Computatioal Linguistics and Chinese Language Processing
vol 3, no. 1 , February 98, pp. 61-74
©Computational Linguisitcs Society of R. O. C.

Towards a Representation of Verbal Semantics -- An Approach Based on Near-Synonyms

Mei-Chih Tsai*, Chu-Ren Huang*, Keh-Jiann Chen*, Kathleen Ahrens+

Abstract

In this paper we propose using the distributional differences in the syntactic patterns of near-synonyms to deduce the relevant components of verb meaning. Our method involves determining the distributional differences in syntactic patterns, deducing the semantic features from the syntactic phenomena, and testing the semantic features in new syntactic frames. We determine the distributional differences in syntactic patterns through the following five steps: First, we search for all instances of the verb in the corpus. Second, we classify each of these instances into its type of syntactic function. Third, we classify each of these instances into its argument structure type. Fourth, we determine the aspectual type that is associated with each verb. Lastly, we determine each verb's sentential type. Once the distributional differences have been determined, then the relevant semantic features are postulated. Our goal is to tease out the lexical semantic features as the explanation, and as the motivation of the syntactic contrasts.

1. Introduction

Radical Lexicalism maintains that all grammatical behaviors are manifestations of lexical features [Karttunen 1986]. Since most lexical attributes are semantic and/or conceptual in nature, taking this lexicon-driven approach to language means that many syntactic properties can be predicted from lexical semantic attributes [Jackendoff 1976, Levin 1985, Dowty 1991, Pustejovsky 1993]. In terms of Natural Language Processing (NLP), surface syntactic structures can be systematically predicted from their lexical semantic representation. From this perspective, the automatic acquisition of lexical knowledge for NLP may be possible, since the relation between syntactic patterns and lexical semantics is predictable to some extent. Dorr & Jones [1996], for example, demonstrate that semantic information can be derived from syntactic cues when the syntactic cues are first divided into distinct groupings that correlate with different word senses.

* Academia Sinica, Nankang, Taipei 115, Taiwan, R.O.C. E-mail: tsmei@mail.nsysu.edu.tw
+ National Taiwan University, Taipei 106, Taiwan, R.O.C.

However, as Levin [1993] points out, there are still many questions to be explored:

> If the hypothesis that syntactic properties are semantically determined is taken seriously, then the task is to determine, first, to what extent the meaning of a verb determines its syntactic behavior, and second, to the extent that syntactic behavior is predictable, what components of verb meaning figure in the relevant generalizations. The identification of the relevant components of meaning is essential if this approach is to be successful.[Levin 1993:14]

Our paper will focus on the last point above. We propose using the distributional differences in the syntactic patterns of near-synonyms[1] to deduce the relevant components of verbal semantics. In particular, we want to identify the semantic features that differentiate verbal syntactic behaviors. Our strong hypothesis is that all lexical semantic features can be identified this way. In contrast, salient semantic features deduced from a shared verb class may or may not be predictive of verbal features because they may simply be descriptions of the meaning. Our method is as follows:

1) Determine distributional differences in syntactic patterns

2) Deduce the semantic features from the syntactic phenomena

3) Test the semantic features in new syntactic frames

How will we determine the distributional differences in syntactic patterns? Our corpus-based approach calls for us to search, sort, and classify all relevant data according to the four following criteria: First, we will classify each of these instances according to the syntactic functions of the verbs themselves (i.e. predicate, complement, adverbial, determinal, nominal). Second, we can classify the corpus data in terms of argument type that the verbs take (i.e. NP subject, VP subject, sentential subject, NP object, NP double-object, sentential object). Third, we determine the aspectual types each verb is associated with (i.e. aspectual markers, aspectual adverbs, resultative complements). Lastly, we examine the sentential modes that each verb occurs in (i.e. passive, imperative, evaluative, declarative, interrogative).

---

1. According to Lyons (1995: 60), synonyms are expressions with the same meaning, whereas near-synonyms are expressions that are more or less similar, but not identical, in meaning. In this respect, many of the expressions listed as synonymous in dictionaries, actually, are near-synonyms.

