framework.

For example, we have labeled the direct relationships of different semantic units within single sentences as follow:



The above two sentences show the semantic model that is based on the analysis of the semantic relations. In (1), for example, the AGENT *women* 'we' is directly related to the action verb *jiaotan* 'talk'. We link them together and label their semantic relations as 'A' representing 'AGENT'. The time word *yizhenzi* 'for a while', it complements the 'core' verb, so we link them use a 'head' separating the 'core' from the 'adjunct', and tag their semantic relations as 'T' representing 'time'. In (2), there are three semantic units, two words and one pattern. They are on the same root level; we link and label the relation between each direct semantic unit as 'A' representing 'Agent' and 'C' representing 'Content', the third semantic unit is composed of pattern which reanalyzed as

two subtrees attached each relation labeling as 'A' and 'J' representing 'Judgment'

Direct Relations Determining

The basic link is the direct link between two semantic units. In addition, a set of general rule for determining the direct relations has been identified. There are summarized into three major conditions.

1. That between Head and Its Modifier as a Case of Direct Relationship

The headword and modifiers that come before it hold a kind of modification relationship, which is one of the typical cases of direct relationships. E.g.,

- (3) a. Gao zige de ren
 tall body DE person
 the person with tall body
 b. ren de gezi gao
 person DE body tall
 'The person's body is tall.'

In the above sentence, *ren* 'person' and *gezi* 'body' hold a modification relationship, but *gao* 'tall' and *ren* 'person' is related indirectly as the relationship between the two words is realized through that of *gezi* 'body'. Therefore, we say that the relationship that *ren* 'person' holds with *gezi* 'body' is the most direct one, but that with *gao* is a rather indirect one.

2. That between an Action Verb and Its Patient as a Case of Direct Relationship

In case a head noun is an AGENT of an action verb within a modifying phrase, then the relationship between the head noun and the action verb is a direct one. The following sentence illustrates the point.

- (4) a. chi pingguo de nuhai.
 Eat apples DE girl
 'the girl who is eating apples.'
 b. nuhai chi pingguo
 girl eat apples
 'The girl is eating apples.'

In the above sentence, *nulai* 'girl' is an AGENT of the action verb *chi* 'eat', the two words are directly related, they thus are considered as holding a direct relationship. Contrastively, the

relationship between *nulai* ‘girl’ and *pingguo* ‘apples’ is of an indirect type.

3. Other Cases of Direct Relationships

In case there is neither modification nor AGENT/PATIENT relationship, the whole phrase, which is still directly related to a following describing phrase, has to be embedded. E.g.,

(5) *ban shiqing yinggai guquan daju.*
 Handle problem should care-about overall situation

‘People should care about the overall situation when they handle problems.’

(6) *chouyan shang shenti.*

Smoke harm health
 ‘Smoking harms health.’

(7) *ta neng daying de shiqing wo ye neng daying.*

He can accept DE issues I also can accept
 ‘The issues that he can accept are also acceptable to me.’

‘Head’ Determining

In this study, we have proposed an approach that combines ‘basic conceptual structure’ and ‘Head-Driven Principle’. By ‘Head-Driven Principle’, most structures are analyzed as having a ‘Head’ modified by various types of modifiers. The exceptions are ‘Subject-Predicate Structure’ and ‘Verb-Object Structure’. In above two sentences, for example, the relation linking between the ‘core’ noun and verb with their ‘adjunct’ is tagged with arrow as ‘head’. ③ and ④ are labeled the ‘head’ on the core noun. ⑤ is labeled the ‘head’ on the core verb. Employing the ‘Head-Driven Principle’ for the construction of semantic model, some ambiguous sentences can be clearly represented.

3 the advantage of semantic model

As for the one-to-more correspondence between syntactic-semantic structures of Chinese sentences, we have found many interesting language phenomena and have developed mechanisms to capture those phenomena in our NLU (Natural

Language Understanding) –oriented theoretical frameworks.

In developing our semantic tree bank, we also have articulated a framework of ‘Noun-Centrality’ as a supplement to the widely assumed ‘Verb-Centrality’ practice. Based on such model, we can successfully disambiguate some troublesome sentences, and minimize the redundancy in language knowledge description for natural language processing. We automatically learn a simpler, less redundant representation of the same information.

