From Collocation to Event Information:
The Case of Mandarin Verbs of Discussion[*]

Mei-chun Liu

National Chiao Tung University

The importance of deciphering lexical knowledge, especially the eventive information encoded on verbs, has been highly recognized and various approaches have been proposed to represent lexical information that plays a crucial role in grammatical realization. In this paper, a subset of Mandarin verbs of communication —verbs of discussion—is explored in detail, with reference to the semantic properties characteristic of the whole class. While quantitative NLP approaches might tend to overlook some semantic details, it is shown that a corpus-based, contrastive analysis of near-synonyms can be most useful in extracting lexical information that is critical in differentiating verbs.

Key words: lexical semantics, Mandarin verbal semantics, verbs of discussion, corpus-based, semantic attributes

1. Introduction

As lexical information is key to natural language processing (NLP), the need for processing a large amount of lexical information has always been an issue in NLP research and applications. From a purely linguistic perspective, the study of lexical semantics has also been a focal area in recent years, as linguistics in general expands its frontiers toward lexicon-driven theories and practices. More and more researchers believe that the lexicon is where most of our linguistic knowledge has been stored. Entries of verbs, in particular, constitute a central part of the lexicon and are crucial for understanding grammatical structures. In the following, I shall give a brief overview of the development of verbal semantics in terms of *why* it is studied, *what* is to be studied, and *how* it has been studied.

1.1 Why study verbal semantics?

The semantics of verbs has always been a core concern in linguistic theories. In the early generative paradigm, the meaning of a verb is generalized into and disguised

---

under the so-called subcategorization frame. Formal theories in general try to find ways to link sentence structures with the argument structures of verbs, which presumably are lexically-specified. Case Grammar looks at argument structures from a purely semantic perspective and defines verb meanings with semantic roles (Fillmore 1968). As huge lexical databases are built to respond to the needs of NLP applications, linguistic research has also shifted directions in recent years, from approaches that are structurally based to those that are lexically based. Verbal semantics has thus gained increased significance with regard to linguistic analysis as well as knowledge representation.

Each verb lexicalizes some unique eventive information with a range of possible 'templates' for argument expressions. As Levin & Rappoport Hovav (1996) put it, it is generally assumed that the syntactic realization of arguments—their syntactic type and grammatical function—is predictable to a large extent from the meaning of verbs. The major goal in verbal semantic studies is then to extract the 'meaning' of a verb from examining the range of its argument expressions and other collocational associations.

1.2 What is to be studied?

Under the assumption that the meaning of a verb determines its syntactic behavior, various models have been proposed to explain the mapping from lexical semantics to syntax. Central to the issue is the task of *identifying and representing semantic 'determinants' that shape the syntactic behavior of a verb*. Given that meanings are multi-faceted and semantic distinctions are basically open-ended, what exactly are the meaning components that are grammatically relevant?

To answer this question, Levin & Rappoport Hovav (1996) provided examples from English to illustrate 'grammatically relevant' aspects of meaning. For most people, 'loudness of speech' seems to be a well-defined semantic feature that is also cognitively salient. However, verbs that differ in loudness of speech, such as *whisper* vs. *shout*, do not show any major differences in argument expression. This shows that the parameter of 'loudness' may not be syntactically relevant. On the other hand, the semantic distinction between 'manner of speaking' and 'content of speaking' proves to be relevant to syntax, as evidenced from their differences in participating in the conative pattern, taking an *at*-phrase:

(1) I whispered/*say at Mary.

It is clear that only certain aspects of meaning will surface in syntactic realizations and constitute the target of investigation. Lexical semanticists are concerned with ways of delimiting these syntactically relevant semantic components from the syntactically irrelevant information.

1.3 How is verbal semantics studied?

In searching for the grammatically relevant aspects of meaning, a number of approaches have been attempted in the past. In the following, I shall briefly introduce some of the major approaches, including frame semantics (Fillmore & Atkins 1992), Alternation-based approach (Levin 1993), conceptual structure and the localist approach (Jackendoff 1990), Generative lexicon (Pustejovsky1995), and the corpus-based approach (Biber et al. 1998).

1.3.1 Frame semantics

The semantic frame-based approach argues that a word's meaning 'can be understood only with reference to a structured background of experience, constituting a kind of conceptual prerequisite for understanding the meaning' (Fillmore & Atkins 1992). Take for example the group of commercial transaction verbs—*buy*, *sell*, *charge*, *spend*, *pay*, and *cost*. These verbs all require an understanding of property ownership and money economy, a knowledge schema involving four major semantic categories: Buyer, Seller, Goods, and Money. These verbs differ in the ways of expressing these categories (The table below is taken from Fillmore & Atkins 1992:79):

(2) Semantic and Syntactic Valence for Verbs in the Transaction Frame

	Buyer	**Seller**	**Goods**	**Money**
Buy	Subj	(from)	Direct-Obj	(for)
Sell	(to)	Subj	Direct-Obj	(for)
Charge	(Indirect-Obj)	Subj	(for)	Direct-Obj
Spend	Subj	NULL	For/on	Direct-Obj
Pay	Subj	(Indirect-Obj)	(for)	Direct-Obj
Cost	(Indirect-Obj)	NULL	Subj	Direct-Obj

In sum, verbs of the same class share the same semantic frame, defined mainly with a set of frame-specific elements. The differences among the verbs can be accounted for with a distinct set of 'highlighted' roles as well as their mapping relations to grammatical functions.

