Corpus-based Lexical Semantic Study of Verbs of Doubt:

HUA⇔IYI⇔ 懷疑 **and** *CA□I* 猜 **in Mandarin**[1]

Mei-chun Liu

National Chiao Tung University

This paper examines the semantic differences between the near-synonymous pair in Mandarin – *ca#i* and *hu⇔aiy⇔i*. As verbs of doubt, they are similar in marking weak epistemic assertion but differ in event structure and presuppositional strength. The analysis is based on collocation patterns of the two verbs displayed in the Sinica Corpus. A Module-Attribute representational framework is adopted in this paper for detailing the semantic distinctions.

Key words: Lexical semantics, verbal semantics, Mandarin verbs, verbs of doubt

1. Introduction

1.1 Background Assumption

As lexical semantics has gained increased attention in recent linguistic research, the assumption put forth in Levin (1993:1) serves as a working hypothesis for studies of verbal semantics:

"… the behavior of a verb, particularly with respect to the expression and interpretation of its arguments, is to a large extent determined by its **meaning**."

By examining verb behavior, researchers aim to extract lexical semantic information that is grammatically pertinent. Lexical studies of verbs try to answer a fundamental question, that is, in what way is verbal syntax related to verbal semantics? In other words, what are the semantic properties coded in the verbal lexicon which potentially influence the syntactic behavior of verbs?

The present study reports a case study of verbal semantics with a focus on verbs of doubt in Mandarin. It aims to explore the semantic distinctions coded in the near-synonym pairs: *HUA⇔I YI*⇔ 懷 疑 and *CA□I* 猜 . The study takes on a

corpus-based approach and adopts the module-attribute representational framework proposed in Huang et al. (2000).

The advantage of using a corpus is that it provides a huge database of natural occurrences from which observational generalization and statistic comparison can be made. The association patterns revealed in the database indicate more clearly the range of variation in the grammatical behavior of verbs. What is characteristic of the corpus-based approach is that distributional tendencies, rather than grammaticality, are taken to be the central concern and evidence for linguistic analysis. As for the target of observation, near-synonym sets prove to be most useful in narrowing the scope of investigation with the benefit of verifiable contrasts in the same semantic field (Grandy 1992). The minimal contrastiveness presumably present between near-synonyms may then serve as a good indicator of meaning components. Tsai et al. (1998) and Chief et al. (2000) convincingly showed the advantages of looking at corpus distribution of near-synonyms.
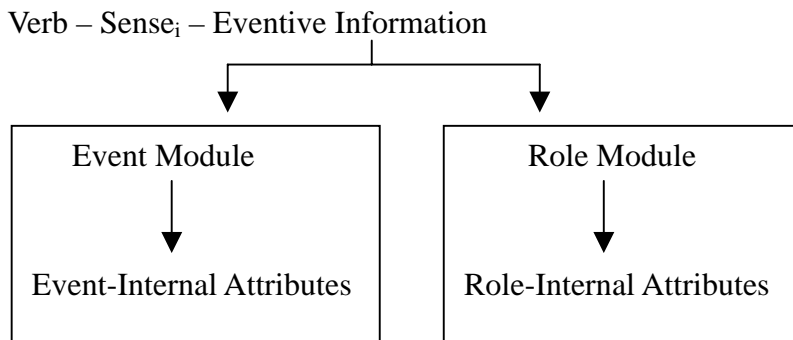
1.2 The Corpus

The corpus that is utilized in this study is the Academia Sinica Balanced Corpus (Sinica Corpus), which is the largest database of both written and spoken contemporary Mandarin, containing a total of 5 million words with part-of speech tagging (Chen et al. 1996). The corpus was developed by the CKIP (Chinese Knowledge and Information Processing) group at Academia Sinica, Taiwan, and it is open to the research community through the internet: http://www.sinica.edu.tw/ ftms-bin/kiwi.sh.

1.3 The Representational Framework

The representational framework adopted here is called the Module-Attribute Representation of Verbal Semantics (MARVS, Huang et al. 2000). The model views each verbal sense as one distinct *event structure* conveying *eventive information* which can be defined by the composition of modules and attributes. The Event Module represents the overall shape of the event structure. The Role Module represents salient participant roles. Within each module, detailed specifications are represented as attributes: Event-Internal Attributes are features pertaining to the whole event and Role-internal Attributes are features further specifying a participant role. The model

can be schematized as follows:

(1) Module-Attribute Representation of Verbal Semantics (MARVS):

Verb – Sense$_i$ – Eventive Information

Event Module	Role Module
↓	↓
Event-Internal Attributes	Role-Internal Attributes

The model takes grammar as information-based and lexicon-driven. It makes explicit three related premises. First, verbs with different senses will have different eventive information. The identification of verbal senses is then dependent on the identification of event structures. Second, the eventive information is based on the sum of all attested instances of event realizations. A complex lexical event may never be fully instantiated. Third, the event modules constitute the basic framework for verbal semantics. The classification of information is twofold: structural vs. attributive. There is thus a two-way distinction between modules and attributes. Pre-packaged structural information is viewed as modules while attached attributes provide more detailed description.

