2008

MARVS Revisited: Operationalizing Sense Frequency and MI Values

Siaw Fong Chung

Kathleen Ahrens
Hong Kong Baptist University, ahrens@hkbu.edu.hk

Citation
Chung, Siaw Fong, and Kathleen Ahrens. "MARVS Revisited: Operationalizing Sense Frequency and MI Values." *Language and Linguistics* 9(2) (2008): 415-434.

MARVS Revisited: Operationalizing Sense Frequency and MI Values

Siaw Fong Chung
National Taiwan University
f91142002@ntu.edu.tw

Kathleen Ahrens
National Taiwan University
kathleenahrens@yahoo.com

Abstract

In MARVS (Module-Attribute Representation of Verbal Semantics), verbs can be differentiated based on eventive information which comprises event modules and role modules. Huang et al. (2000) used MARVS to examine near-synonyms and suggested that it is a model that can highlight the difference between synonymous sets. However, this paper found that there are weaknesses in the methodology of MARVS which can be improved by adding two additional steps. These steps include establishing the shared senses of the near-synonyms through corpus analyses and using collocations in terms of Mutual Information Values to operationalize the methodology. This paper demonstrates the effectiveness of these two steps and suggests that these steps should be adopted in MARVS.

Keywords: Near-synonyms, MARVS, sense, Mutual Information Value.

1.0 Introduction

A great number of semantic and computational models have been suggested and several have been used in distinguishing near-synonyms (such as Cruse (1986), Lyons (1995), Kay (1988), DiMarco, Hirst and Stede (1993), and Edmonds (1999)). These models range from the traditional descriptive approach to the quantitative calculation of similarities between synonyms. Recent comparisons of Chinese near-synonyms also utilize examples from corpora (such as Liu (2003) and Tsai et al. (1998)). This is also done in the Module-Attribute Representation of Verbal Semantics (MARVS) model developed by Huang et al. (2000. This model takes into consideration elements related to an event such as argument type, syntactic function, selectional restriction, etc. However, this model does not currently use quantitative analyses in to select argument types or to analyze meaning difference between near-synonyms.

This work aims to suggest two additional steps to MARVS in order operationalize the steps for verbal semantic analysis. The first step includes analyzing instances from corpus so as to establish the similarities and differences between near-synonyms. The second step suggests using Mutual Information (MI) values to look for argument types that collocate with the verbs. In addition, this paper also lays out the criteria for the selection of collocates using MI values, as the usual results from the calculation of MI values contain noises which have to be filtered out manually. This paper will address these issues in details through using the examples of *bai3* and *fang4*, which was discussed in Huang et al. (2000) as well as in the follow-up work of Ahrens, Huang and Chuang (2003).

2.0 Near-Synonyms: A Review of Methodologies

There are several ways to examine near-synonyms. One can analyze near-synonyms based on descriptive comparison or on quantitative analysis. Descriptive analyses usually depend on intuition with the aid of additional references such as dictionaries. Quantitative analyses, on the other hand, usually attempt to find out the differences between near-synonyms through comparing the behaviors of the synonymous pairs such as in comparing the argument types of the pairs attested.

Earlier work on synonyms tend to use descriptive approach, such as that demonstrated by Collinson (1939, cited in Harris, 1973: 14), which attempted to list the possible differences between synonyms using nine elements. These elements are general/specific applicability, intensity, emotion, moral approbation, professionalism, written/non-written, colloquialism, local/dialect and child talk. Today, the commonly agreed differences between synonyms are found within features such as connotations, implications, selectional restrictions and syntactic variations (by linguists such as Cruse 1986 and Lyons 1995 and lexicographers such as Kay 1988, DiMarco, Hirst

and Stede 1993, and Edmonds 1999). Cruse (1986:278-279) considered selectional and collocational restrictions the "main effect of presupposed semantic traits of a lexical item," which brings out the "syntagmatic companions" of words. Near-synonyms, according to Cruse (1986:267), are "lexical items whose senses are identical in respect of 'central' semantic traits, but differ…in 'minor' or 'peripheral traits'." In fact, most synonyms are near-synonyms by sharing certain central similarities and peripheral differences. Perfect synonyms are rare, as Taylor (2002:265) said: "perfect synonym is vanishingly rare, methodologically proscribed, or a logical impossible" (Taylor, 2002:265).