One semantic structure corresponds to more syntactic structures

Please consider the following examples: (8)(9)(10)

Ta da-le wo She beat me ‘She beat me.’	Ta ba wo da-le She BA me beat ‘She beat me.’	Wo BEI Ta da-le I BEI she beat ‘I have been beat by her.’
--	--	---

The above three sentences, their syntactic structures are clearly different from each other. That is, the direct object *wo* ‘me’ appears right after the main verb in (8) whereas the same logical object have moved to a pre-verbal position with a help of a special Chinese preposition *BA* in (9) and to a sentence-initial position with the help of *BEI* in (10). But underlying the difference syntactic structures, they share the same basic semantic structure, using semantic represented expression. (11) The Same Semantic Structure of Different Sentences in (8), (9) and (10)

AGENT	<i>Ta</i> ‘she’
PATIENT	<i>Wo</i> ‘me’
ACTION	<i>Da</i> ‘beat’

Several different sentences which should be analyzed as having the same syntactic structure may have fundamentally different semantic structures. The following three sentences (12), (13) and (14), for example, should be analyzed as having the syntactic structure (11), but their semantic structures are nevertheless represented as (12’), (13’) and (14’) respectively in our framework.

NP + V + Adj + NP

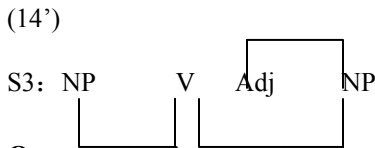
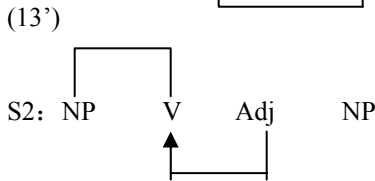
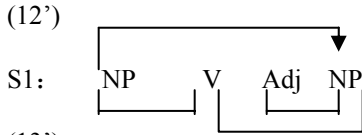
(12) S1 = Ta ku-hong le yanjing

he cry-read ASP eye

‘He cried so much that his eyes are read.’

(13) S2=Wo kan-tou le ni
 I see-through ASP you
 'I understand you thoroughly.'

(14) S3=Ta da po-le beizi
 She broke up the cup
 She broke up the cup.

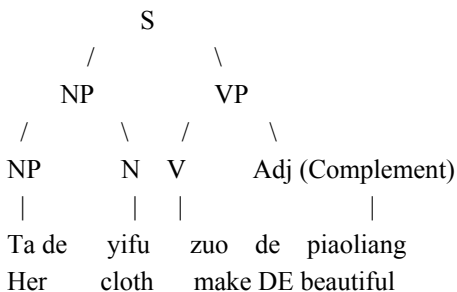


One syntactic structure corresponds to more semantic structures

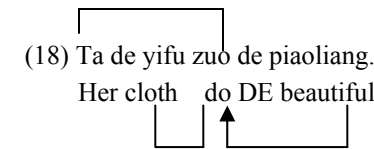
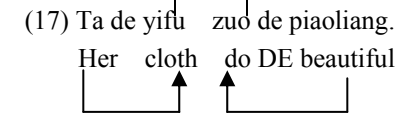
In the Chinese language, one syntactic structure may correspond to two or more semantic structures, that is, various forms of structural ambiguity are widely observed in the language. Disregarding the semantic types will cause syntactic ambiguity. The semantic types of constituents often rise to differences in semantic structure. If this type information is not available during parsing, important clues will be missing, and loss of accuracy will result. Please see the following example.

(15) Ta de yifu zuo de piaoliang.
 Her cloth do DE beautiful
 'She has made the cloth beautifully OR (Somebody) has made her cloth beautifully.'

Syntactically, the sentence, with either one of the above two semantic interpretations, should be analyzed as (16)



But the two semantic structures have to be properly represented in a semantics-oriented treebank. We do so as in (11) and (12) respectively.



So under our proposal, the above two different types of semantic relations can be clearly represented.

3 Conclusions

In this paper we have demonstrated how the semantic model can be created to analyze and represent the semantic relationships of Chinese sentence structures. The semantic model project is producing a structured tree bank with a richer set of semantic and syntactic relationships. Of difference words on the basis of the analysis of lexical meanings and conceptual structures that underlie those lexical meanings.

Reference

Baker, C.F., C.J.Fillmore & J.B.Low. (1998) The Berkeley FrameNet project. Accepted for August 1998 ACL/COLING.
 Lower J, Baker C, Fillmore C 1997 A frame-semantic Approach to Semantic Annotation. In *Tagging Text with Lexical Semantics: Why, What, and How? Proceeding of the Workshop, Special Interest Group on the Lexicon*, Associa