This process is time-consuming. However, because we are dealing with near-synonyms, we expect there to be many shared syntactic behaviors that can be ignored for the purpose of this study. This will facilitate the identification of (sometimes unexpected) grammatical contrasts that instantiates deeper lexical semantic contrasts of the near-synonym pairs. The crucial difference will be found in the small number of instances where they are in complementary distribution in terms of one of the above four types of syntactic information.[2] In what follows we will present our 3 - step methodology (i.e. determine syntactic difference, deduce semantic feature, test for reliability of semantic feature) for each of the 4 different types of syntactic information (i.e. syntactic functions (Section 2), argument structure (Section 3), aspectual type (Section 4), sentential type (Section 5). In the concluding section (Section 6), we discuss the advantages of this method as compared to an account that is based on differentiating semantic classes of verbs [Levin 1993].

2. Syntactic functions

In this section, we look at what type of syntactic functions a verb can occur with, including predicate, adverbial, complement, nominalization, etc.

2.1 Distributional differences

The distributional contrasts in terms of the syntactic functions between the two state verbs LEI 'be tired' and PIJUAN 'be tired' are that LEI functions as a (resultative) complement in 6% of the cases, but never occurs in a nominal phrase, while PIJUAN serves as a noun in 9% of the instances, but never occurs in a (resultative) complement position. The data from the Academia Sinica Balanced Corpus[3] (abb. Sinica Corpus) is given in Table 1 and the relevant examples are given in (1) and (2). (The numbers next to the verbs in the table indicate the number of instances of occurrence in the entire Sinica Corpus.)

Functions	Complement	Nominalization
LEI 174	11 (6%)	--
PIJUAN 33	--	3 (9%)

Table 1. *Table 1. Differences in syntactic functions: LEI vs. PIJUAN*

2. Due to space limit, only three pairs of verbs are illustrated in this paper: LEI-PIJUAN, 'be tired', QUAN-SHUIFU, 'persuade', GAOXING-KUAILE, 'be happy'. The last pair GAOXING and KUAILE are not included in Teng's synonyms dictionary because their corresponding terms in English are different.

3. Academia Sinica Balanced Corpus is the largest balanced corpus of both written and spoken contemporary Mandarin, developed by CKIP group in Academia Sinica, Taiwan, containing 3.5 million words.

(1) Resultative complement

(1a) ta zou de hen lei[4]
he walk DE very be-tired
'He walked so much that he was tired.'

(1b) # ta zou de hen pijuan
he walk DE very be-tired

(2) Nominalized object

(2a) shuimian shi zhi pijuan zuihaode fangfa
sleep be treat be-tired best method
'Sleeping is the best method to treat the tiredness.'

(2b) # shuimian shi zhi lei zuihaode fangfa
sleep be treat be-tired best method

2.2 Semantic feature

One semantic feature that would distinguish the meaning of these two verbs is [+/-effect]. In other words, though both are states that predicate of people, LEI has the additional meaning that is an effect state of an (unspecified) event, while PIJUAN does not specify this. It is obvious that an effect state occurs as a resultative complement, and represents the effect of another predicate. On the other hand, there seems to be a tendency against nominalized complex verbs in Chinese (e.g. all verb-resultative compounds cannot be nominalized). Thus, an effect state has the semantic implicature of a complex event and cannot be nominalized.

2.3 Prediction/Verification

After looking at near-synonyms to determine the semantic feature that differentiates them, we need to test our hypothesis. The following two examples demonstrate that it is much easier for LEI than for PIJUAN to occur with the perfective aspect marker (ASP) *-le*. The statistics shown in Table 2 indicate the relatively high percentage of LEI co-occurring with *-le* when compared with the zero utterance of PIJUAN.

(3) Perfective aspect marker

(3a) tamen lei le jiu lai ci he pijiu
they be-tired ASP then come here drink beer
'When they get tired, they come here to drink some beer.'