Frame semantics goes hand in hand with the general assumption in cognitive semantics that meaning can only be understood in relation to a set of background information or a cognitive model (termed as Idealized Cognitive Model (ICM) in Lakoff 1987). However, besides a mere listing of frame-specific categories and their

grammatical realization patterns, frame semantics fails to provide further analysis on aspectual information pertaining to the understanding of event structure.

1.3.2 Diathesis alternation

Levin (1993) provides a complete study of English verbs and verb classes, based on diathesis alternation patterns, i.e., alternations in the expression of arguments. The basic assumption behind the work is that the behavior of a verb, particularly with respect to the expression of its arguments, is to a large extent determined by its meaning. And the group of verbs exhibiting the same alternation patterns shares the same meaning components and belongs to the same semantic class. For example, the Locative Alternation in English can be used to distinguish the *spray/load* group of verbs from the *fill/cover* group, since only the *spray/load* group can participate in Locative Alternation, as exemplified below:

(3)   Locative Alternation in English
  a. *He sprayed water on the plants*
  b. *He sprayed plants with water.*

A more complicated example is found with a set of prototypical transitive verbs: *touch*, *hit*, *cut*, and *break* (Levin 1993:6-10). These verbs vary in terms of their participation across four different transitive alternations:

(4)   English Transitive Alternations

	touch	*hit*	*cut*	*break*
Conative *'X hits Y.'* *'X hits at Y.'*	No	Yes	Yes	No
Body-part Possessor Ascension *'X hit Y's head.'* *'X hits Y on the head.'*	Yes	Yes	Yes	No
Middle *'X cut Y.'* *'X cuts easily.'*	No	No	Yes	No
Causative *'X broke Y.'* *'Y broke.'*	No	No	No	Yes

Each of the above alternations is in principle associated with certain meaning components. Verbs that may participate in a given alternation are said to have the

associated meaning components. Based on their differences as shown above, the verbs can be categorized into different verb classes, with distinct semantic features:

(5) Four Verb Classes:
    a. *touch:* pure verb of <u>contact</u> (no implication for change of state)
    b. *hit:* a verb of <u>contact by motion</u>
    c. *cut:* a verb of causing a <u>change of state</u> by <u>moving</u> something into <u>contact</u> with the entity that changes state.
    d. *break:* a pure verb of <u>change of state</u>

In order to classify verbs with alternations successfully, the semantic prerequisites for each alternation pattern will have to be clearly defined. And the alternation-based approach faces the problem of identifying all the necessary and sufficient semantic conditions for each alternation. In other words, the direct associations of alternation patterns to semantic components are difficult to establish. While relying heavily on argument-expression possibilities for the extraction of semantic properties, it is not clear what pattern of argument expression indicates what meaning component.

1.3.3 Conceptual structure and the localist approach

According to Jackendoff (1990), 'meaning' in natural language is an information structure that is mentally and spatially encoded. The semantic primitives that may be utilized to construct the mental representation include the following:

(6) Semantic primitives:
    a. [STATE]   BE ([THING], [PLACE])
    Jim is in the pub. (spatial) → The party is on Sat. (non-spatial)
    b. [EVENT]   GO ([THING], [PATH])
    Jim went into the pub. (spatial) → The prize went to Kate. (non-spatial)
    c. [PLACE]   IN ([THING])
    d. [PATH]   TO ([PLACE])

This mental structure approach is de-compositional in nature. While it attempts to rebuild verbal meanings upon cognitively primitive constructs, no clear indications are made as to how complex event structures can be represented with the limited set of primitives.

1.3.4 Generative lexicon

Besides the inherent sense of a verb, it is noted that a verb may predicate different activities when used in different contexts. For example, *enjoying a book* and *enjoying a cup of coffee* actually involve two distinct events, that is, reading vs. drinking. As an attempt to deal with both lexically-specified information and contextually-derived meaning coërcions, Pustejovsky (1995) proposed some compositional mechanisms for meaning generation. In his framework, there are multiple levels of semantic representation:

(7)　Levels of Semantic Representation:
　　a.　Argument Structure
　　　　eg. *Build*　ARG1=animate, ARG2 = artifact
　　b.　Event Structure (lexical aspect)
　　　　eg. *Build*　E1= Process, E2=state
　　c.　Qualia Structure (eg. *book*)
　　　　-Formal : type of things　　　　　　'This book is thick.'
　　　　-Agentive : how to 'come to being'　　'He wrote a book.'
　　　　-Telic : purpose　　　　　　　　　'He read a book.'
　　　　-Constitutive: parts/content　　　　'The book in interesting.'
　　d.　Inheritance Structure

The generative approach attempts to solve the issue that verbs may have different meaning 'facets' when combined with other clausal elements. The qualia structure is useful in disambiguating possible readings of the same verb phrase. For example, in Mandarin, *gan xi* 'rush (a) play' may have an agentive reading pertaining to the act of producing a play, or a telic reading referring to the watching of the play (Liu, to appear).