Later work on distinguishing near-synonyms used both descriptive as well as quantitative analyses. For instance, Taylor's (2002) analysis of 'tall' and 'high' was carried out through psycholinguistic experimentation (acceptability rating tasks) in addition to descriptive analysis of the two adjectives. Taylor claimed that 'tall' and 'high' can be differentiated using MacLaury's (1997, 2002) Vantage theory, which distinguish near-synonyms in terms of dominant/recessive meanings. For both these adjectives, the dominant meaning emphasizes on the similarity of "a fixed landmark which is the human body sanctions the application of the word to a limited range of prominently upright entities." 'Tall,' however, has restriction (i.e., recessive meaning) on dimensional uses whereas 'high' has restriction on positional uses.

More statistical approaches to near-synonyms can be seen in the computational field. For example, a statistical analysis of near-synonym by Church et al. (1994) used Mutual Information (MI) as well as substitutability in T-scores to differentiate between the near-synonyms 'request' and 'ask for.' MI value measures the co-occurrences of a word neighboring terms so as to determine whether a word is a constant collocate to another word. They found twenty-eight significant objects that collocate with both 'request' and 'ask for,' among which are 'aid,' 'assistance,' 'copy,' 'dismissal' and 'extension.' As for substitutability, 'request' is found to have the highest substitutability value than 'ask for' when substituted by words such as 'seek,' 'grant' and 'demand.'

Also important in the computational approaches to near-synonyms is that model suggested by Pustejovsky (1991). In this model, the meanings of the verbs can be generated from the nominals surrounding the verbs by examining the Qualia structures of the verbs, i.e., the structure of the nominals are co-compostitional by four types of role (Pustejovsky 1991: 426-427).

a) Constitutive role (the relation between an object and its proper parts such as 'narrative' for a novel)
b) Formal role (role that distinguishes the object within a larger domain such as 'book' or 'disk' for a novel)
c) Telic role (the purpose and function of the object such as 'read' for a novel)
d) Agentive role (factors involved in bring about the object such as 'artifact' or 'write' for a novel)

Based on these four roles, the author claims that the load of distinguishing verb meanings can be distributed to the nominals (or adjectives) surrounding the verbs. This model has later been used in various linguistic theories. The strength of this model comes from its prediction of the possibilities to distinguish a term from another based on their 'part-of-relation' (constitutive role), 'kind-of-relation' (formal role), 'function relation' (telic role) and 'origin relation' (agentive role) (also suggested in Croft and Cruse, 2004:137). When near-synonyms are concerned, the peripheral differences between a synonymous pairs can occur at any of these four aspects. In other words, these four aspects provide an alternative way of stating the differences between a synonymous set, in addition to stating the differences in semantic features such that was done in traditional semantics (i.e., [± female], [±animate], etc.) or that in the work of Collinson (1939).

Compared to the model of Pustejovsky (1991) as well as using the semantic features, the approach suggested by Church et al. (1994) has the advantage of using corpus data, which help reduce subjectivity in the analysis. MARVS has the advantage of combining the quantitative approaches and the semantic features used in traditional semantics.

3.0 MARVS

MARVS provides a way to state the peripheral differences between synonyms is through identifying the eventive information that differs between the synonymous set attested. It lays out this eventive information in terms of event modules and role modules (See Figure 1 below taken from Huang et al., 2000:24).

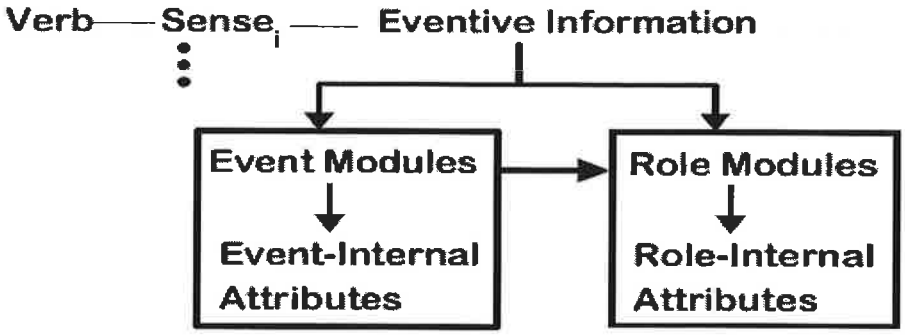

Figure 1 *Module-Attribute Representation*

Under each module, there are attributes which further define the behaviors of the module. Some examples of role-internal attributes are [sentience], [volition], [affectedness] and [design]. Examples of event-internal attributes are such as [control], [effect].