4. The abbreviations used in the glosses are the following: ASP 'aspect maker', BEI 'passive maker,' CL 'classifier,' PAR 'sentential-final particle.' Examples begin with a # are either unnatural or inacceptable.

(3b) # tamen pijuan le jiu lai ci he pijiu
they be-tired ASP then come here drink beer

Collocation	*-le*
LEI 174	38 (22%)
PIJUAN 33	--

Table 2. *Differences in collocations: LEI vs. PIJUAN*

According to Smith [1991], perfective *-le* appears only in dynamic sentences, presenting closed non-stative situations. When stative verbs occur with this morpheme, the sentences have only inchoative reading with focus on the initial point of the state. Thus, the collocation to *-le* reveals that the state expressed by LEI results in a change of state. In other words, LEI is an effective state, i.e. [+ effect]. PIJUAN, on the other hand, is a genuine state, i.e. [- effect].

In addition to the perfective aspect marker, LEI and PIJUAN also differ in the association with durational complements. While LEI takes a durational complement in 2% of the cases, PIJUAN never does.

(4) Durational complement

(4a) tamen lei le yi xiawu
they be-tired ASP one afternoon
'They have tired themselves all afternoon.'

(4b) # tamen pijuan le yi xiawu
they be-tired ASP one afternoon

Collocation	Durational Complement
LEI 174	4 (2%)
PIJUAN 33	--

Table 3. *Differences in collocations: LEI vs. PIJUAN*

As durational complements are used to locate an interval during which the prediction holds true [cf. Paris 1988], it is expected that there be endpoints in the state LEI. Again the feature [+/- effect] distinguishes the two state verbs in question.

3. Argument selection

The distributional differences for argument selection involve determining whether the verb occurs with an NP subject, VP subject, sentential subject, NP object, double NP

object, sentential object, etc.

3.1 Distributional differences

In the case of GAOXING and KUAILE 'be happy', GAOXING can take a sentential object in more than 7% of the cases, while KUAILE cannot, as shown in Table 4 and example (5).

Collocation	Sentential Object
GAOXING 280	20 (7.1%)
KUAILE 365	--

Table 4. *Differences in argument selection: GAOXING vs. KUAILE*

(5) Sentential Object

(5a) tamen hen gaoxing Zhangsan mei zou
they very be-happy John not go away
'They were glad that John did not go away.'

(5b) # tamen hen kuaile Zhangsan mei zou
they very be-happy John not go-away

3.2 Semantic feature

The semantic feature that can be deduced from this distributional difference is [+/- effect], where GAOXING is an effect state triggered off by the cause expressed in the sentential object.

3.3 Prediction/Verification

We observe from the data that only GAOXING can be associated with the perfective aspect marker -*le* in 0.7 % of the instances, as demonstrated below.

(6) Perfective aspect marker

(6a) keren gaoxing le jiu gei xiaofei
customer be-happy ASP then give tip
'When customers are pleased, they give tips.'

(6b) # keren kuaile le jiu gei xiaofei
customer be-happy ASP then give tip

Collocation	*-le*
GAOXING 280	2 (0.7%)
KUAILE 365	--

Table 5. *Differences in collocations: GAOXING vs. KUAILE*

The contrast between (6a) and (6b) is correctly predicted, because it is possible for GAOXING to represent a changed state brought out by some cause, but not for KUAILE. It is then justified to say that GAOXING is an effect state, i.e. [+ effect], whereas KUAILE is [- effect].

4. Aspectual types

The distributional difference for aspectual types involve looking at the aspect markers, aspectual adverbs and resultative complements the verbs co-occur with.

4.1 Distributional differences

In the case of QUAN and SHUIFU 'persuade', only QUAN occurs with the durative aspect marker *-zhe*[5] in 1.8% of the cases, SHUIFU never does.

Collocation	*-zhe*
QUAN 112	2 (1.8%)
SHUIFU 50	--

Table 6. *Differences in collocations: QUAN vs. SHUIFU*

(7) Durative aspect marker

(7a) ta yimian zou, yimian quan-zhe Zhangsan
he one-side walk one-side persuade ASP John
'He persuaded John as he walked.'