1.3.5 The corpus-based approach

As large corpora of natural language data were built and made available to researchers, linguistic studies could make use of the distributional tendencies observable in a corpus. Proponents of the corpus-based approach (e.g., Biber et al. 1998) believe that the meaning of a verb is projected in its 'association patterns' found in a large corpus. Statistical counting of 'association patterns' then serves to reveal the semantic distinctions. For example, the difference between the two adjectives *strong vs. powerful* is best illustrated by their collocational tendencies: *strong coffee vs. powerful car*. And the statistical findings on the complement types between the two verbs *begin* and *start* also demonstrate their semantic differences (Biber et al. 1998):

(8) Corpus-based Statistics: *begin vs. start*

	begin	*start*
+*To* -V	60%	17%

It is shown that *begin* is followed much more often by an infinitival phrase, indicating that *begin* is a complement-taking verb, while *start* predicates a simple event.

The association patterns observable from a large corpus help to 'delimit' the grammatical distinctions that require a semantic account. However, linguistic analyses are still needed to 'define' the relevant meaning facets behind the grammatical differences. It is shown in this paper that some association patterns of verbs may not be readily discoverable and semantic analyses of the grammatical patterns may be the key to interesting discovery.

1.4 Current program on Mandarin verbal semantics

The above approaches have been successfully applied to the study of English verbs. But are they equally applicable to the study of Mandarin verbs? A team of researchers have worked on Mandarin verbal semantics in recent years and found that a contrastive analysis of near-synonym pairs, based on corpus observations proves to be effective in 'extracting' verb meanings that are syntactically relevant. The program aims to identify eventive information that is *syntactically* relevant by comparing the association patterns of near-synonym sets. It makes use of the largest balanced corpus of Taiwan Mandarin, the Sinica Corpus, containing 5 million words, as the primary source of data (http://www. sinica.edu.tw/ftms-bin/kiwi.sh/).

Various works on Mandarin verbal semantics have been published that focus on various classes of verbs (Tsai et al. 1998, Chang et al. 2000, Liu 1999, Liu 2000; for a collection of results, see Liu 2002). In this paper, I shall focus on a sub-type of communication verbs, namely, verbs of discussion (商量 vs. 討論), and demonstrate how a detailed linguistic analysis of near-synonyms can help find the crucial semantic distinctions between the pair of verbs, while other currently available resources with a top-down approach to the verb system tend to miss these important distinctions.

2. Verbs of discussion: Two-way communication

2.1 Frame-based analysis

Communication verbs constitute a basic domain in our lexicon as they encode the most fundamental aspect of human activities. From the perspective of frame semantics

(Fillmore & Atkins 1992), verbs of communication can only be understood with respect to some knowledge schema or conceptual structure pertaining to human communication, i.e., a 'frame'. In order to capture the different frames that give rise to various lexical entries of communication, 13 frames are distinguished in FrameNet I[1] (http://www.icsi. berkeley.edu/~framenet), as shown below with their representative lemma:

(9) Frames in the Communication Domain:
   a. Conversation: *talk, discuss, argue*, etc.
   b. Statement: *address, claim, comment*, etc.
   c. Encoding: *express, voice, put*, etc.
   d. Questioning: *question, inquire, ask*, etc.
   e. Request: *ask, beg, command*, etc.
   f. Response: *answer, reply, respond*, etc.
   g. Hear: *hear, overhear, read*, etc.
   h. Commitment: *promise, swear, threaten*, etc.
   i. Manner: *babble, shout, whisper*, etc.
   j. Volubility: *(be) talkative, chatty, silent*, etc.
   k. Candidness: *(be) candid, frank*, etc.
   l. Noise: *groan, scream, bark*, etc.
   m. Gesture: *gesture, nod, beckon*, etc.

These frames are meant to characterize the distinct senses encoded in verbs of communication. Each frame is clearly defined with its prototypical 'frame elements', that is, core participant roles involved in the event. Among the 13 frames, the Conversation frame is unique in that it captures the nature of two-way communication. In all the other frames, a Speaker and an Addressee are put forth as the major participants, depicting a uni-directional path of communication. The Conversation frame, on the other hand, involves 'Interlocutors', a term used to depict co-agency of the conversationalists. Below is a list of the Frame Elements in Conversation:

(10) Frame Elements in the Conversation frame
   a. Interlocutor 1
   b. Interlocutor 2
   c. Interlocutors
   d. Topic
   e. Medium

---

[1] FrameNet is a frame-based lexical database of verbal information in English (Baker et al. 1998).

Regarding the lemma in the Conversation frame, the verb *discuss* is among the most representative ones, as it illustrates all the core frame elements:

(11) *They* [Interlocutors] *discussed the issue* [Topic] *over the phone* [Medium].