The advantage of the MARVS framework is that by maintaining the module vs. attribute distinction, it allows us to represent finer semantic properties within the same event structure or with the same participant role. Two verbs may share the same type of event structure or the same list of participant roles but differ in event-internal or role-internal attributes.

2. Verbs of Doubt

In the Sinica Corpus, there is a total of 137 occurrences of *CA☐I* 猜, and 369 occurrences of *HUA⇔IYI⇔* 懷疑. This study will show that the semantic contrast between the two verbs can be captured by aspectual, event-type distinctions and varied strength of epistemic assertion.

2.1 Initial Observation

　　Both *HUAIYI* and *CAI* are cognition verbs that mark epistemic modality. In the corpus, the two verbs are mainly used as predicates taking a complement-theme. Other grammatical uses of the two verbs are found with very little distributional significance. The table in (1) below shows the overall distribution of grammatical functions for the two verbs:

(2) Distribution of Grammatical Functions of *HUAIYI* and *CAI*

Function	*HUAIYI*		*CAI*	
Verbal				
+NP	55	(14.9%)	35	(25.5%)
+Clause	138	(37.4%)	39	(28.5%)
+zero	133	(36%)	42	(30.7%)
+DE-comp.	0		13	(9.5%)
Adjectival	7	(1.9 %)	7	(5.1%)
Nominalized	34	(9.2%)	1	(0.7%)
Adverbial	2	(0.5%)	0	

　　The two verbs share similar verbal patterns, taking either a nominal or a clausal complement. At first glance, they share certain meaning components and can both occur in the following contexts with either an affirmative (3) or a negative (4) complement clause:

(3) 我 懷疑/猜 他是兇手

　　wo HUAIYI / CAI ta shi xiongshou

　　I HUAIYI / CAI he be murderer

　　' I suspect that he is the murderer.'

(4) 我 懷疑/猜 果汁不是純的

　　wo HUAIYI / CAI guozhi bu shi chun de

I HUA≅IYI⇔ / CA□I juice not be pure DE

'I suspect that the juice is not pure.

Despite the similarity in the above examples, they behave differently in other structures. For example, *HUA⇔IYI⇔*, but not *CA□I*, can be modified by a preverbal degree adverbial *he◊n* 很 'very', as in (5), but *HUA⇔IYI⇔* cannot take a postverbal resultative, as in (6):

(5) 我很 懷疑/*猜

　　 wo◊ he◊n HUA⇔IYI⇔/ CA□I*

　　 I very HUA≅IYI⇔/* CA□I

　　 'I'm very doubtful.'

(6) 我 猜/*懷疑 得出

　　 wo◊ *CA□I / *HUA⇔IYI⇔ de chu##*

　　 I CA□I / *HUA⇔IYI⇔ DE out

　　 'I can guess it.'

Apparently, the verb *HUA⇔IYI⇔* is more like a stative predicate, allowing a degree modifier; while *CA□I* is more like an active predicate, allowing a potential resultative complement.

The two examples above reflect some preliminary differences in their event structures. In the following, their collocational patterns will be closely examined with generalizations on their semantic distinctions.

2.2 Collocation and Aspectual Distinction

As observed in example (4), the verb *HUA⇔IYI⇔* can co-occur with degree modifiers *he◊n/shi⇔fe⇦n/fe⇦icha⇔ng* 很/十分/非常 or the evaluative marker *zhi⇔de⇔* 值得 'worth' that usually only take stative verbs. The distributional frequency of the two verbs with evaluative elements is shown below:

(7) Collocation with evaluative markers

	Degree modifier *he⟨n/shi⟺fe⟸n/fe⟸icha⟺n* 很/十分/非常 'very'	Evaluative marker *zhi⟺de⟺* 值得 'worth'
HUA⟺IYI⟺	24 (6.5%)	6 (1.6%)
CA□I	0	0

This shows that the two verbs differ fundamentally in event type: *HUA⟺IYI⟺* is more stative, subject to degree modification and evaluation.