(7b) # ta yimian zou, yimian shuifu-zhe Zhangsan
he one-side walk one-side persuade ASP John

4.2 Semantic Feature

As the marker *-zhe* indicates that an event is on-going [cf. Li & Thompson 1981], the fact

5. Some authors consider *-zhe* as imperfective aspect marker [Ma 1985, Smith 1991].

that QUAN can take such a marker and SHUIFU never can suggests that there are aspectual differences between these two verbs. On the one hand, QUAN denotes an extensible, atelic event. On the other hand, SHUIFU denotes a bounded, telic event. The semantic feature that would distinguish the meaning of these two verbs is [+/- telic].

4.3 Prediction/Verification

If our hypothesis is correct, we expect that only QUAN is compatible with adverbs indicating the durative aspect. Consider the following examples.

(8) Durative aspectual adverb

(8a) ta yizhi quan Zhangsan jiehun
he all-the-time persuade John get-married
'All the time he persuaded John to get married.'

(8b) # ta yizhi shuifu Zhangsan jiehun
he all-the-time persuade John get-married

The adverb *yizhi* 'all the time' in the above examples can only occur with QUAN but not with SHUIFU. This means that only the event denoted by QUAN can be in progress. The difference between these two verbs in telicity is then justified.

A second argument in support of the claim that QUAN differs from SHUIFU in verbal aspect is related to the fact that only QUAN admits, in 3.6% of instances, resultative complements which indicate completion or termination [cf. Smith 1991]. Consider the examples in (9).

Collocation	Resultative Complement
QUAN 112	4 (3.6%)
SHUIFU 50	--

Table 7. *Differences in collocations: QUAN vs. SHUIFU*

(9) Resultative complement

(9a) ta quan de Zhangsan xin hen fan
he persuade DE John mood very be-bored
'He kept trying to persuade John until John was bored to death.'

(9b) # ta shuifu de Zhangsan xin hen fan
he persuade DE John mood very be-bored

It is reasonable that telic verbs like SHUIFU exclude the possibility of taking resultative complements, since we cannot terminate an event which is already terminated. But for atelic verbs like QUAN, it is natural that they take resultative complements, indicating that events are accomplished. Thus the feature [+/- telic] can account for the contrastive use of aspectual type between these two items.

5. Sentential types

In this section, we look at what type of sentences a verb can join, including passive sentence, imperative sentence, wish sentence, evaluative sentence, etc.

5.1 Distributional differences

One of the distributional contrasts between QUAN and SHUIFU involves the possibility of forming passive sentence. It seems that SHUIFU occurs more frequently in passive construction (6%) than QUAN does (0.9%). The examples in (10) show that QUAN is not allowed in the passive construction without a resultative complement.

Collocation	Passive Sentences
QUAN 112	1 (0.9%)
SHUIFU 50	3 (6%)

Table 8. *Differences in collocations: QUAN vs. SHUIFU*

(10) Passive sentence

(10a) # Zhangsan bei ta quan le
John BEI he persuade PAR

(10b) Zhangsan bei ta shuifu le
John BEI he persuade PAR
'John was persuaded by him.'

(10c) Zhangsan bei ta quan-zou le
he BEI he persuade go-away PAR
'John was persuaded to leave by him.'

In case of GAOXING and KUAILE 'be happy', the following distributional contrasts in terms of the sentential types are noticed from the Sinica Corpus: GAOXING never constitutes wish sentences but admits evaluational sentences (1.8%), while KUAILE occurs in wish sentences (2.2%) but never appears in evaluational sentences.

Collocation	Wish Sentences	Evaluational Sentences
GAOXING 280	--	5 (1.8%)
KUAILE 365	8 (2.2%)	--

Table 9. *Differences in collocations: GAOXING vs. KUAILE*

(11) Wish sentence

(11a) zhu ni kuaile!
wish you be-happy
'I wish you be happy.'

