

A Bidirectional Study of Mandarin Conversation Verbs*

Yiching Wu

d898702@oz.nthu.edu.tw

Institute of Linguistics National Tsing Hua University

Abstract

This study examines verbs of conversation, from two directions, bottom-up and top-down, e.g. 交談 *jiao1 tan2* ‘talk’, 商量 *shang1 liang2* ‘discuss’, 吵架 *chao3 jai4* ‘quarrel’, and 聊天 *liao2 tian1* ‘chat’ etc. In addition to the inductive bottom-up method, inducing generalization on the semantic properties of a lexical item by identifying its syntactic behavior and collocations, the deductive top-down approach, deducing semantic attributes from domain ontology is found to be helpful in systematically accounting for the linguistic phenomena.

1. Introduction

There are two common strategies used to determine truth from facts, induction and deduction. Studying lexical semantics is no exception. Linguists also probe lexicons from bottom-up or top-down perspectives.

1.1 Bottom-up approach: from lexical items to semantic fields

By following this approach, linguists may study from either a single lexical item (e.g. Fillmore and Atkins 1992), a pair or a set of near synonyms (e.g. Tsai et al 1996, 1998, Chief et al 2000, Liu et al 2000, Wu and Liu 2001, Liu 2000, 2002a, 2002b, 2003, etc.), or a class of lexical items (Chang et al 2000b, Lien 2001 & 2002, etc.) in order to capture the generalization of semantic components, constraints and rules for a semantic field, thereby constructing their theories. Generalizations may be derived from an observation of syntactic behavior and collocations of the items. The linguistic data may be collected from linguists’ own intuition, informants’ judgment, dictionaries, or from electronic thesauri e.g. WordNet (<http://www.cogsci.princeton.edu/~wn/index.shtml/>), and corpora such as British National Corpus (BNC) at <http://www.hcu.ox.ac.uk/BNC/>, and Sinica Corpus at <http://www.sinica.edu.tw/ftms-bin/kiwi.sh/>.

1.2 Top-down approach: from upper classes to lexical items

Using this approach, linguists start from an upper class, probe their way through the subclasses, and then to specific lexical entries. In general, the aim of this method is to facilitate language processing by constructing a taxonomy or ontology of the human lexicon. Semantic hierarchy and inheritance relations are the two main research targets. HowNet (http://www.keenage.com/html/c_index.html/) and Suggested Upper Merged Ontology

* This study is supported in part by NSC 91-2411-H-009-012-ME. I am indebted to my colleagues in the project and two anonymous reviewers for their insightful comments.

(SUMO at <http://ontology.teknowledge.com/>) are two of the online representatives. They contain a nearly complete hierarchy for Chinese and English words respectively. VerbNet (<http://www.cis.upenn.edu/verbnet/>) based on Levin (1993) and FrameNet I (<http://www.icsi.berkeley.edu/~framenet/>) are two of the other less exhausted cases. In Levin (1993), there are forty-eight verb classes grouped by a variety of syntactic alternations, but these classes are not structured by other upper classes. Though the concept of domains is obliterated in FrameNet II, FrameNet I contains fourteen domains with subordinate frames and lemmas, but these domains are not subsumed to other superior classes.

The problem of the bottom-up approach is that the semantic properties of each lexical item may be extracted and the overt syntactic behavior may be accounted for, but the inheritance relationship, with its parent and ancestor classes, remains opaque. In contrast, the problem surrounding the top-down approach is that the inheritance relationship among the different levels may be clear enough to account for the covert syntactic behavior, but the detailed semantic attributes may be missed. To compensate for this drawback, SUMO combines its ontology with WordNet synsets. (Pease et al 2002), and researchers are now pursuing a multi-lingual semantic network (Huang et al 2002). A prototype of the Chinese-English bilingual interface of general and domain-specific ontologies, constructed by the Chinese Knowledge Information Processing Group (CKIP), is now also available at <http://godel.iis.sinica.edu.tw/CKIP/ontology/>.

This study aims to provide a bidirectional approach, incorporating the above two methods in order to explore a detailed analysis of the finer semantic distinctions of conversation verbs.