In addition, they display other distinctions in aspectual properties, as evidenced from their collocation with various aspect-marking elements. *HUA⟺IYI⟺* can be preceded by the inception verb *ka⟸ishi◊* 'start', denoting an inchoative change, but *CA□I* cannot:

(8) 她開始 懷疑/*猜 果汁到底是不是純的.

　　ta⟸ ka⟸ishi◊ HUA⟺IYI⟺/ CA□I guo◊zhi⟸ da⟘odi◊ shi⟘ bu⟺ shi⟘ chu⟸n de*

　　she start HUA≅IYI⟺/ * CA□I juce to bottom be not be pure DE

　　'She started wondering if the juice was pure.'

Below is their distributional frequency with two inchoative-marking devices, preverbal *ka⟸ishi◊* 'start' and the postveral *qi◊la⟺i* 'up'. It is clear that *HUA⟺IYI⟺* allows an *event focus* on the starting point[2]. An 'event focus' refers to the profiled event component of a complex event (For details, see Liu 1999).

(9) Collocation with inchoative marking

	Inceptive verb	Inchoative –*qi◊la⟺i*

---

[2] It seems to be intuitively true that *CA□I* may also occur with the verb *ka⟸ishi◊* indicating an inceptive-progressive aspect, as in the imperative *ka⟸ishi◊ ca#i!* 'You may start guessing'. This use is, however, not found in the corpus. Given that *CA□I* is essentially an activity verb, it is potentially compatible with inceptive markers.

	ka⇐ishi◊	
HUA⇔⇔IYI⇔	14 (3.5%)	1 (0.2%)
CA□I	0	0

However, with regard to the marking of an endpoint, *CA□I* can be followed by the adverbial *WA⇔N* 'finish', denoting the completion of an event, but *HUA⇔IYI⇔* cannot:

(10) 你到底 猜/*懷疑 完了沒?

　　 ni◊ da⫞odi◊ CA□I / HUA⇔IYI⇔　 wa⇔n le me⇔i*

　　 you to bottom CA□I /* HUA≅IYI⇔ finish LE no

　　 'Have you on earth finished guessing?'

Moreover, while *CA□I* can co-occur with a durational phrase of time, *HUA⇔IYI⇔* cannot:

(11) 他已經 猜/*懷疑了三天三夜了

　　 ta⇐ yi◊ji⇐ng CA□I / HUA⇔IYI⇔　 le sa⇐n tia⇐n sa⇐n yie⫞ le*

　　 he already CA□I /* HUA≅IYI⇔ LE three day three night LE

　　 'He has already been guessing for three days.'

The above examples (10)-(11) show that the event structure of *CA□I* involves a process, which may be bounded by an 'endpoint'. When the endpoint is profiled, it predicates the stative result of 'guessing', the acquisition of a certain weak 'belief', as exemplified in (3)-(4) above. The meaning extension from a process to a stative prediction seems to be characteristic of verbs of cognition. For example, the process of 'thinking' may result in obtaining a particular 'thought', expressed as a clausal complement to the verb 'think'.

The distributional differences between *HUA⇔IYI⇔* and *CA□I* as discussed above indicate clearly their distinction regarding aspectual composition, which can be represented with different event modules in MARVS: The verb *HUA⇔IYI⇔* may

co-occur with a degree or evaluative adjunct, indicating its stativity; and it allows a predicative focus on the starting point of the state, indicating a change of state or *inchoative state.* On the other hand, *CA□I* may be used with a perfective marker or a durative phrase, typical of an 'activity' situation (cf. Smith 1991). The event *CA□I* refers to is a potentially on-going process that may have a final point, a resultative state of the cognitive activity:

(12) Even Modules of *HUA⇔IYI⇔* and *CA□I*:

    HUA⇔IYI⇔: inchoative state   • ____

    CA□I: bounded process   • ////// •

The above distinction in the event module also bears implications on the role Module: the subject of *HUA⇔IYI⇔* is more like an experiencer, while *CA□I* takes a volitional agent.