(11b) # zhu ni gaoxing!
wish you be-happy

(12) Evaluational sentences

(12a) zhei-jian shi zhide gaoxing.
this CL thing be-worth be-happy
'This thing is worth enjoying.

(12b) # zhei-jian shi zhide kuaile
this CL thing be-worth be-happy

5.2 Semantic Feature

The semantic feature that would distinguish the meaning of QUAN and SHUIFU is [+/-effect]. Though both are events, SHUIFU has an additional meaning of effect which corresponds to the affectedness property of passive sentences, while QUAN does not have.

As for GAOXING and KUAILE, the distinctive feature of their meaning is [+/-control]. Though both are states, only the controllable one can GAOXING express the calculated reaction in evaluational sentences and refuses the impredictive nature of wish sentences.

5.3 Prediction/Verification

We have seen in (9) above that it is possible for QUAN but not for SHUIFU to take a resultative complement. This collocational difference constitutes a good argument for the claim that the meaning of QUAN and SHUIFU can be distinguished by the feature of effect. One point needs to be clarified: why QUAN cannot occur in passive sentences

alone without a resultative complement behind? Given that resultative complements not only indicate the accomplishment of the main event, but also express the affected state of the participant, then, the use of such elements can contribute to QUAN additional properties like completion and affectedness, which are inherent to SHUIFU.

Now let us turn to the semantic feature [+/- control]. To support the claim that GAOXING can be controlled and KUAILE cannot, consider the use of imperative sentence illustrated below.

Collocation	Imperative Sentences
GAOXING 280	3 (1.1%)
KUAILE 365	--

Table 10. *Differences in collocations: GAOXING vs. KUAILE*

(13) Imperative sentence

(13a) bie gaoxing!
don't be-happy
'Don't be happy!'

(13b) # bie kuaile!
don't be-happy

The data show that GAOXING can form imperative sentences in 1.1% of the instances, while KUAILE never can. This means that the hearer can only change the state of GAOXING, but not the state of KUAILE. In other words, only the state of GAOXING is controllable.

6. Conclusion

The notion that the syntactic behavior of verbs is semantically determined has been examined extensively, especially for English verbs (please see Levin 1993 for relevant references). The technique that has been used quite productively is one that determines the distinctive behavior of verb classes. Levin summarizes this method:

> The assumption that the syntactic behavior of verbs is semantically determined gives rise to a powerful technique for investigating verb meaning that can be exploited in the development of a theory of lexical knowledge. If the distinctive behavior of verb classes with respect to diathesis alternations arises from their meaning, any class of verbs

> whose members pattern together with respect to diathesis alternation should be a semantically coherent class: its members should share at least some aspect of meaning. Once such a class is identified, its members can be examined to isolate the meaning components they have in common. Thus diathesis alternations can be used to provide a probe into the elements entering into the lexical representation of word meaning. [Levin 1993:14]

However, this technique is not easily implemented in Mandarin, because extensive study of diathesis alternations has not been done in Mandarin. Perhaps one reason is because Mandarin allows both subject and object omission, which means that it is very difficult to get a handle on what is a relevant 'alternation.' The work that has been done on semantic interpretations of syntactic structures (and the verbs that may occur in these structures) in Mandarin, such as in the case of pre-posed objects (such as BA and BEI), while interesting, is inconclusive because the wide variety of contexts and possible meanings defies a unified explanation. [Cf. Thompson 1973, Mei 1978, Bennett 1981, Ren 1991, Sun 1995, etc]

Moreover, the diathesis alternation technique does not allow for a very fine grained analysis of semantic features, because verbs may belong to more than one (seemingly unrelated) alternation class[6], and because different verb classes may share the same alternation[7]. Thus, it is difficult to extract the common semantic feature that predict the difference between the classes. When we look at near-synonyms, on the other hand, we are able to set up a controlled study of lexical semantic contrasts and their grammatical effects[8]. We hope that this fine-grained approach will aid us in identifying the semantic features or attributes that dictate the syntactic differences of verbs.