It seems that by characterizing the discussants as Interlocutors and highlighting the role Topic, the meaning of *dicuss* is well embedded in the conversation frame. However, in Mandarin, there are two distinct lexical entries that may both be glossed as 'discuss'—討論 *tǎolùn* and 商量 *shāngliáng*. And they both share the same set of frame elements:

(12) 他們 [Interlocutors] 在電話上 [Medium] 商量/討論 這件事 [Topic]。
    'They discussed the issue over the phone.'

What are the semantic distinctions between them? It is clear that a more in-depth analysis beyond the frame-based approach is needed to identify finer semantic details that make them distinct.

In the following, I shall first examine information in existing resources and see what they can offer. These resources are more or less built upon a top-down approach to verbal meanings.

2.2 Dictionary definitions—Verb Usage Dictionary (動詞用法詞典)

In *Verb Usage Dictionary* (Meng et al. 1985), a dictionary focusing solely on verbal knowledge, the two verbs are defined as:

(13) Dictionary definition:
    討論：就某一問題交換意見或進行辯論
    商量：交換意見

The definitions seem to be interchangeable and do not show any critical differences. If we look at their respective usages, again, the information is roughly the same:

(14) Usages for 討論 and 商量 in Verb Usage Dictionary
    a. 討論
    [名賓]　～科研計畫
    [動賓]　～怎麼處理這件事
    [小句賓]　～這個人怎麼安排

[動時量]　～一下兒
[了著過]　～了一上午/還～著呢/～過這個問題
[重疊]　～～計畫生育問題
【名賓類】[對象] ~問題|~計劃|~措施|~方案
【動結】~//清楚|~//完~得/不了~//成　~好
【動趨】~上~//下去~出　~[不]出來~得/不過來~起

b. 商量
[名賓]　～一件事
[動賓]　～怎麼辦
[小句賓]　～此事讓誰處理
[動時量]　～一下兒
[了著過]　～了一上午/還～著呢/～過這個問題
[重疊]　～～計畫生育問題
【名賓類】[對象] ~問題|~計劃|~措施|~方案
【動結】~//清楚|~//完~得/不了~//成　~好
【動趨】~上~//下去~出　~[不]出來~得/不過來~起

From the above, we see that the ranges of usage for the two verbs are also identical. The information in the dictionary does not appear to be helpful.

2.3 The Grammatical Knowledge-Base of Contemporary Chinese: A complete specification (現代漢語語法信息辭典詳解)

Next, we look at resources that are designed for NLP applications. In the *Grammatical Knowledge-Base of Contemporary Chinese* (現代漢語語法信息辭典詳解, Yu et al. 1998), there are two differences found: first, 討論 can be the object of 'have' as in 有一個討論 'have a discussion', but 商量 cannot; second, 討論 can be nominalized and preceded by a noun as in 政策討論 'policy discussion'. The two differences are highlighted below:

(15) Information in the Grammatical Knowledge-Base of Contemporary Chinese

詞語	討論	商量
准謂賓	准	准
有賓	有	
前名	可	
後名	可	可
體謂准	體謂	體謂

句賓	句	句
動趨	趨	趨
著了過	著了過	著了過
在	在	在
重疊	ABAB	ABAB
謂	可	可
備註	時事~/~時間/~這事怎樣處理	~對策/~結果/~誰去辦理/這件事~過了

Besides listing the syntactic differences here, no further account is given as to the significance of these differences.

2.4 HowNet

In HowNet (Dong 1998a, b), both verbs are defined with the concept 'discuss' (DEF=discuss|商討), and the concept 'discuss|商討' is linked with the following hierarchical sense relations:

(16) Hierarchical Relation of the Concept 'discuss|商討' in HowNet
　　　discuss, 商討, 1, {agent, partner, content}
　　　－event,事件－act,行動－ActSpecific,實動
　　　－AlterSpecific,實變－AlterState,變狀態
　　　－AlterMental,變精神－AlterKnowledge,變感
　　　－MakeOthersKnowledge,使他人感知
　　　－communicate,交流－discuss,商討

Basically, the two verbs are regarded as sharing the same concept without being further differentiated in the framework of HowNet.

2.5 Chinese Synonyms Usage Dictionary (近義詞用法詞典)

With a primary concern on near-synonym sets, *Chinese Synonyms Usage Dictionary* (Teng 1994), compares the two verbs in greater details and significant differences can be found. First, their definitions differ:

(17) Definitions in Dictionary of Near-synonyms:
　　　討論：to discuss something so as to establish pros and cons; talk over
　　　商量：to discuss something so as *to settle an issue or to achieve an aim*

Secondly, they may collocate with different grammatical patterns. Six differences can be found in the book:

(18) Collocational Variations between 討論 and 商量

	討論	商量
Modifier ～會	Yes	No
Noun 這一次～ 進行～	Yes	No
V ～出主意 ～辦法	No	Yes
VR ～好 ～得怎麼樣	No	Yes
Manner 熱烈～	Yes	No
Abab	No	Yes

As shown in the table, only 討論 may be nominalized to function as a modifier or head noun, and it may take a manner adjunct, while only 商量 may coöccur with certain resultatives. It appears that by focusing on near-synonyms, Teng (1994) is able to discover some important distinctions between the verbs. However, the above data are based mainly on intuitive judgments. We might well wonder whether corpus data will provide any new insights.