In Huang et al. (2000), two verbs of 'put' in Chinese (*bai3* and *fang4*) are found to differ at the [design] of the role-internal attribute because the way of putting is different in the two verbs, with *bai3* entails the act of "putting following a certain plan" as well "a resultant state" but not in *fang4*.

> *..since the plan which the putting action follows entails a resultant state to be attained, bai3 can take a resultant object while fang4 has no such entailment and cannot take such an object* (Huang et al. 2000:35).

The following diagram shows the differences between *bai3* and *fang4* (Huang et al. 2000:36).

(1) MARVS for *bai3* and *fang4*

bai3 •____ <Agent, Theme, Location>
|
[design]

fang4 •____ <Agent, Theme, Location>

The methodology of MARVS is corpus-based and it also contrasts the differentiation between the near-synonyms in terms of attributes (which can be said as a type of descriptive element). Therefore, it has the combination of both the descriptive approach and the corpus-based approach. However, as suggested in the next subsection, there are weaknesses of MARVS that can be strengthened to make the model more complete.

4.0 Weaknesses of MARVS

Scholars such as Tognini-Bonelli (2001) distinguished between corpus-based and corpus-driven analyses. Corpus-based analysis uses corpus as resources of examples for verifying intuition. Corpus-driven analysis, on the other hand, allows the discovery of new sentence patterns for the purpose of research. MARVS is one that is corpus-based. Its weakness, therefore, can be attacked from its selective use of sentences from corpus. This is similarly carried out in Ahrens, Huang and Chuang (2003), which is an extension of the work by Huang et al. (2000).

In Ahrens et al. (2003) suggested that the different meanings of the English 'set' and the Chinese *bai3* can be represented in MARVS. However, they only took examples of sentences from the corpus, which is a data-driven methodology.

The original steps of MARVS were provided by Ahrens et al. (2003: 470).

> *How do we determine these collateral differences? First, we examine these near-synonym pairs by first combing a corpus for all relevant examples of the words in question. These examples are then categorized according to their syntactic function.* ***Third, each instance is classified into its argument-structure type****. Fourth, the aspectual type associated with each verb is determined, and fifth, the sentential type for each verb is also determined.*

The underlined step above shows that the study did not collect and analyze sample sentences from the corpus but only extracted relevant examples needed. By analyzing the meanings of sentences for the synonymous pairs, one can obtain information such as a) the meaning shared by the pairs (i.e., the 'central semantic traits') and b) the differences in meanings between the pairs (i.e., the 'peripheral traits').

The second weakness of MARVS is that, if corpora are accessible, the model should utilize information such as collocations in terms of MI values. This information is obtainable from the Sinica corpus for the analysis of the Chinese synonyms as well as in the British National Corpus for the analysis of 'set' in English. By adding this information, one can reduce manual work in generating the argument types for the synonyms (as was bolded in the quotation of methodology by Ahrens et al. (2003) above).

Finally, the most important weakness of a semantic model is the arbitrariness of the attributes. However, since this is also a problem for traditional semantics as well as in most feature-identifying models, this weakness will not be discussed extensively in this work. Based on the two weaknesses that were outlined above, this study suggests two additional steps to the original methodology of MARVS. This paper also

re-analyzes *bai3* and *fang4* so as to show the additional steps are needed to operationalize the comparison of near-synonyms.

5.0 Reanalysis of *bai3* and *fang4*

The revised steps (with inclusion of two steps and the modification of the first step) are given in (2) below. The additional two steps are in italic bold face.

(2) First, examine the near-synonym pairs by analyzing **at least the first 100 examples from the corpus**.

Second, analyze the senses either according to intuition or the meanings in WordNet so that the similarities (i.e., the pair is near-synonyms) and differences of sense can be identified.

Third, categorize the examples according to their syntactic function.

Fourth, classify its argument-structure type based on their collocation restrictions discovered through MI values.

Fifth, determine the aspectual type associated with each verb.

Sixth, determine the sentential type for each verb.

