Event Selection and Coercion of Two Verbs of Ingestion: A MARVS perspective

Jia-Fei Hong

Graduate Institute of Linguistics National Taiwan University

No. 1, Sec. 4, Roosevelt Road 106, Taipei, Taiwan R.O.C

d95142001@ntu.edu.tw

Chu-Ren Huang

Institute of Linguistics Academia Sinica

No. 128, Section 2, Academia Road 115, Taipei, Taiwan R.O.C

churen@gate.sinica.edu.tw

Kathleen Ahrens

Graduate Institute of Linguistics National Taiwan University

No. 1, Sec. 4, Roosevelt Road 106, Taipei, Taiwan R.O.C

kathleenahrens@yahoo.com

Abstract

Event semantics in general and event type coercion in particular have been a challenging yet rewarding topic in verbal semantics (Pustejovsky, 1995). However, there have been few corpus-based empirical accounts discussing the range of event type coercions based on the lexical meanings of the verbs. In this paper, we explore the possible types of event coercions for two verbs of ingestion in Mandarin Chinese. In particular, we will show that different types of coercions can be predicted by the bifurcation of event-internal and role-internal attributes proposed in the MARVS theory (Huang et al, 2000). Data examined are taken from the Chinese Gigaword Corpus (LDC 2005) and accessed through Chinese Word Sketch (Huang et al, 2005).

1 Introduction

In this paper, we will explore the possible types of event coercions for two verbs of ingestion *chi1* (to eat) and *he1* (to drink) in Mandarin Chinese. We select our data from the Chinese Gigaword Corpus through Chinese Word Sketch Engine in order to confirm and refine the argument role and lexical semantic information for these two verbs.

In addition, we will use the MARVS theory to explain and examine the different event types and coercions of these two verbs of ingestion. When we further explore the same event with these verbs of ingestion, we can gain understanding into different event types.

The next section introduces the GigaWord Corpus and Word Sketch Engine. We will then present background of the MARVS theory. Thirdly, we show empirical data such as the manual sense analysis from the Chinese Wordnet Group (CWN Group). Fourthly, we analyze and discuss the possible types of event coercions for two verbs of ingestion in Mandarin Chinese. Finally, we conclude with predictions and future work.

2 Chinese Word Sketch

Explanations of Gigaword Corpus and Chinese Word Sketch (CWS) can be found in Kilgarriff et al. (2005), Huang et al. (2005), Ma and Huang (2006) and Hong and Huang (2006).

The database for CWS is collected from Chinese Gigaword Corpus, which contains about 1.1 billion Chinese characters, including more than 700 million characters from Taiwan's Central News Agency, and nearly 400 million characters from China's Xinhua News Agency. The segmentation and tagging was performed automatically with automatic and partially manual post-checking. The precision accuracy is estimated to be over 96.5% (Ma and Huang 2006).

The Sketch Engine (SKE, also known as the Word Sketch Engine) is a novel Corpus Query System incorporating word sketches, grammatical relations, and a distributional thesaurus.

The advantage of using Sketch Engine as a query tool is that it pays attention to the grammatical context of a word. In order to show the cross-lingual robustness of Sketch Engine, as well as to propose a powerful tool for collocation extraction based on a large scale corpus with minimal pre-processing, we constructed Chinese Word Sketch Engine (CWS) by loading the Chinese Gigaword to the Sketch Engine (Kilgarriff et al., 2005). All components of the Sketch Engine were implemented, including *Concordance, Word Sketch, Thesaurus and Sketch Difference*.

3 MARVS Theory

The MARVS theory is the Module-Attribute Representation of Verbal Semantics which is a theory of the representation of verbal semantics that is based on Mandarin Chinese data (Huang et al. 2000). In MARVS, there are two different types of modules: event structure modules and role modules, as well as two sets of attributes: event-internal attributes and role-internal attributes which are linked to the event structure module and role module respectively, (Figure 1).

In the Module-Attribute Representation of Verbal Semantics (MARVS) theory, lexical knowledge is classified into two types: structural information is represented by means of the composition of atomic modules while content information is represented by means of attributes attached to these modules. In addition, the roles that participate in the event are represented in the Role Modules. So, the semantic attributes pertaining to the whole event are called the *Event-Internal Attributes*. Moreover, the semantic attributes pertaining to each role are termed *Role-Internal Attributes*.