In the next section, findings based on corpus observations will be presented.

3. Corpus-based contrastive analysis

In The Sinica Corpus (as introduced above), there are 1,431 instances of 討論, but only 80 tokens of 商量. This huge difference in frequency of use readily indicates that 商量 might be semantically more restricted and marked, and hence more restricted in use. Observations of the corpus data also show that the two verbs are associated with quite different patterns in terms of grammatical functions, semantic properties of the participant roles, and manner modification.

3.1 Grammatical functions

In terms of their grammatical functions, the two verbs are mainly used as a predicate or as a noun (due to nominalization). Their distribution over the two major functions is shown below:

(19) Distribution over grammatical functions:

Grammatical Functional ＼ Verbs	討論 (1431)	商量 (80)
Predicate	1025 (72%)	77 (96%)
Nominalization	406 (28%)	3 (4%)

The predicate use of the two verbs (72% and 96%, respectively) is relatively higher than their nominal use (28% and 4%, respectively). As a predicate, they occur in three types of argument structure: without an overt object (20a), followed by an NP-object (20b), or followed by a clausal complement (20c):

(20) a. 我想跟你們 討論/商量 一下.
     'I'd like to discuss with you for a while.'
   b. 我想跟你們 討論/商量 [折扣的問題].
     'I'd like to discuss [the issue of discount] with you.'
   c. 我想跟你們 討論/商量 [該怎麼解決折扣的問題].
     'I'd like to discuss with you about [how to solve the issue of discount].'

The three environments for their predicate use are distributed as follows:

(21) Distribution of Argument Types

Argument type ＼ Verb	討論 (1431)	商量 (80)
+zero (no overt object)	553 (39%)	37 (46%)
+NP	359 (25%)	17 (21%)
+Clause	113 (8%)	23 (29%)

Among the three options for argument expression in (21), we note that, quite counter-intuitively, the two verbs are used most frequently as intransitives, without taking any overt object (39% and 46%, respectively). This seems to indicate that both verbs tend to profile the unique 'act' of two-way communication between interlocutors. The 'topic' or 'content' of the discussion, although semantically a default argument, may not be coded syntactically as a core argument.

Moreover, between the NP object and the clausal complement, 討論 prefers an NP over a clausal complement (25% vs. 8%), while 商量 shows a slightly higher percentage with nominal than with clausal complements (29% vs. 21%). As will be clear in the next section, these distributional tendencies appear to be correlates of their semantic distinctions.

Next, we consider the nominal uses of the two verbs. As shown in (19) above, 討論 has a much higher percentage of nominalization (28%), while 商量 is rarely nominalized (4%), confirming somewhat the intuitive judgment in (18) that only 討論, but not 商量, can be used as a nominal modifier (e.g. 討論會) or a noun (e.g. 這一次 討論). When nominalized, 討論 may be used freely to refer to the nominal activity pertaining to the process of discussion, taking a wide range of modifications, such as a durational modifier (22a), a manner modifier (22b), or a modification on topic (22c), but 商量 can only occur in a rare environment as a fixed expression (23):

(22) a. 長時間的討論
         'a long discussion'
     b. 熱烈的討論
         'a lively discussion'
     c. 對哲學的討論
         'a discussion on philosophy'
(23) 打個商量，好嗎？
         '(Let's) make a deal, OK?'

The different tendencies in nominalization indicate that 討論 may function as a full-fledged noun, referring to the activity nominal predicated by 討論, but 商量, on the other hand, may not function as a full-fledged noun since it cannot be used freely to refer to the nominal activity predicated by 商量. This distinction in nominalization may have something to with their distinct event structures, as will be discussed in the following sections.

In sum, the distributional variations on grammatical patterns, as shown above, are believed to correlate with and reflect some of the semantic properties of the verbs. In the next section, we shall take a closer look at the semantic details of their participant roles.

3.2 Semantic analysis of association patterns: Participant roles

As mentioned in 2.1 above, from the perspective of frame semantics, the event of 'discussion' inherits a common set of frame elements from the Conversation frame, characterizing the basic properties of a two-way communication. The frame elements are repeated here:

(10) Frame Elements in the Conversation frame
    a. Interlocutor 1
    b. Interlocutor 2
    c. Interlocutors
    d. Topic
    e. Medium

Observations of the grammatical behavior of 討論 and 商量 (3.1 above) confirms that the two verbs normally take an agent-subject, which may be characterized as the Interlocutors when referring to both parties of the discussants (e.g.「他們」一起討論), or Interlocutor 1 and Interlocutors 2 when the two parties are coded separately (e.g.「他們」和「我們」討論). And roughly half the time, they also take a theme-object (NP or clause), which corresponds to the Topic element. What is beyond the labeling of participant roles is that the two verbs display quite different association patterns in terms of the semantic types of Topics and Interlocutors they may take.