First, we suggest that a sample of sentences have to be collected from the corpus. The number of examples should be consistent for all the synonymous set attested. In the second step, we suggest that these examples are then analyzed manually or by using a reference such as WordNet. The purpose of this is to find out the similarities and differences in meanings between the synonymous set. This step is also important in that it proves that the items in the synonymous set are indeed synonyms, i.e., they share at least one similarity in meaning despite the other differences. The fourth step suggests that collocations and MI values can be used as criteria to determine the arguments of the synonyms. As Palmer (2000) said, consistent concrete criteria have to be stated clearly for discovering sense distinction; the aim of adding these steps is to make the model (MARVS) more operationalized and applicable to other verbs. The followings will take *bai3* and *fang4* as an example and demonstrate how these two additional steps can be conducted.

5.1 Sense Analyses

First, to prove the synonymous meaning, the sentences from the corpus (the Sinica Corpus) are analyzing according to different senses. There are two ways of doing this: one is through grouping the data in the Sinica Corpus according to the senses provided in WordNet 1.6 (through the Sinica Bow, Huang et al. 2004); the other is to assign different meanings to the sentences and group them according to these meanings. This paper takes the second option because the definitions from WordNet 1.6 do not help with our analysis. For example, the senses of *bai3* in 3(a) below are too narrow (that only the meaning of 'arrange' is derived) whereas the senses of *fang4* (3b) are too general (that 'put into a certain place' comprises 'put,' 'lay,' 'position,' 'pose' and 'place.')

(3) (a) *bai3*: 1: arrange thoughts, ideas, temporal events, etc.
2: an apparatus consisting of an object mounted so that it swings freely under the influence of gravity

(b) *fang4* 1: discharge or direct or be discharged or directed as if in a continuous stream
2: put into a certain place: "Put your things here"
3: locate

In addition, in examples (4) and (5) below, the use of *bai3* in 4(a) is not the same as in 4(b), and both not easily represented using the meanings in 3(a). Similarly, example 5(a) is a meaning extension of *fang4* but it is not related to the real sense of 'locate' or 'put in a certain place.' Since MARVS only chooses the relevant sentences from the corpus, some other senses of *bai3* and *fang4* could have been left out. However, through collection sample sentences from the corpus, this problem is necessarily addressed, as all examples (not a selected few) appearing in the sample collected have to be dealt with .

(4) (a) *bai3 ge zi1 shi4* 擺個姿勢 'to pose'
(b) *bai3 qi2zi3* 擺棋子 'to lay a piece in a board game'

(5) (a) *fang4 zhe feng1zheng1* 放著風箏 'to fly kite'
(b) *fang4 yi3zi3* 放椅子 'to put a chair'

In order to collect a sample of sentences, this paper takes the first 100 sentences for each verb from the Sinica Corpus (from the total of 233 for *bai3* and 1021 for *fang4*). The results are shown in Table 1.

Table 1: Sense Analysis for *Bai3* and *Fang4*[1]

bai3	**%**	***fang4***	**%**
metaphor	27	Metaphors	43
arrange for display	24	**put (things)**	**21**
put (things)	**12**	let go (animals, person, hand, prey, etc.)	13
lay (baby, basin, book, dishes, garden, etc.)	12	discharge (bomb, fire, firework, kite, etc.)	8
set up	12	non-classified	6
pose	5	keep (meat, tea leaves, things)	3
move	4	play (record, music, etc.)	3
non-classified	4	add	2
		locate (building)	1
Total	**100**	**Total**	**100**

From Table 1, there is an overlapped meaning of 'put' that appear for both *bai3* and *fang4*. Examples of this sense can be seen in (6) below.

(7) (a) 錢就擺在房間某件東西裡面
'the money is put inside something (a container) in the room'

(b) 杜象把這個作品放在一個木箱裡
'Duxiang put this piece of arts inside a wooden case'

As one can see from (7), the use of *bai3* and *fang4* in (7) can be substituted with one another. This overlapped meaning shows that the pair *bai3* and *fang3* is near-synonymous. The analysis in Table 1 also singled out metaphorical expressions such as in (8) below.

(8) (a) 把全民利益擺在第一
'to put the interests of the whole nation at priority'

(b) 霸著話題不放
'to dominate the topic of discussion (without letting go)'

[1] Since only 100 examples were analyzed, the percentages also reflect the number of instances found for each sense.

These metaphorical uses were excluded for this analysis because they will create noise in the data, if they were included as part of the other meanings. The "non-classified" meaning refers to instances where *bai3* and *fang4* are either used as nouns, as in 7(a), or when the immediate contexts of the keywords are ambiguous or incomplete in meanings, as in 7(b).