---

6. For example, according to Levin (1993), 'hit' belongs to verbs of throwing, verbs of contact by impact as well as verbs of existence, whereas 'cut' belongs to seven classes--verbs of cutting, verbs of separating and dissembling, verbs of creation and transformation, verbs of psychological state, verbs of bodily state and damage to the body, verbs of grooming and bodily care and meander verbs.

7. For example, 'hit' and 'cut' share the conative alternation.

8. Effectively, more larger scale experiment would be needed to deduce the semantic features for more verbs as well as to determine to what extent can this approach be generalized.

References

Bennett, P., "The Evolution of Passive and Disposal Sentences," *Journal of Chinese Linguistics* 9, 1981, pp. 61-89.

Dorr, B. J., and D. Jones, "Role of Word Sense Disambiguation in Lexical Acquisition: Predicting Semantics from Syntactic Cues," in *Proceedings of 16th International Conference on Computational Linguistics* (*COLING 96*), 1996, pp. 322-327.

Dowty, D. R., "Thematic Proto-Roles and Argument Selection," *Language* 67, 1991, pp. 547-619.

Jackendoff, R. S., "Towards an Explanatory Semantic Representation," *Linguistic Inquiry* 7, 1976, pp. 89-150.

Karttunen, L., *Radical Lexicalism*, CSLI-86-68.

Levin, B. (Ed.), *Lexical Semantics in Review*, *Lexicon Project Working Papers 1*, Center for Cognitive Science, MIT, 1985.

Levin, B., *English Verb Classes and Alternations: A Preliminary Investigation*, University of Chicago Press, 1993.

Liu, M., "Lexical Meaning and Discourse Patterning - The three Mandarin cases of 'build'," paper presented in the 3rd Conference on Conceptual Structure, Discourse, and Language, Boulder CO., 1997.

Lyons, J., *Linguistic Semantics: An Introduction*, Cambridge: Cambridge UP, 1995.

Ma, J.-H., *A Study of the Mandarin Chinese Verb Suffix ZHE*, Taipei: Crane Publishing. 1985.

Mei, K., "Ba Sentences in Mandarin Chinese," *Wen-shi-zhe Xuebao* 27, 1978, pp. 145-180. Taiwan University, Taipei.

Paris, M.-C., "Durational Complements and Verb Copying in Chinese," *Tsing Hua Journal of Chinese Studies New Series* 28.2, 1988, 423-439.

Pustejovsky, J., *The Generative Lexicon*, MIT Press, 1993.

Pustejovsky, J., S. Bergler, and P. Anick, "Lexical Semantic Techniques for Corpus Analysis," *Computational Linguistics* 19.2, 1993, pp. 331-358.

Ren, X., "The Post-Verbal Constituent in Chinese Passive Forms," *Journal of Chinese Linguistics* 19, 1991, 221-241.

Smith, C. S., *The Parameter of Aspect*, Kluwer Academic Publisher, 1991.

Sun, C., "Transitivity, the Ba Construction and Its History," *Journal of Chinese Linguistics* 23, 1995, pp. 159-195.

Teng, S.-H., *Chinese Synonyms Usage Dictionary*, Taipei: Crane Publishing, 1994.

Thompson, S. A., "Transitivity and the Ba Construction in Mandarin Chinese," *Journal of Chinese Linguistics* 15.1, 1973, pp. 208-221.

Tsai, M.-C., C.-R. Huang, and K.-J. Chen, "You jinyici bianyi biaozhun kan yuyi jufa zhi hudong. (From near-synonyms to the interaction between syntax and semantics)," in *Proceedings of 5th International Symposium on Chinese Languages and Linguistics* (*IsCLL 5*), 1996, pp. 167-180.

Acknowledgment

This paper is partially sponsered by National Science Council (87-2811-H-001-004). The authors would like to thank the CKIP group members for stimulating discussions on some of the verbal semantic features put forth here. We are also grateful to two anonymous reviewers for their valuable comments.