With regard to the Topic slot, when filled by an NP, certain restrictions are found for the use of 商量. In (24a), both 討論 and 商量 may take an NP-Topic that names some unsettled issue; but in (24b), when the NP-Topic refers to a generic, preëxisting agenda, 商量 cannot be used:

(24) a. Unsettled issues:
      討論/商量   結婚的事/作戰計畫
      'to discuss   the issue of marriage/the plan of war'
   b. Generic, predetermined agenda:
      討論/*商量   人我關係/語言學/文學
      'to discuss   interpersonal relation/linguistics/literature'

As for the Clausal-Topic, when a solution is sought as in (25a), both 討論 and 商量 can be used, but when the clausal topic talks about unpredictable future outcome as in (25b), 討論 is more likely used:

(25) a. A 'solution' is sought:
      討論/商量   如何克服目前的困境
      'to discuss   how to overcome the current difficulty'
   b. Unpredictable future:
      討論/*商量   戰爭對股市可能造成什麼影響
      'to discuss   what may be the potential influences of war to the stock market'

The semantic restrictions on the use of 商量 reminds us of the definitions put forth by Teng (1994), where the purpose of 'settling an issue or achieving an aim' is clearly specified for the meaning of 商量 (see 17 above), while 討論 encodes the meaning of 'exchanging ideas' without specifying a default purpose.

Besides taking different types of topics, the verbs differ in terms of the semantic prerequisite of the Interlocutors (discussants). For 商量, only those Interlocutors who have some kind of authority or some degree of control over the issue being discussed can occur as the subject. Thus, in (26a), both 討論 and 商量 can be used since the Interlocutors ('board of regents') are supposed to be the policy makers, but in (26b), 商量 is "dis-preferred", given the fact that 'laborers or manufacturers' in a company have little say or responsibility on policy making:

(26) Semantic Restrictions on Interlocutors
    a. 董事們開會　討論/商量　公司的經營策略
       'The board of directors have a meeting to discuss the company's policies.'
    b. 工人們開會　討論/*商量　公司的經營策略
       'The laborers have a meeting to discuss the company's policies.'

The requirement of agent involvement in the use of 商量 can been seen as an inferential consequence stemming from the inherent purpose of undertaking the activity of 商量. Since the purpose of 商量 is to seek a solution over an unsettled issue, the Interlocutors involved should at least have a say regarding the solution to be sought. In other words, the prerequisite for undertaking the activity of 商量 is the Interlocutors' personal commitment to taking a part or having a say in the possible solution being sought in the event of 商量. This also explains why, when nominalized, 討論 may refer to the objective process of exchanging ideas, but 商量 in 打個商量 implies personal give and take.

It is clear from the above discussion (Note: the term 'discussion' here can only be translated as 討論, not 商量) that although sharing the same set of participant roles, the two verbs differ in their semantic specifications on role-internal properties.

3.3 Semantic analysis of association patterns: Nominalization and modification

As also mentioned previously (3.1), the two verbs show different tendencies in nominalization and manner modification. The verb 討論 can undergo nominalization freely and function as the head of a possessive phrase as in (27a), can be modified by a postverbal manner complement as in (27b), or a preverbal manner adverb as in (27c):

(27) a. 他們的 討論/?商量 很熱烈
   'Their discussion is lively.'
   b. 他們 討論/?商量 得很熱烈
   'They discussed with enthusiasm.'
   c. 他們熱烈地 討論/?商量
   'They enthusiastically discussed.'

The wide range of adjectival and adverbial modifications allowed in the use of 討論 can be viewed as deriving from the fundamental difference in its event structure from that of 商量. Although both are process verbs that may take durational adjuncts, 討論 is a prototypical activity verb, encoding a simple, aspectually stable act of communication. Given its simple, homogeneous eventuality, 討論 may be used as a common noun to receive a variety of modifications. On the other hand, 商量 is semantically more complex, encoding a process with a 'semantic' endpoint. The verbal activity of 商量 is meant to achieve a goal or solution. When nominalized, 商量 tends to profile the semantic endpoint, rather than the verbal activity. This may be seen from the attested case of nominal 商量 in the Sinica Corpus, as exemplified in (23) above, repeated here:

(23) 打個商量，好嗎？
   '(Let's) make a deal, OK?'

The important point beyond our intuitive understanding of the verbs is that they are associated with different semantic requirements of the participant roles. In terms of the semantic characters of the Topic (an NP or a clause), 討論 takes a wider, unmarked range of topics, but 商量 is quite restricted in selecting topics: it only takes those that name some kind of a problem and call for a solution the discussants have a certain degree of control or authority over. The coming about of the solution depends completely on the solution-seeking process of 商量. Therefore, the topic associated with 商量 is highly marked and implicates personal involvement or commitment on the part of the discussants.