(7) (a) 「放的哲學」
"The Philosophy of Fang4"

(b) ?這家店過年過節有送禮的，所以可以擺。
'? This shop gives away gifts during festival, so (one) can place (something) there"

Finally, the results in Table 1 also show that *bai3* and *fang4* are similar in one meaning but they differ in many other meanings. These differing meanings give clues as to how one synonym differs from one another. In the case above, one can find out which senses are found in one of the synonyms but not the in the others. In the next section, we will demonstrate the use of MI values to find the argument types for the synonyms.

5.2 Mutual Information Value

In order to find out the MI values for the arguments that collocate most frequently with each verb, the MI values for all the search results (233 for *bai3* and 1031 for *fang4*) were calculated by the internal system of the Sinica Corpus. The window size is set from -4 to 4 (i.e., 4 words on the left or right of the key word). The MI list shown has several columns, as shown in example (8) below.

(8) Examples of MI values for *bai3*

	MI	freq(y)	freq(x,y)	y:詞/詞類
(a)	10.126	1	1	缸數(Na)
(b)	10.126	1	1	咬鳥卦(Na)
(c)	**9.839**	**4**	**3**	**扭腰(VA)**
(d)	9.433	2	1	花椒(Na)
(e)	**9.279**	**7**	**3**	**炮竹(Na)**
(f)	2.380	11562	5	的(T)

Freq (y) is the number of times the words on the rightmost column appear in the whole corpus (including texts other than *bai3*). Freq (x,y) refers to the number of times the words y co-occur with the target word (x= *bai3*). MI values refer to the probability of the words y collocate with x (cf. Church and Hanks, 1990). For Sinica Corpus, the definition of the MI value is the calculation "between a key and the characters occurring in the specified window (i.e., the left and/or right context)" in the Sinica corpus (Huang, Ahrens and Chen, 1998: 157). The MI values calculated by the Sinica corpus is as in (9) below, "where **N** is the size of the corpus and **m** is the size of the selected window" (Huang et al., 1998: 157):

(9)

$$I(x, y) = \text{Log}_2 P(x,y) / P(x) \cdot P(y)$$

$$\approx \text{Log}_2 \frac{f(x,y) / m \cdot N}{f(x)/N \cdot f(y)/N}$$

$$= \text{Log}_2 f(x,y) \cdot N / m \cdot f(x) \cdot f(y)$$

Even though MI values are indices showing whether x and y are associated, this paper suggests that one should not refer to MI values per se. This is to avoid including data that we do not need. For instance, 8(a), when both x and y occurs once respectively, the probability of the two co-occurring together will be absolute and the MI value will be high. In order to avoid these examples, this paper sets two criteria for choosing the collocated arguments for the verbs (x). These two criteria are: a) the freq (x, y) must be higher than 3 (i.e., the x and y co-occur at least three times in the whole corpus of *bai3*); and b) the MI value should not be lower than 5. These threshold levels were set based on our observation of the data.

These two criteria can help avoiding selecting a common term such as *de* 的 (example 8(f)) which occurs so often in the whole corpus that the MI value becomes very low (even though the number of times it co-occurs with *bai3* is more then 5). Based on these criteria, the final selected arguments for *bai3* are shown in (10) below.

(10) Collocated Arguments for *bai3*

MI	freq(y)	freq(x,y)	y:詞/詞類
9.839	4	3	扭腰(VA)
9.279	7	3	炮竹(Na)
8.072	39	5	地攤(Nc)
8.006	25	3	平(VC)
7.511	41	3	書架(Na)
7.487	56	4	桌(Nf)
6.991	184	8	桌(Na)
6.991	92	4	桌子(Na)
6.507	112	3	姿勢(Na)
5.932	199	3	門口(Nc)
5.881	279	4	中間(Ncd)
5.680	256	3	左(Ncd)
5.627	270	3	右(Ncd)
5.605	644	7	起(Di)
5.393	341	3	東(Ncd)
5.282	381	3	西(Ncd)
5.088	1080	7	往(P)

Compared to (9), one can see that these two criteria remove items lexical items such as 缸數 'the number of tubs' 咬鳥卦 'a type of fortune-telling card picked up by a bird.' Items such as these two have absolute association with *bai3* as the only time they appear in the corpus co-occur with *bai3*. Using the two criteria set above, this paper filtered out unwanted examples and the list in (10) shows more plausible arguments of *bai3*. When the same criteria applied to *fang4*, the results in (11) were obtained.