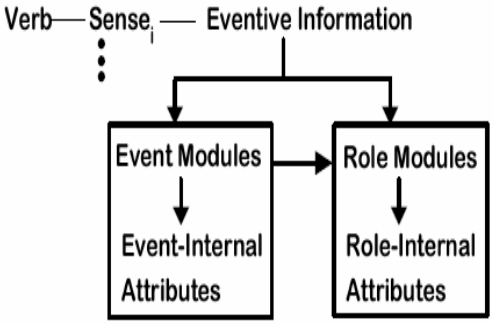


Figure 1: Module-Attribute Representation

In this paper, we will use this theory to explain event selection and coercion of two verbs of ingestion *chi1* and *he1* and define their different event types.

4 Motivation and Goals

Ahrens et al. (2003) examines this theory in light of English data and Mandarin Chinese data. In their paper, they look at the near synonym contrast of the verbs “put” and “set” based on data from the sampler of the British National Corpus. Moreover, they also examine the distributional differences that exist for *bai3* and *fang4* in Mandarin Chinese.

The English verbs “put” and “set” seems synonymous and interchangeable in most contexts such as:

(1a) Put/set the book on the table.
(1b) He set/put the pin on the cushion.
(2a) ta zhengzai fang/*bai shu zai zhuo-shang
s/he DURATIVE put/set book at table-top
(2b) She is putting/setting the book on the table.

From this data analysis, they find out the near synonym pair “put” and “set” or *bai3* and *fang4* have near complimentary distributions which clearly indicate their semantic contrasts.

In their study, Ahrens et al. (2003) regarded “put” as the roles of Agent, Theme, and Location in its event structure with location having a role-internal attribute of [+bounded]. “Set” has four senses: one that involves the roles of Agent, Theme and Location with a role-internal attribute of [+position] on the Theme, a second one which involves roles of Agent, Theme, and Location with a role-internal attribute of [+position], a third one which has three roles of Agent, Theme, and Predicate and a fourth one which has two roles, Agent and Theme, with a role-internal attribute of [+restricted] on the theme.

In this paper, we will use the MARVS theory to explore event Selection and Coercion of two verbs of ingestion *chi1* (to eat) and *he1* (to drink). What is interesting for *chi1* and *he1* is that they have overlapping meaning where one is an elaboration of the other, called troponym by WordNet; in addition, they have very rich semantic extensions. We take MARVS to explain the possibility of different event module-attribute representations of the complex meaning extensions based on automatically analyzed distributional data from CWS.

5 Data Collection and analysis

We collect the data from the smaller (5 million words) Sinica Corpus which shows that the dominant sense of *chi1* is to ingest solid food, while *he1* is to ingest liquid. We also check related data from the Gigaword Corpus with Chinese Word Sketch. Based on these data and detailed manual analysis, the sense inventory of *chi1* and *he1* are given below (Huang et al. 2006) in Example 1 and Example 2.

5.1 Sense Analysis

吃→chi1→ㄔ

詞義1：[及物動詞，VC] 使食物經過口中吞入體內。同義詞「用(0400)」。{eat, 00794578V，下位詞}

□例句：一隻猴子會分辨什麼果子能<吃>，什麼果子不能吃，這屬於本能。

□例句：沒落期的羅馬人認為，人生最快樂的事情就是<吃>東西。譬如，吃烤乳豬的那一剎那最快樂。

□例句：每天睡到下午兩點起床，之後就約一些朋友<吃>晚飯、談生意，一直到半夜兩、三點才回家。

□例句：其實儒家本身符合人性，也無所謂教條，而「禮教<吃>人」根本是政治上在利用人性的弱點。

Example 1: Original sense of *chi1*

喝 1 → he1→ㄏㄜ

詞義1：[及物動詞，VC；名詞，nom] 將液體或糊狀的食物經由口中進入體內並吸收養分。{drink,00795711V}

□義面1：[及物動詞，VC] 將液體或糊狀的食物經由口中進入體內並吸收養分。{drink,00795711V}

□□例句：好好吃一頓飯，歡喜<喝>一碗茶；一日喜樂無惱，一夜安眠無夢，又是價值多少？

□□例句：想成宵夜的人，也可以試試餓的時候吃水果、或是<喝>麥片也行，因為稠稠的，所以會有飽足感。

□□例句：上級要徵糧，公社幹部便將應用作社員口糧的糧食拿去湊數，公社食堂當然就只能給社員<喝>稀飯了。

□□例句：人們在端午節這天，都要在門上插蒲艾，貼神符，在身上帶香包，並且要<喝>雄黃酒。為的是消災除病，驅邪迎福。

Example 2: Original sense of *he1*

Our data analysis focused on idioms of the two verbs of ingestion *chi1* and *he1* from CWS for this study such as:

	More usage for「chi1」	Common usage	More usage for「he1」
object	yao4 (medicine)、dong1 xi1 (foodstuff)、shi2 wu4 (foodstuff)…	xi1 fan4 (porridge)、xi3 jiu3 (wedding banquet) 、nai3 shui3 (milk)、leng3 yin3 (cooling drink)…	jiu3 (wine)、cha2 (tea)、ku3 shui3 (complaints)…

Table 1: The common patterns for "chi1" and "he1"

The above data collection and analysis in Table 1 suggests that one clear clue for a meaning extended or metaphorical use is the violation of selectional restrictions. This is, in turn, facts that should be extendable from CWS.

5.2 Neutralized Selectional Restriction

One set of challenging facts for selectional restrictions involves cases where they are neutralized. For example, when an object has both solid and liquid attributes, an object will be selected both by "chi1" and "he1" such as below (Hong and Huang, 2006):

(1) *chi1 xi1 fan4*
to eat porridge
(2) *he1 xi1 fan4*
to drink porridge

These neutralization effects can also be found with metaphoric uses. For instance, both wedding banquet (*xi3 jiu3*) and afternoon tea time (*xia4 wu3 cha2*), can be selected by both verbs *chi1* and *he1*.

Therefore, the widely shared intuition that the two verbs of ingestion *chi1* and *he1* select solid and liquid food respectively is supported by various dictionaries and preliminary observation of corpora.

In terms of eventive verbal semantics, this fact suggests that the general event of ingestion is further classified according to the nature of the patients involved. However, corpus data show that there are significant counter-examples, such as below:

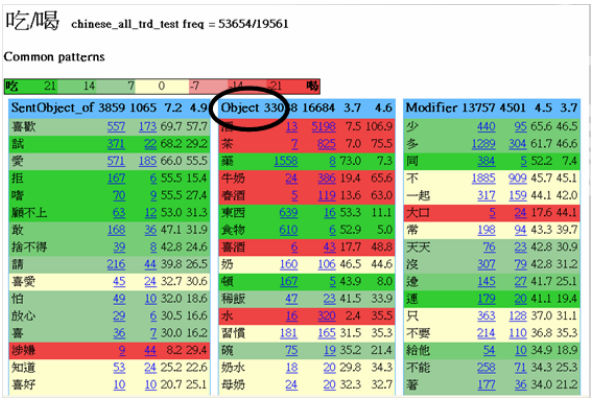
(3) *chi1/he1 xi1 fan4*
to eat/ drink porridge
(4) *chi1/he1 xi3 jiu3*
to eat/ drink wedding banquet
(5) *chi1/he1 nai3shui3*
to eat/ drink milk

Further examination of the situations involved in each of these counter-examples show that they represent three different kinds of event coercion as discussed below.

6 Chinese Word Sketch Results

The Sketch Difference is one of the most powerful functions of the Word Sketch Engine. This is a very efficient tool for doing contrastive studies. This function will be shown as the common pattern for *chi1* and *he1*, the *chi1* only pattern and the *he* only pattern.

In addition, in "object" argument role, we can find more significant comparisons as compared with other argument roles such as sentobject_of or modifier. This will support the explanation involving the neutralized selectional restrictions.



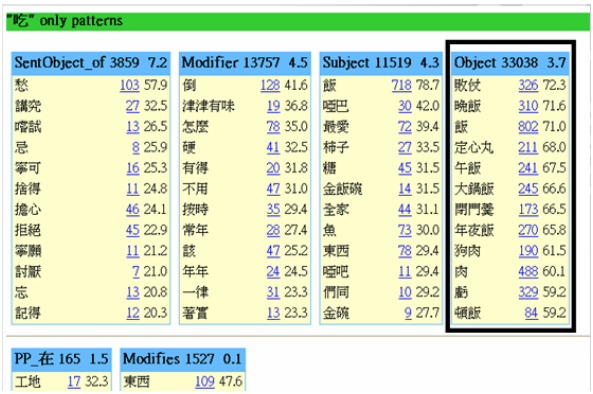
吃/喝 chinese_all_trd_test freq = 53654/19561