Their distinction in event types is summarized as follows:

(28) Event Structures
   a. 討論: a simple Process
   <Verbal Activity>: to exchange views on any TOPIC
   e.g. 討論問題 'to discuss a problem'

    b. 商量: a semantically bounded Process
        <Verbal Activity> + <a semantic endpoint>:
        to exchange views on an unsettled issue that calls for a SOLUTION the
        discussants are personally relevant to.
        e.g. 商量[對策] 'to come up with a solution'

3.4 Overt coding of a semantically implicated 'product'

The semantic implication of an end product (a solution) arising from the event of 商量 is not an isolated phenomenon as far as communication verbs are concerned. Another sub-set of communication verbs, verbs of explaining (解釋 or 說明), display a similar behavior. When nominalized, the verb 解釋 may refer to either the process of the verbal activity (the process of explaining) or the semantically implicated verbal product (the explanation arising from the process). In short, it may refer to a nominal activity (29a) or a nominal entity (29b):

(29) a. Nominal activity: the process of explaining
        三個小時的解釋還是不夠
        'Three hours of explaining isn't enough.'
    b. Nominal entity: the verbal product 'an explanation'
        我不接受[他的解釋].
        'I don't accept his explanation.'

In some sense, the end product derived from an explaining activity resembles the role of Incremental Theme (Dowty 1991), as it can be overtly coded as a separate argument from the topic-theme:

(30)   Theme   Incremental Theme
         | |
     將[他的提議] 解釋為 [自私自利的作法]
     '[His proposal] was interpreted as [a selfish move].'

The arising of some kind of an end product through the process of verbal communication seems to be characteristic of the group of communication verbs that semantically presuppose a salient purpose.

When communication verbs are nominalized, they may refer to different portions of the event. With a simple eventuality, such as 討論, nominalization normally refers to the on-going process, a nominal activity. With a complex, semantically bounded event,

such as 商量 or 解釋, nominalization may refer to the end product of the verbal activity. As the contrived example in (31) shows, the manner modifier 熱切地 'whole-heartedly' may characterize the process of explaining, focusing on the activity nominal, while the quality attribute 令人滿意的 'satisfactory' only applies to the verbal product:

(31) Nominalization with Different Event Focuses

[verbal activity]   [verbal product]

| |

經過熱切地[解釋]，他給了一個令人滿意的[解釋].

'Through explaining whole-heartedly, he gave a satisfactory explanation.'

With verbs of communication, the grammatical process of nominalization profiles different 'event focuses', that is, the focus of the nominalization, as revealed by the scope of the modifier, may fall upon different portions of the event structure.

4. Semantic representation with MARVS

To systematically represent the semantic distinctions in verbs, a representational scheme is proposed in Huang et al. (2000), namely, the Module-Attribute Representation of Verbal Semantics (MARVS). Under the assumption that each verbal sense corresponds to some unique eventive information, the model aims to translate all detectable verbal information into two categories: information pertaining to event types (the Event Module) vs. information pertaining to participant roles (the Role Module). And further semantic distinctions within each module can be coded as Event-internal or Role-internal attributes (See Huang et al. 2000 for detailed illustrations). The above-mentioned semantic distinctions characteristic of the Mandarin verbs of discussion (討論/商量) can be represented within the MARVS framework as follows:[2]

(32) Module-Attribute Representation for 討論/商量

Verb	Event Module Event Internal Attribute	Role Module Role Internal Attribute
討論	Simple process ////	\<Interlocutors> \<Topic>
商量	bounded process with a semantically implicated endpoint ////·	\<Interlocutors>: [+control] \<Topic>: [+solution]

---

[2] The framework of MARVS is still under development. The representation put forth here is a preliminary attempt to formulate the semantic distinctions.

The two verbs differ in event types, one is a typical process verb; the other is a semantically bounded process with a potential endpoint. As for core participants, they both require Interlocutors as agent and a Topic-theme. But 商量 imposes further semantic requirements on both roles. The agent of 商量 is specified with the feature [+control], to show that they have to hold authority over or personal commitment to the issue being discussed and the Topic for 商量 has to implicate a problem that awaits a solution.

5. Conclusions

Sharing the same semantic frame, the pair of verbs 討論 vs. 商量 appears to be indistinguishable at first sight, but careful examination of their corpus distribution reveals their semantic contrasts in event type and role-internal specifications. The semantic distinctions between the verbs bear significant consequences for their syntactic realizations. The approach demonstrated here proves to be useful in detecting lexical information of individual verbs as well as defining class-specific properties. Several generalizations can be drawn from the above discussions:

First, semantic contrasts between near-synonyms, based on different association patterns, may reveal crucial semantic characteristics regarding other verbal sets of the same domain.

Second, collocational skewings of verbal behavior need to be translated into semantic distinctions in linguistically describable terms. The information may be viewed as specifications pertaining to event structure (the Event module) and/or participant roles (the Role module).

Third, 'concepts' or 'conceptual units' in the sense of HowNet may be utilized to define semantic domains and frames for verb classification. But further characterization of frame-elements (Role-internal attributes) may be needed to capture the unique meaning components of individual verbs.