(11) Collocated Arguments for *fang4*

MI	freq(y)	freq(x,y)	y:詞/詞類
8.080	7	4	長假(Na)
7.851	11	5	水燈(Na)
7.828	9	4	倉(Na)
7.541	9	3	成交價(Na)
7.486	19	6	在一塊(VH)
7.423	27	8	假(Na)
7.173	13	3	紅龜(Na)
6.991	26	5	四海(Nc)
6.905	17	3	盆(Na)
6.815	31	5	架子(Na)
6.711	55	8	風箏(Na)
6.560	24	3	心念(Na)
6.519	25	3	武松(Nb)
6.480	52	6	人質(Na)
6.283	116	11	重心(Na)
6.257	184	17	桌(Na)
6.228	78	7	口袋(Na)
6.024	41	3	書架(Na)
5.905	77	5	心思(Na)
5.846	49	3	炸彈(Na)
5.767	53	3	枕頭(Na)
5.727	92	5	桌子(Na)
5.565	238	11	火(Na)
5.560	87	4	敢(Nf)
5.489	70	3	肩膀(Na)
5.461	96	4	化妝品(Na)
5.412	126	5	抓住(VC)
5.376	183	7	羊(Na)
5.362	106	4	客廳(Nc)
5.362	106	4	牛奶(Na)
5.343	135	5	心力(Na)
5.326	577	21	重點(Na)
5.173	96	3	浴室(Nc)
5.113	102	3	包袱(Na)
5.034	184	5	封(Nf)
5.022	484	13	下(VC)

Comparing the lists in (10) and (11), one sees that more argument types were found for *fang4*. This is another advantage of using the criteria, as one can compare the productivity of two near-synonyms in terms of their argument-taking behavior. By comparing (10) and (11), one can see that *fang4* is more productive than *bai3* in terms of the types of argument it takes. The comparison is made more obvious by laying out the collocated arguments for the two verbs, as shown in Table 2 below..

Table 2: Collocated Arguments for *bai3* and *fang4*

Bai3		*Fang4*			
扭腰(VA)	門口(Nc)	長假(Na)	架子(Na)	心思(Na)	羊(Na)
炮竹(Na)	***中間(Ncd)***	水燈(Na)	風箏(Na)	炸彈(Na)	客廳(Nc)
地攤(Nc)	***左(Ncd)***	倉(Na)	心念(Na)	枕頭(Na)	牛奶(Na)
平(VC)	***右(Ncd)***	成交價(Na)	武松(Nb)	桌子(Na)	心力(Na)
書架(Na)	起(Di)	在一塊(VH)	人質(Na)	火(Na)	重點(Na)
桌(Nf)	***東(Ncd)***	假(Na)	重心(Na)	畝(Nf)	浴室(Nc)
桌(Na)	***西(Ncd)***	紅龜(Na)	桌(Na)	肩膀(Na)	包袱(Na)
桌子(Na)	往(P)	四海(Nc)	口袋(Na)	化妝品(Na)	封(Nf)
姿勢(Na)		盆(Na)	書架(Na)	抓住(VC)	下(VC)

By identifying the selectional restriction through this way one can verify the following statement by Huang et al. (2000:35) that "the orientation of the placed object [of *bai3*] can be specified while only location can be specified for *fang4*." This is seen in Table 2 above for orientations of *zhong1 jian1* 中間 'middle,' *zuo3* 左 'left,' *you4* 'right,' etc. (highlighted in Table 2), all of which are not found in the list for *fang4*. When carried out using these steps, one then make more data-driven proposals within the MARVS model.

4.0 Conclusion

Two additional steps are suggested and the analysis of *bai3* and *fang4* in this work will help strengthen the applicability of MARVS to other verbs. The addition of these two steps are advantaged by a) being able to operationalize the steps to identify contrasts in near-synonymy; and b) using quantitative data (frequency and MI values) to state the differences between two verbs. Furthermore, no study in literature has yet stated the criteria clearly so that the more precise MI values can be selected. Thus, this study not only contributes methodology-wise to computational linguistic research but also provides clarification to a previously established model.

For future research, the use of WordSketchin finding out the collocation of grammatical relations will also be possible (Kilgarriff and Tugwell, 2001). In addition,

we suggest extending the methodology discussed herein to other synonymous pairs in Chinese as well as in other languages. If the methodology can be applied to other languages as well, it will also prove the workability of the model.