2 Conversation verbs

To extract Chinese conversation verbs, several resources were consulted. Firstly, Conversation is one of the fourteen frames of the Communication domain in FrameNet I, and there are both Chinese and English words, as well as definitions, in HowNet. By retrieving the corresponding Chinese words and definitions of the English lemmas subsumed to the Conversation frame in FrameNet, a set of possible Chinese candidates is obtained. Secondly, the resultant set of candidates was checked with the lexical items in CKIP's Chinese-English bilingual ontologies. Any items that are used only in mainland China were temporarily ruled out. Thirdly, dictionaries, thesauri, and the intuition of native speakers were consulted. Finally, entries and their frequency in Sinica Corpus were taken into consideration. In this way, a set of target Chinese conversation verbs was obtained, e.g. 交談 *jiao1 tan2*, 談話 *tan2 hua4*, 會談 *hui4 tan2* 'talk', 閒聊 *xian2 liao2* 'gab' and 聊天 *liao2 tian1* 'chat', 交流 *jiao1 liu2*, and 溝通 *gou1 tong1* 'communicate', 商量 *shang1 liang2*, 討論 *tao3 lun4*, and 商討 *shang1 tao3* 'discuss', 吵架 *chao3 jia4* 'quarrel' and 爭辯 *zheng1 bian4* 'debate', etc.

After setting the target items, their syntactic behavior and collocations were probed. In

addition, their upper class, the domain of communication was also investigated. In what follows, we will first illustrate how the near synonyms were analyzed from a bottom-up approach and then elaborate on a top-down method.

3. Analysis of near synonyms

In this section, we will use three verbs of ‘talk/converse’ as an example to illustrate the bottom-up approach: *jiao1 tan2*, *tan2 hua4* and *hui4 tan2*, literally meaning ‘talk to each other’, ‘talk words’ and ‘meet and talk’ respectively.

3.1 Grammatical function distribution

As shown in table 1 below, sixty percent of the *jiao1 tan2* tokens function as a predicate, the main verb of a clause. In contrast, the majority of the tokens of *tan2 hua4* and *hui4 tan2* are used as a head noun. The three lexical items have approximately the same functional percentage as a modifier.

Lemma \ Function	交談 <i>jiao1 tan2</i>	談話 <i>tan2 hua4</i>	會談 <i>hui4 tan2</i>
Predicate	71 (60%)	57 (31%)	50 (33%)
Head Noun	30 (25%)	101 (54%)	82 (54%)
Modifier	17 (15%)	28 (15%)	20 (13%)
Total	118 (100%)	186 (100%)	152 (100%)

Table 1: Grammatical function distribution of *jiao1 tan2*, *tan2 hua4*, and *hui4 tan2*

3.2 Collocation

All three verbs can be modified by a duration, e.g. *Ta1 men jiao1 tan2/hui4 tan2 le shi2 fen1 zhong1* and *Ta1 men tan2 le shi2 fen1 zhong1 de hua4* ‘They have talked for ten minutes’. The three verbs can all collocate with the progressive (正 *zheng4*) 在 *zai4* and the experiential 過 *guo4*, e.g. *Ta1 men zheng4 zai4 jiao1 tan2/tan2 hua4/hui4 tan2* ‘They are talking to each other,’ and *Ta1 men jiao1 tan2 guo4 /tan2 guo4 hua4/hui4 tan2 guo4* ‘They have talked to each other.’ In addition, they can all be followed by the inchoative particle 了 *le*, e.g. *Ta1 men (kai1 shi3) jiao1 tan2/tan2 hua4/hui4 tan2 le!* ‘They start to talk!’ From the above facts, and by following the methodology used by Chang et al (2000a), we can induce the generalization that these verbs are bounded process verbs. However, these verbs contrast with ‘discuss’ verbs such as 商量 *shang1 liang2* and 討論 *tao3 lun4* in that they do not take a Topic directly, e.g. ‘**jiao1 tan2/*tan2 hua4/*hui4 tan2/shang1 liang2/tao3 lun4 shi4 qing2* ‘*converse/discuss about something’. Furthermore, they do not take a Message in the same manner as other saying verbs, e.g. ‘*Ta1 men *jiao1 tan2/*tan2 hua4/*hui4 tan2/shuo1 ta1 men mei2 you3 qian2* ‘They *conversed/*talked/said they had no money.’

In addition, the subject agent, the Speaker, of the three verbs must be plural, e.g. *Ta1*

*gen1 wo3 /wo3 men/ *wo jiao1 tan2 le ban4 xiao3 shi2/tan2 le ban4 xiao3 shi2 de hua4/hui4 tan2 le ban4 xiao3 shi2* ‘He and I /we/*I have talked for half an hour.’ This symbolizes the reciprocity of a conversation event, in which both the speaker and the listener do the speaking and listening. However, *hui4 tan2* differs from the other two in that its speakers are mostly officials. When *hui4 tan2* functions as a predicate, only 18% (9/50) of the Speakers are common people. Most Speakers (82%) are government officials, representatives of countries or parties, or school officials. In addition, among the nine instances of non-officials there are two doctor-and-patient pairs, and two businessmen pairs.