2.3 Presupposition and Epistemic Distinction

Another important and interesting distinction between the two verbs has to do with their epistemic marking capacity and contextual presupposition. Both verbs indicate low epistemic assertion of a proposition, distinct from strong epistemic assertion verbs such as *ren∃we⇔i* 'think, or *xia#ngx∃in* 'believe'. The extremely weak assertional strength in the use of *HUA⇔IYI⇔* may even lead to the opposite – a 'denial' of a proposition that is contextually presupposed. The verb *HUA⇔IYI⇔* expresses such a strong doubt that it may implicate a denial of the truth-value of the proposition. Take (3) as an example, repeated here in (13): *wo◊ HUA⇔IYI⇔ ta⇐ shi�q̂xio⇐ngsho◊u* 我懷疑他是兇手 'I doubt that he is the murderer'. The sentence is potentially ambiguous[3]. Besides the possible reading as making a weak assertion, similar to the use of *CA□I,* in saying 'I doubt that he is the murderer', one can actually intend to deny a commonly-held presupposition that he IS the murder, given an appropriate context. As

---

[3] Out of context, the sentence (Example 13) may have two different interpretations. It may be used to claim a weakly asserted belief 'I think he is the murder' or to challenge such a presupposition. The point here is that the possibility of using *HUA⇔IYI⇔* to challenge a presupposition reveals its peculiar status in epistemic marking.

shown in (13), it is possible to use *HUA⇔IYI⇔* to challenge a statement of presupposition:

(13) Challenge to a presupposition

Statement of a Presupposition

我　懷疑　　[他是兇手]

wo◊ *HUA⇔IYI⇔*　[*ta⇐shi⫽xio⇐ngsho◊u*].

I HUA≅IYI⇔ he be murderer

'I doubt that **he is the murderer.'**　(Contrary to the entertained belief, I don't think he is the murderer.)

The above interpretation is made possible because the verb *HUA⇔IYI⇔* has such low epistemic strength that it may function as a negator to challenge a pre-existing assumption. The key to this pseudo-negation use is the existence of a contextual presupposition. Without any contextual presupposition, *HUA⇔IYI⇔* serves to make a weak assertion with very low epistemic certainty; it functions to make an *irrealis assertion*, as defined in Givon (1993: 170): 'The proposition is weakly asserted as either possible or likely; but the speaker is not ready to back it up with evidence or other strong grounds; and a challenge from the hearer is readily entertained or even explicitly solicited.'

However, when occurring with a clearly inferred contextual presupposition, *HUA⇔IYI⇔* serves to mark it with such a low degree of certainty that it is nearly negated. This particular use of *HUA⇔IYI⇔* resembles a *negative assertion* in its presuppositional requirement. Givon (1993:189) has made it clear that a negative assertion is made 'on the tactic assumption that the hearer either has heard about, believes in, is likely to take for granted, or is at least familiar with the corresponding affirmative proposition.' This observation is crucial to the understanding of the epistemic distinction between the two verbs.

3. MARVS Representation of *HUA⇔IYI⇔* and *CA□I*

Besides their module distinction in event types, the two verbs also differ in event-internal characteristics: while they both encode very low epistemic strength, *HUA⇔IYI⇔* may mark an assertion as extremely weak that it almost negates or counters the proposition. Below is a schematic representation of the semantic distinctions between the two verbs using the Module-Attribute model:

(14) MARVS for *HUA⇔IYI⇔* 懷疑 and *CA□I* 猜:

Verb	Event Module Event-Internal Attributes	Role Module Role-Internal Attributes

HUA⇔IYI⇔
懷疑
	inchoative state • _____
	
	                  |
	
	        [irrealis assertion]◄------------► [challengeable]
	            [denial] ◄-----------► [presuppositional]
		<Experiencer, Theme>
		
		                      |

CA□I 猜	bounded process • ////// •
	                 |
	
	        [low epistemic strength]
		<Agent, Theme>
		
		               |
		
		            [challengeable]

4. Conclusion

Verbs of doubt are semantically a subset of verbs of cognition. The verbs in this class normally involve two kinds of participant roles: the Initiator of Cognition (Experiencer or Agent) and the Content of Cognition (Theme). Different verbs encode a different event structure that selects a different view of aspectual composition and different details of the participant roles. Moreover, as can be seen from the above discussion, with this particular subset, the epistemic assertional strength encoded in a particular verb may be the source for potential pragmatic implication. The lexical semantic distinctions made in this study can also be further studied as to their interaction with pragmatic implicature.

The importance of lexical semantics has been highly recognized in recent years as linguistics in general is pushing its frontier toward lexicon-driven theories and practices.

The study on verbal semantics reported here is far from conclusive or perfect. It aims, however, to demonstrate a corpus-based approach that utilizes the minimal contrastiveness of near-synonyms. A series of such studies has been done as the result of a research program funded by the National Science Council, Taiwan. (For a comprehensive overview, see Liu 2001). Additional collaborative research will be needed for subsequent, long-term endeavors that will ultimately unveil the lexical fundamentals as well as the complexity of Mandarin verbs.