Acknowledgements

We would like to acknowledge the NSC project grant to Kathleen Ahrens #94-2411-H-002-038 and the 深耕 project grant "Lexicon-driven Ontology and Conceptual Structure" to Professor Chu-Ren Huang for supporting the discussion herein. We also thank Professor Chu-Ren Huang for his comments on this work.

References

Kathleen Ahrens, Huang Chu-Ren and Shirley Chuang. 2003. "Sense and Meaning Facets in Verbal Semantics: A MARVS Perspective." *Language and Linguistics*. 4(3). Taipei: Academic Sinica. pp. 468-484.

Church, Kenneth Ward, William Gale, Patrick Hanks, Donald Hindle, and Rosemund Moon. 1994. "Lexical Substitutability." In B. T. S. Atkins, and A. Zampolli (Eds.). *Computational Approaches to the Lexicon*. pp. 153-177.

Collinson, W. E. 1939. "Comparative Synonymics." *Transactions of the Philosophical Society*. pp. 54-77.

Cruse, D. A. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.

DiMarco, Chrysanne, Graeme Hirst and Mandred Stede. 1993. "The Semantic and Stylistic Differentiation of Synonyms and Near-Synonyms." In *Proceedings of the AAAI Spring Symposium on Building Lexicons for Machine Translation*. pp. 114-121.

Edmonds, Philip. 1999. *Semantic Representations of Near-Synonyms for Automatic Lexical Choice*. Ph.D thesis, Department of Computer Science, University of Toronto. Available at http://www.cs.toronto.edu/compling/Publications/ Abstracts/ Theses/EdmondsPhD-thabs.html.

Harris, Roy. 1973. *Synonym and Linguistic Analysis*. Oxford: Basil Blackwell.

Huang, Chu-Ren, Kathleen Ahrens, Chang Li-li, Chen Keh-jiann, Liu Mei-chun, and Tsai Mei-Chih. 2000. "The Module-Attribute Representation of Verbal Semantics: From Semantics to Argument Structure." In *Biq* (Ed.) *Special Issue on Chinese Verbal Semantics. Computational Linguistics and Chinese Language Processing*. Vol.5.1.19-46.

Huang, Chu-Ren, Kathleen Ahrens, and Keh-jiann Chen. 1998. A Data-driven Approach to the Mental Lexicon: Two Studies on Chinese Corpus Linguistics. *Bulletin of the Institute of History and Philology*. 69.1.151-179.

Kay, Mairé Weir (Ed). 1976. *Webster's Collegiate Thesaurus*. Springfield, Mass.: G. & C Mariam Co.

Kilgarriff, Adam and Tugwell, David. 2001. "WORD SKETCH: Extraction and Display of Significant Collocations for Lexicography". In the *Proceedings of the ACL Workshop COLLOCATION: Computational Extraction, Analysis and Exploitation*. Toulouse. pp. 32-38.

Liu, Mei-chun. 2003. "From Collocation to Event Information: The Case of Mandarin Verbs of Discussion." *Language and Linguistics*. 4(3). pp. 563-585.

Lyons, John. 1995. *Linguistic Semantics: An Introduction*. Cambridge University Press.

MacLaury, Robert E., 1997b. Vantage theory in cognitive science. In: Ramscar, M. (Ed.), Proceedings of the Interdisciplinary Workshop on Similarity and Contrast.

Dept. of Artificial Intelligence, University of Edinburgh, pp. 157–163.

MacLaury, Robert E., 2002. Introducing vantage theory. Language Sciences 24, 493–536.

Palmer, Martha. 2000. Consistent Criteria for Sense Distinctions. *Computers and the Humanities*, 34. 217-222.

Pustejovsky, James. 1991. "The Generative Lexicon." *Computational Linguistics*. 17 (4). pp. 409-441.

Taylor, John R. 2002. "Near Synonyms as Co-extensive Categories: 'High' and 'Tall' Revisted. *Language Sciences*. 25. pp. 263-284.

Tognini-Bonelli, Elena. 2001. Corpus Linguistics at Work. (Studies in Corpus Linguistics: 6). Amsterdam/Atlanta, GA: John Benjamins.

Tsai, Mei-Chih, Chu-Ren Huang, Keh-Jian Chen and Kathleen Ahrens. 1998. "Towards a Representation of Verbal Semantics: An Approach Based on Near-Synonyms." In *Proceedings of the tenth Conference on Computational Linguistics and Speech Processing (ROCLING X)*. pp. 34-48.