When the Speakers are realized as Interlocutor_1 and Interlocutor_2, being an argument in a matrix clause or in a subordinate clause as a pre-nominal modifier, they may be linked with or without an overt connective such as 與 *yu3*, 和 *he2/han4*, and 跟 *gen1*, e.g. 戈巴契夫與葉爾辛/美國國務卿貝克和伊拉克總統海珊/我們跟所有相關的人士/辜汪 *ge1 bal qi4 fu1 yu3 ye4 er3 xin1/mei3 guo2 guo2 wu4 qing1 bei4 ke4 he2/han4 yi1 la1 ke4 zong3 tong3 hai3 shan1/wo3 men gen1 suo3 you3 xiang1 guan1 de ren2 shi4/gu1 wang1* ‘Gorbacheve and Yeltsin/the American Secretary of State, James Baker, and the President of Iraq, Saddam Hussein/we and all the related people/Koo and Wang’. Among these three overt connectives and the covert linker, *gen1* is the most colloquial and is often used in daily conversation, whereas *yu3* and the covert linker usually appear in formal texts. There are seventy-one instances of Interlocutor_1 and Interlocuteor_2 using *hui4 tan2* in Sinica Corpus. The distribution of the four linking devices is shown in table 2 below.

pattern count	Interlocutor_1 conj. Interlocutor_2			Interlocutor_1 Interlocutor_2
	與 <i>yu3</i>	和 <i>he2/han4</i>	跟 <i>gen1</i>	covert linker
total	46	9	1	15

Table 2: Linking devices

Yu3 and the covert linker connect forty-six and fifteen pairs of speakers respectively. 和 *he2/han4* links nine, but 跟 *gen1* combines only one. This shows that *hui4 tan2* is a formal conversation event.

3.3 Lexical Distinctions Redefined as the MARVS Representation

The above generalizations can be represented by the Module-Attribute Representation of Verbal Semantics (MARVS) proposed by Huang and Ahrens (1999) and Huang et al (2000).

Module/Attributes	交談 <i>jiao1 tan2</i>	談話 <i>tan2 hua4</i>	會談 <i>hui4 tan2</i>
Event Module	●/////●	●/////●	●/////●
Inherent Attributes	[Reciprocal]	[Reciprocal]	[Reciprocal] [formal]
Role Module	<Speaker, Medium>	<Speaker>	<Speaker>
Role-Internal Attributes	[Plural] [language]	[Plural]	[plural][representative]

Fig. 1: MARVS Representation of the semantic differences among conversing verbs

From the above discussion, we can induce the following generalizations. Firstly, each of the above three items denotes a bounded process event which refers to a reciprocal communication activity. Since it is a reciprocal event, the Speaker role must have a minimum of two agents. Secondly, *hui4 tan2* is a more formal conversational event in contrast with the other two, and thus its Speakers tend to be representatives of a country or an organization. Thirdly, *jiao1 tan2* is inclined to take a language Medium whereas *tan2 hua4* and *hui4 tan2* do not. In addition, we know that the ‘talk/converse’ verbs do not collocate with a Topic as with the ‘discuss’ verbs, nor do they co-occur with a Message as with the ‘say’ verbs. However, as we cannot adequately account for them so far, we will attempt an alternate approach in the next section.

4. From a domain, frames, to subframes

In this section, we will take a top-down perspective to investigate the verbs of conversation. In FrameNet, there are fourteen frames within the domain of Communication. To capture the conceptual structure for understanding events in the domain of communication, Liu and Wu (2003) propose a schematic representation as shown in Fig. 2 below:

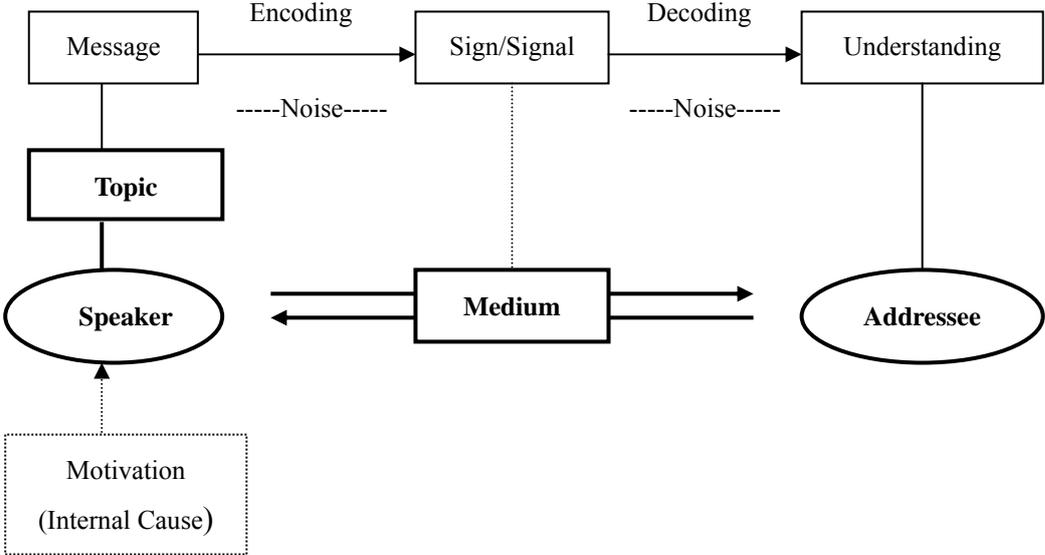


Fig. 2: Schematic Representation of Conversation

Communication in general is realized as an information-exchange process, where a Speaker, from certain motivation, sends a Message on a given Topic, through a process of packaging (Encoding), and an Addressee receives the package, decodes it, and reaches a certain understanding. The process is reciprocal and is carried out via a Medium (face-to-face, phone, TV, or email and fax, etc.).

Speaker, Addressee, Topic, Message, Sign/Signal, and Medium are the core frame

elements (FEs) of Communication. Each of the fourteen frames of Communication profiles certain frame elements. The Conversation frame focuses on the bilateral communication between the Speaker and the Addressee which are realized as Interlocutor_1, Interlocutor_2 and Interlocutors. Therefore, in addition to the three roles, only Medium and Topic are possible participant roles. The above schema may also account for the reason why Topic is not obligatory to all Chinese verbs, and Message is not a core element in the Conversation frame. Since the central focus is on the reciprocal communication process, Topic may not be profiled in every case, and Message may be suppressed.

Conversation verbs may be further classified into four subtypes according to their different purposes and manners:

Subframe	Purpose	Manner	Highlighted FEs
1 Converse	to exchange information	unmarked	Medium-language
2 Discuss	to solve a problem	serious	Topic
3 Quarrel	to exchange different opinions	heated	Cause
4 Chat	for fun	causal	Accompanying activities

Table 3: Subframes of Conversation

The Converse subframe is unmarked with a purpose to exchange information, e.g. *jiao1 tan2*, *hui4 tan2* ‘talk’, 交流 *jiao1 liu2* and 溝通 *gou1 tong1* ‘communicate’, etc. Hence, the Converse subframe verbs tend to co-occur with a language medium, e.g. *yi3 he2 lan2 hua4 jiao1 tan2* ‘converse in Dutch’. In the Discuss subframe, interlocutors communicate in a more serious manner in order to solve problems, e.g. 商量 *shang1 liang2*, 討論 *tao3 lun4* and 商討 *shang1 tao3* ‘discuss’, therefore the verbs tend to collocate with a Topic, e.g. *shang1 liang2 jie2 hun1 de shi4* ‘discuss a wedding affair’ and *tao3 lun4 nong2 ye4 wen4 ti2* ‘discuss issues on agriculture’. In the Quarrel subframe, interlocutors exchange different opinions in a heated manner, e.g. 吵架 *chao3 jia4* ‘quarrel’ and 爭辯 *zheng1 bian4* ‘debate’. Verbs in this subframe tend to collocate with a cause that results in the disagreement, e.g. *wei4 le qian2 chao3 jia4* ‘quarrel about money’. In the Chat subframe, interlocutors communicate in a casual manner for fun, e.g. 閒聊 *xian2 liao2* ‘gab’ and 聊天 *liao2 tian1* ‘chat’, etc., and hence the verbs tend to co-occur with accompanying recreational activities such as drinking coffee, e.g. *he1 kal fei1 liao2 tian1* ‘drink coffee and chat’.

From this point of view, the collocation of a Topic with ‘discuss’ verbs, as well as other highlighted participant roles in the subframes, may also be systematically accounted for.

5. Conclusion

The conversation verbs studied here serve to illustrate a hybrid approach to lexical semantics. The bottom-up approach provides a detailed generalization from studying specific lexical items. The top-down approach, aided by the domain schema, provides an overall outlook of the properties of the whole domain, helping to offer a systematic account for the linguistic phenomena. Although each of the methods has both positive and negative aspects, by incorporating the two approaches, detailed semantic features and outlined semantic properties can be expected.