Common patterns

吃 21 14 7 0 -7 -14 -21 喝

SentObject_of 3859 1065 7.2 4.9				Object 33038 16684 3.7 4.6				Modifier 13757 4501 4.5 3.7			
喜歡	557	173	69.7 57.7	飯	13	5198	7.5 106.9	少	440	95	65.6 46.5
試	371	22	68.2 29.2	茶	7	825	7.0 75.5	多	1289	304	61.7 46.6
愛	571	185	66.0 55.5	菜	1558	8	73.0 7.3	同	384	5	52.2 7.4
拒	167	6	55.5 15.4	牛奶	24	386	19.4 65.6	不	1885	909	45.7 45.1
嗜	70	9	55.5 27.4	香酒	5	119	13.6 63.0	一起	317	159	44.1 42.0
難不上	63	12	53.0 31.3	東西	639	16	53.3 11.1	大口	5	24	17.6 44.1
敢	168	36	47.1 31.9	食物	610	6	52.9 5.0	常	198	94	43.3 39.7
捨不得	39	8	42.8 24.6	喜酒	6	43	17.7 48.8	天天	76	23	42.8 30.9
請	216	44	39.8 26.5	奶	160	106	46.5 44.6	沒	307	79	42.8 31.2
喜愛	45	24	32.7 30.6	領	167	5	43.9 8.0	逸	145	27	41.7 25.1
怕	49	10	32.0 18.6	稀飯	47	23	41.5 33.9	連	179	30	41.1 19.4
放心	29	6	30.5 16.6	水	16	320	2.4 35.5	只	363	128	37.0 31.1
喜	36	7	30.0 16.2	習慣	181	165	31.5 35.3	不要	214	110	36.8 35.3
涉嫌	9	44	8.2 29.4	碗	75	19	35.2 21.4	給他	54	10	34.9 18.9
知道	53	24	25.2 22.6	奶水	18	20	29.8 34.3	不能	258	71	34.3 25.3
喜好	10	10	20.7 25.1	母奶	24	20	32.3 32.7	著	177	36	34.0 21.2

Table 2: Objects Shared by *chi1* and *he1*

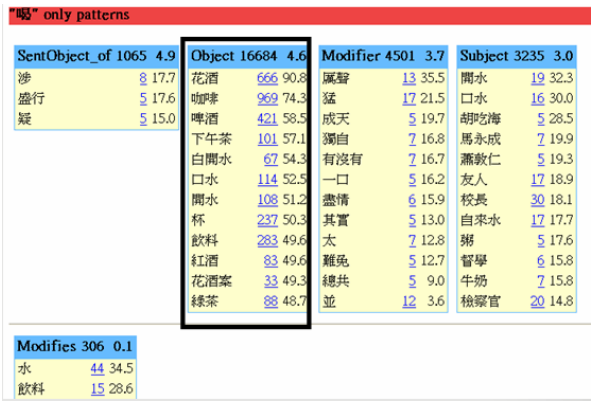


"吃" only patterns

SentObject_of 3859 7.2		Modifier 13757 4.5		Subject 11519 4.3		Object 33038 3.7	
愁	103 57.9	倒	128 41.6	飯	718 78.7	散仗	326 72.3
講究	27 32.5	津津有味	19 36.8	啞巴	30 42.0	晚飯	310 71.6
嚐試	13 26.5	怎麼	78 35.0	最愛	72 39.4	飯	802 71.0
忌	8 25.9	硬	41 32.5	柿子	27 33.5	定心丸	211 68.0
寧可	16 25.3	有得	20 31.8	糖	45 31.5	午飯	241 67.5
捨得	11 24.8	不用	47 31.0	金飯碗	14 31.5	大鍋飯	245 66.6
擔心	46 24.1	按時	35 29.4	全家	44 31.1	開門餐	173 66.5
拒絕	45 22.9	常年	28 27.4	魚	73 30.0	年夜飯	270 65.8
寧願	11 21.2	該	47 25.2	東西	78 29.4	狗肉	190 61.5
討厭	7 21.0	年年	24 24.5	啞吧	11 29.4	肉	488 60.1
忘	13 20.8	一律	31 23.3	們同	10 29.2	勵	329 59.2
記得	12 20.3	著實	13 23.3	金碗	9 27.7	頓飯	84 59.2