Fourth, the currently available semantic ontologies/taxonomies, constructed with a top-down approach for NLP purposes, do not equate to and cannot replace the linguistically based, semantic analyses of individual verbs and verb classes. The two approaches should complement each other in completing the mission of identifying and presenting lexical information that is grammatically relevant.

References

Baker, Collin F., Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. *Proceedings of the COLING-ACL*. Montreal, Canada.

Biber, Douglas, Susan Conrad, and Randi Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Dong, Zhendong. 1988a. Enlightment and challenge of machine translation. *Shanghai Journal of Translators for Science and Technology* 1988.1:9-15.

Dong, Zhendong. 1988b. Knowledge description: What, how and who? *Proceedings of International Symposium on Electronic Dictionary*. Tokyo, Japan.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67.3: 547-619.

Fillmore, Charles J. 1968. The case for case. *Universals in Linguistic Theory*, ed. by Emmon Bach and Robert T. Harms, 1-88. New York: Holt, Rinehart and Winston.

Fillmore, Charles J., and Beryl T. Atkins. 1992. Toward a frame-based lexicon: The semantics of RISK and its neighbors. *Frames, Fields, and Contrasts*, ed. by Adrienne Lehrer and Eva Feder Kittay, 75-102. Hillsdale, NJ: Lawrence.

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. *International Journal of Computational Linguistics and Chinese Language Processing* 5.1:19-46. Also appeared in *Proceedings of the Symposium on Selected NSC Projects in General Linguistics from 1998-2000*, 119-46, 2001.

Jackendoff, R. S. 1990. *Semantic Structures*. Cambridge: MIT Press.

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

Levin, Beth, and Malka Rappaport Hovav. 1996. From lexical semantics to argument realization. Manuscript. A revised and expanded version will appear in the Cambridge Research Surveys in Linguistics Series.

Liu, Mei-chun. 1999. Lexical meaning and discourse patterning—The three Mandarin cases of 'build'. *Cognition and Function in Language*, ed. by Barbara Fox, Dan Jurafsky, Laura Michaelis, 181-199. Stanford: CSLI.

Liu, Mei-chun. 2000. Categorical structure and semantic representation: Mandarin verbs of communication. Paper presented at the 5th Conference on Conceptual Structure, Discourse and Language. Santa Barbara: University of California.

Liu, Mei-chun. 2002. *Mandarin Verbal Semantics: A Corpus-based Approach*. Taipei: Crane Publishing Co.

Liu, Mei-chun. (to appear). Lexical information and beyond: Constructional inferences in semantic representation. *Journal of Chinese Linguistics*.

Liu, Mei-chun, 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. Cambridge: Harvard University.

Liu, Mei-chun, 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.

Meng, Cong, Huaide Zheng, Qinghai Meng, and Wenlan Cai. (eds.) 1985. *Dongci Yongfa Cidian* [*Verb Usage Dictionary*]. Shanghai: Shanghai Cishu Chubanshe.

Pustejovsky, James. 1995. *The Generative Lexicon*. Cambridge: MIT Press.

Teng, Shou-hsin. (ed.) 1994. *Chinese Synonyms Usage Dictionary*. Taipei: Crane Publishing Co.

Tenny, Carol. 1992. The aspectual interface hypothesis. *Lexical Matters*, ed. by Ivan A. Sag and Anna Szabolcsi, 1-27. Stanford: CSLI.

Tsai, Mei-chih, Chu-Ren Huang, Keh-jiann Chen, and Kathleen Ahrens. 1998. Towards a representation of verbal semantics. *International Journal of Computational Linguistics and Chinese Language Processing* 3.1:61-74.

Yu, Shiwen, Xuefeng Zhu, Hui Wang, and Yunyun Chang. 1998. *Xiandai Hanyu Yufa Xinxi Cidian Xiangjie* [*The Grammatical Knowledge-Base of Contemporary Chinese: A Complete Specification*]. Beijing: Tsinghua University Press.

Website Resources

FrameNet. http://www.icsi.berkeley.edu/~framenet/

HowNet (知網). http://www.keenage.com/html/c_index.html/

Sinica Corpus (中央研究院現代漢語平衡語料庫). http://www.sinica.edu.tw/ftms-bin/kiwi.sh/

Dept. of Foreign Languages and Literature &
Graduate Institute of Linguistics
National Chiao Tung University
1001 Ta Hsueh Road
Hsinchu 300, Taiwan
mliu@cc.nctu.edu.tw

從共現結構看事件訊息：
「討論類」動詞的詞彙語意研究

劉美君

國立交通大學

　　動詞語意的研究一直是語言學與自然語言處理及應用的一大課題。本文旨在綜論詞彙語意的剖析方法，並指出現有以處理大量訊息為目標的語意資料，無法呈現近義詞的語意區隔。唯有以語料庫為本，對照剖析動詞的共現結構、功能分布及語意限制，才能清楚界定語意和語法間的互動關係，進而釐清近義詞的關鍵屬性。動詞語意研究的最終目的即在找出具有影響語法表現的語意屬性。

關鍵詞：詞彙語意，中文動詞語意，討論類動詞，語料庫，語意屬性