References

Chen, Keh-Jiann, Chu-Ren Huang, Li-Ping Chang, Hui-Li Hsu. 1996. SINICA CORPUS: Design Methodology for Balanced Corpora. In Proceedings of the 11th Pacific Asia Conference on Language, Information and Computation (PACLIC 11). Seoul, Korea.

Chief, Lian-Cheng, Chu-Ren Huang, Keh-Jiann Chen, Mei-Chih Tsai, and Li-Li Chang. 2000. What Can Near Synonyms Tell Us. *International Journal of Computational Linguistics and Chinese Language Processing* 5.1: 47-60.

Givon, T. 1993. *English Grammar: A Function-based Introduction,* Vol. 1. Amsterdam/Philadelphia: John Benjamins Publishing Co.

Grandy, Richard E. 1992. Semantic Fields, Prototypes, and the Lexicon. In *Frames, Fields, and Contrasts: New Essays in Semantic and Lexical Organization*, ed. by A. Lehrer and E. Kittay, 103-122. Hillsdale: Lawrence Erlbaum.

Huang, Chu-ren, Mei-chun Liu, and Mei-Chi Tsai. 1999. From Lexical Meaning to Event Structure Attributes: Across Semantic Classes of Mandarin Verbs. In *Working Papers on Chinese Verbal Semantics, Vol. 1*, ed. by K. Ahrens, C-R Huang, and M-C Tsai, 15-28. Taipei: Academia Sinica.

Huang, Chu-ren, Kathleen Athens, Li-Li Chang, Keh-Jiann Chen, Mei-chun Liu, Mei-chih Tsai. 2000. The Module-Attribute Representation of Verbal Semantics: From Semantics to Argument Structure. *International Journal of Computational Linguistics & Chinese Language Processing* 5 (1): 19-46.

Levin, Beth. 1993. *Verb Classes and Alternation*. Chicago: Univ. of Chicago Press.

Liu, Meichun. 1996. Conceptual Manipulation and Semantic Distinction: the Case of Mandarin Postverbal DE Complement. In *Conceptual Structure, Discourse, and Language*, ed. by A. Goldberg, 333-46. Stanford: CSLI.

-----. 1999. Lexical Meaning and Discourse Patterning - the Three Mandarin Cases of 'Build'. In *Cognition and Function in Language*, ed. by B. Fox, D. Jurafsky, & L. Michaelis, 181-199. Stanford: CSLI.

-----. 2000. Categorical Structure and Semantic Representation: Mandarin Verbs of Communication. Paper presented at the 5th Conference on Conceptual Structure, Discourse and Language. University of California, Santa Barbara.

-----. 2001. *Mandarin Verbal Semantics: A Corpus-based Approach*. Taipei: Crane Publishing Co.

Liu, Meichun, Chu-Ren Huang, and Zhi-Ling Chang. 1999. The Locus-Locatum

Alternation: a Lexical Semantic Study of Mandarin Verbs of Surface Contact. Paper presented at the North American Conference on Chinese Linguistics, Harvard Unversity.

Liu, Meichun, Chu-Ren Huang, Charles Lee, and Ching-Yi Lee. 2000. When Endpoint Meets Endpoint: a Corpus-based Lexical Semantic Study of Mandarin Verbs of 'Throwing'. *International Journal of Computational Linguistics and Chinese Language Processing* (5) 1:81-96.

Smith, C. S. 1991. *The Parameter of Aspect*. Dordrecht: Kluwer Academic Publishers.

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

Dept. of Foreign Languages and Literatures
Graduate Institute of Linguistics and Cultural Studies
National Chiao Tung University
Hsinchu, TAIWAN
mliu@cc.nctu.edu.tw

中文認知動詞「懷疑」和「猜」的詞彙語意探討

劉美君

國立交通大學

本文藉由觀察語料庫中「懷疑」和「猜」的語法表現,來界定此二動詞的語意區別。並利用「模組屬性」的表徵模式,將二者在事件結構及參與角色上的細部內涵表達出來。「懷疑」和「猜」均屬低認定度的認知動詞,「懷疑」甚至可出現否定某論述的用法。在事件結構上,「猜」屬動作性的事件,可有進程,可有結果。「懷疑」則屬狀態性

的事件，可描述認知狀態的變化。本文試圖將詞彙語意的區別化約爲可清楚界定的語意屬性。

關鍵詞: 詞彙語意，動詞語意，中文動詞，猜疑類動詞