References

- Chang, Li-Li, Keh-Jiann Chen and Chu-Ren Huang. 2000a. A Lexical-Semantic Analysis of Mandarin Chinese Verbs: Representation and Methodology. *International Journal of Computational Linguistics and Chinese Language Processing*. 5(1).1-18.
- Chang, Li-Li, Keh-Jiann Chen and Chu-Ren Huang. 2000b. Alternation Across Semantic Field: A Study of Mandarin Verbs of Emotion. *International Journal of Computational Linguistics and Chinese Language Processing*. 5(1).61-80.
- Chief, Lian-Cheng, Chun-Ren Huang, Keh-Jiann Chen, Mei-Chih Tsai, and Li-Li Chang. 2000. What Can Near Synonyms Tell Us? *International Journal of Computational Linguistics & Chinese Language Proceeding*. 5.1, 47-60.
- Fillmore, Charles J., and Atkins, Beryl T. 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. New Jersey: Lawrence.
- Huang, Chu-Ren, and Kathleen Ahrens. 1999. The Module-Attribute Representation of Verbal Semantics. *Working Papers on Chinese Verbal Semantics 1*, ed. by Kathleen Ahrens, Chu-Ren Huang, and Mei-Chih Tsai. 1-14. Taipei: Academia Sinica.
- Huang, Chu-Ren, I-Ju E. Tseng, and Dylan Tsai. 2002. "Translating Lexical Semantic Relations: The First Step Towards Multilingual Wordnets," presented at SemaNet'02: Building and Using Semantic Networks, A COLING2002 Post-Conference Workshop. Aug. 31. Academia Sinica. Taipei.
- Huang, Chu-Ren, Kathleen Athens, 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.
- Lien, Chinfa. 2001. Verbs of Saying in Li Jing Ji. The 4th International Conference on Classical Chinese Grammar. University of British Columbia, 15-17 August.
- Lien, Chinfa. 2002. Lexicalization and Grammaticalization in Taiwan Southern Min—A Case

- Study of Verbs of Commercial Transaction. To appear in *Joy of Linguistics*. Taipei: The Crane Publishing Co., Ltd.
- 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 & Chinese Language Proceeding*. 5.1, 81-96.
- 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. University of California, Santa Barbara.
- Liu, Mei-Chun. 2002a. Corpus-based Lexical Semantic Study of Verbs of Doubt: Huayi and Cai in Mandarin. *Concentric*. 28.2.
- Liu, Mei-Chun. 2002b. Mandarin Verbal Semantics: A Corpus-based Approach. 2nd ed. Taipei: Crane.
- Liu, Mei-Chun. 2003. From Collocation to Event Information: the Case of Mandarin Verbs of Discussion. To appear in *Language and Linguistics*.
- Liu, Mei-Chun and Wu, Yiching. 2003. Beyond Frame Semantics: Insight from Mandarin Verbs of Communication. Paper presented at the 4th Chinese Lexical Semantics Workshop. (第四屆漢語詞彙語義學研討會), Department of Chinese, Translation, and Linguistics, City University of Hong Kong, Hong Kong. June 22-July 11. (<http://icl.cityu.edu.hk/conference/4CLSW/BIG5/home.htm>)
- Niles, I., and Pease, A. 2001. Origins of the Standard Upper Merged Ontology: A Proposal for the IEEE Standard Upper Ontology. In Working Notes of the IJCAI-2001 Workshop on the IEEE Standard Upper Ontology, Seattle, Washington, August 6, 2001.
- Pease, A., Niles, I., and Li, J. 2002. The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications. In Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web, Edmonton, Canada, July 28-August 1, 2002.
- Tsai, Mei-Chih, Chun-Ren Huang, Keh-Jiann Chen, Kathleen Ahrens. 1998. Towards a Representation of Verbal Semantic: An Approach Based on Near-Synonyms. *International Journal of Computational Linguistics & Chinese Language Proceeding*. 3.1, 62-74.
- Tsai, Mei-Chih, Chun-Ren Huang, Keh-Jiann Chen. 1996. You Jinyici Bianyi Biao Zhun Kan Yuyi, Jufa Zhi Hudong (由近義詞標準看語義、句法之互動 From near-synonyms to the interaction between syntax and semantics), paper presented at IsCLL-5, Taipei, Taiwan.
- Wu, Yi-Ching, and Liu, Mei-Chun. 2001. The Semantic Distinction and Information Representation of Psychological Verbs: Xiang, Renwei, Yiwei and Juede—a Corpus Based Analysis. *Proceedings of Research on Computational Linguistics Conference XIV*. 317-336. Tainan: National Cheng Kung University.