PP_在 165 1.5		Modifies 1527 0.1	
工地	17 32.3	東西	109 47.6

Table 3: *chi1* only pattern



"喝" only patterns

SentObject_of 1065 4.9		Object 16684 4.6		Modifier 4501 3.7		Subject 3235 3.0	
涉	8 17.7	花酒	666 90.8	風聲	13 35.5	開水	19 32.3
盛行	5 17.6	咖啡	969 74.3	猛	17 21.5	口水	16 30.0
疑	5 15.0	啤酒	421 58.5	成天	5 19.7	胡吃海	5 28.5
		下午茶	101 57.1	獨自	7 16.8	馬永成	7 19.9
		白開水	67 54.3	有沒有	7 16.7	瀋敦仁	5 19.3
		口水	114 52.5	一口	5 16.2	友人	17 18.9
		開水	108 51.2	盡情	6 15.9	校長	30 18.1
		杯	237 50.3	其實	5 13.0	自來水	17 17.7
		飲料	283 49.6	太	7 12.8	粥	5 17.6
		紅酒	83 49.6	難免	5 12.7	督學	6 15.8
		花酒案	33 49.3	總共	5 9.0	牛奶	7 15.8
		綠茶	88 48.7	並	12 3.6	檢察官	20 14.8

Modifies 306 0.1	
水	44 34.5
飲料	15 28.6

Table 4: *he* only pattern

The classical theory of selectional restriction stipulates that a verb checks certain semantic features in an argument and some features in this selection. We assume the features are [+/- food], [+/- solid], [+/- liquid] for two verb of ingestion *chi*1 and *he1*.

However, the collocational patterns extracted from Chinese Word Sketch strongly suggests otherwise. For this reason, neutralized selection as well as metaphorical and metonymic extensions point to the inadequacy of a feature checking account.

7 Different Levels of MARVS Theory Representation

The MARVS (Module-Attribute Representation of Verbal Semantics) theory offers a straightforward way to account for the three different types of event coercion. The two types of modules in MARVS: event modules and role modules allow two sets of attributes, event-internal attributes and role-internal attributes, to be described. Intuitively, *chi1* and *he1* select [+solid] and [+liquid] feature for the role-internal attributes of their patients respectively.

7.1 Xi1-Fan4 (Porridge)

When a noun such as *xi1fan4* (porridge) appears as the patient, it satisfy the event representation requirements of both *chi1* and *he1* simultaneously since it has both [+solid] and [+liquid] attributes.

For *xi1 fan4* (porridge), the patient involved contains both [+solid] and [+liquid] substances. Hence the event type is neutralized.

Within our lexical knowledge of porridge, it contains two important ingredients: rice (solid) in soup (liquid). Conventionally, it is a kind of rice diluted by water (*xi1-fan4*, where fan is rice) by lexical combination. Hence in a merged lexical ontology, it should inherit properties from both solid and liquid materials, depending on whether the focus on the rice or the soup. This kind of representation not only accounts for the fact that both verbs are allowed, but also picks up the subtle focus on the liquid type.

Through above explanations, we take MARVS theory to explicate the characteristic and feature of *xi1-fan4* and then we regard *chi1/he1 xi1-fan4* as role-internal attributes.

7.2 Xi3-Jiu3 (Wedding Banquet)

In regard to *xi3jiu3* (wedding banquet), literally lucky+wine', its metaphorical meaning is eventive and coerced the above event representation to shift from entity-type patient to event-type patient, with the [+solid] and [+liquid] attributes inherited. However, since a wedding banquet necessary include sub-events of eating solid food and drinking wine, it can select both verbs *chi1* and *he1*.

As for *xi3 jiu3* (wedding banquet), the metaphorical extension of the patient refers to a complex event type which contains separate sub-events that involves ingestion of solid and liquid foods respectively.

We use the wine drunk at the wedding banquet to refer to the event. In the wedding banquet, both eating and drinking are the most salient activities. The activity involves both eating and drinking, so both verbs *chi1* and *he1* are allowed.

Since two verbs of ingestion *chi1* and *he1* are allowed for *xi3 jiu3*, there are [+food] and [+liquid] in this event. Hence, we can identify *chi1/he1 xi3 jiu3* as event-internal attributes by MARVS theory.

7.3 Nai3-Shui3 (Milk)

In the case of "nai3shui3 (milk)", it presents the most intriguing situation. The patient involved is clearly liquid. Milk is liquid food in the literal interpretation. However, with a literal reading, the two verbs of ingestion are only substitutable only when the agent is an infant or young child. Again, when metaphoric uses are involved, "*nai3 sui3*" refers to nourishment for either the body or the soul (Hong and Huang, 2006).

We observe that the differentiation of liquid and solid food is only significant for adults as infant and young children can ingest only liquid food. In other words, when the agents are infant or young children, the liquid/solid classification of ingestion events are not applicable, since both types of events are sufficient to meet the ingesting goal of adequate nutrition.

Last but not least, the example "*nai3shui3* (milk)" showed that the classifications of events are dependent on the intention of the subject. Even though the solid/liquid contrast does exist in the physical/scientific world, the contrast is not significant for ingestion events involving infants and young children. Hence these two types of events are coerced and neutralized in the intentional context of these subjects.

In this case, a speaker uses both verbs of ingestion. This is because the metaphor involves nurture and nourishment. And therefore,"*chi1/he1 nai3shui3*" can be easily reprensented in MARVS by subject-internal attributes specifying that it allows the liquid/solid contrast to be neutralized. That is the Agent Role has the feature [INGEST OBJECT [+/- Solid]].

8 Conclusion

In this study, we collect empirical data from Sinica Corpus, show the pattern and usage of them from Gigaword Corpus by Chinese Word Sketch Engine and then take MARVS theory to explicate different levels of representations for event selection and coercion of two verbs of ingestion.

When we examine these two verbs of ingestion *chi1* and *he1*, we are able to support MARVS theory to identify different event types and different internal attributes based on the distributional data. We are also able to establish a model of event type selection and coercion, where we can predict the meaning of a non-typical event type object, as well as predict metaphoric meaning and event module-attributes.

References

Ahrens, Kathleen, Shirley Yuan-hsun Chuang, and Chu-Ren Huang, 2003. Sense and Meaning Facets in Verbal Semantics: A MARVS Perspective. Languages and Linguistics. 4.3.469-484.

Hong, Jia-Fei and Chu-Ren Huang. 2006. Using Chinese Gigaword Corpus and Chinese Word Sketch in linguistic research. The 20th Pacific Asia Conference on Language, Information and Computation (PACLIC-20). November 1-3. Wu-Han: China Huazhong Normal University.

Huang, Chu-Ren, Adam Kilgarriff, Yicing Wu, Chih-Min Chiu, Simon Smith, Pavel Rychlý, Ming-Hong Bai, and Keh-Jiann Chen. 2005. Chinese Sketch Engine and the Extraction of Collocations. In Proceedings of the Fourth SIGHAN Workshop on Chinese Language Processing, 48-55. October 14-15. Jeju, Korea.

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

Kilgarriff, Adam, Chu-Ren Huang, Pavel Rychlý, Simon Smith, and David Tugwell. 2005. Chinese

Word Sketches. ASIALEX 2005: Words in Asian Cultural Context. Singapore.

Ma Wei-yun, and Chu-Ren Huang. 2006. Uniform and Effective Tagging of a Heterogeneous Giga-word Corpus. Presented at the 5th International Conference on Language Resources and Evaluation (LREC2006). Genoa, Itlay. 24-28 May, 2006.

Pustejovsky, James. 1995. The generative lexicon. Cambridge: MIT Press. Rohsenow, J. S. 1978. Perfect *–le*: Aspect and relative tense in Mandarin Chinese. *Proceedings of symposium on Chinese linguistics: 1977 Linguistic Institute of LSA*, ed. by Robert Cheng, Y. C. Li and T. C. Tang, 267-291. Taipei: Student Book

Tsai, Mei-Chih, Chu-Ren Huang, Keh-Jiann Chen, and Kathleen Ahrens. 1998. Towards a Representation of Verbal Semantics--An Approach Based on Near Synonyms. *Computational Linguistics and Chinese Language Processing*. 3(1): 61-74.

黃居仁 主編。洪嘉馡, 陳韻竹. 2006. 執行總編輯。《意義與詞義》系列《中文詞彙意義的區辨與描述原則》. 中央研究院語言所詞網小組與詞庫小組技術報告.

Website Resources

Chinese Word Sketch Engine: http://corpora.fi.muni.cz/chinese_all/

English Word Sketch Engine: http://www.sketchengine.co.uk/

Lexical Data Consortium. 2005. Chinese Gigaword Corpus 2.5.: http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2005T14

Sinica Corpus. http://www.sinica.edu.tw/SinicaCorpus/