

行政院國家科學委員會專題研究計畫 期中進度報告

中文動詞詞彙訊息的建構：認知感官動詞研究(1/2)

計畫類別：個別型計畫

計畫編號：NSC94-2411-H-009-014-

執行期間：94年08月01日至95年07月31日

執行單位：國立交通大學外國語文學系

計畫主持人：劉美君

共同主持人：梁婷

計畫參與人員：周明輝、江亭儀、徐雅苓、林姿君、胡佳音、鄭守益

報告類型：精簡報告

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫可公開查詢

中 華 民 國 95 年 6 月 2 日

行政院國家科學委員會專題研究計畫研究報告-期中進度報告

中文動詞詞彙訊息的建構：認知感官動詞研究(1/2)

Construction of the Mandarin VerbNet: Cognition-Perception Verbs

計畫編號：NSC 94-2411-H-009-014

執行期限：94 年 8 月 1 日至 95 年 7 月 31 日

主持人：國立交通大學外國語文系 劉美君

共同主持人：國立交通大學資訊科學學系 梁婷

計畫助理：周明輝、江亭儀、徐雅苓、林姿君、胡佳音、鄭守益

摘要

本計畫為兩年計劃中的第一年研究，著重於認知動詞。本計劃團隊在現有的「中文動詞詞網」基礎下，延續過去幾年對「溝通動詞」14 個框架及次框架的研究，繼續分析中文認知動詞的語法、語意特徵，釐清由「認知領域」(Cognition Domain)到此領域下的六個「框架」，再到「次框架」等的概念結構及其體現在語言當中的句法與語意的互動關係。

關鍵詞：中文動詞詞網、認知動詞、框架語意、次框架、概念結構、句法與語意互動

Abstract

This project is part of a 2-year research program on the construction of Mandarin VerbNet. Based on the research of Mandarin verbs of communication, the research group continues to study Mandarin verbs of cognition, with regard to their semantic and syntactic interdependencies. In this report, a frame-based representation of the cognition domain is provided, with a preliminary analysis of the six frames and their subframes.

Keywords: Mandarin VerbNet, Cognition verbs, Frame Semantics, subframes, conceptual structure, syntax-semantics interface

1 Introduction

1.1 Goal of Research

It is observed that cognition verbs and perception verbs have a lot in common with communication verbs (e.g. the discussion of Perception-cognition-utterance verbs in Givón (1993)). Based on the research of Mandarin communication verbs, the project attempts to further investigate the semantic and syntactic properties of Mandarin cognition and perception verbs. A preliminary analysis of Mandarin cognition verbs are presented in Section 2.

1.2 Literature Review

The study of lexical semantics has been growing with the need of analyzing and representing human knowledge. A number of lexically-based information networks have been constructed, such as WordNet (Miller et al. 1990), HowNet (Dong et al.), SUMO (Das et al 2002, Pease et al 2002, Niles and Pease 2003), Sinica BOW (Huang et al.), and FrameNet (Baker, Fillmore and Cronin 2003). While all these networks serve as valuable resources for knowledge representation, only FrameNet is based on linguistic theory of Frame Semantics (Fillmore and Atkins 1992). However, the structure of FrameNet does not include a detailed account of the interrelations of the proposed frames. To meet the gap, Liu and Wu (2003) proposed a five-layered hierarchical structure:

Domain -> Frame -> Subframe -> Near-synonym Set -> Lemma

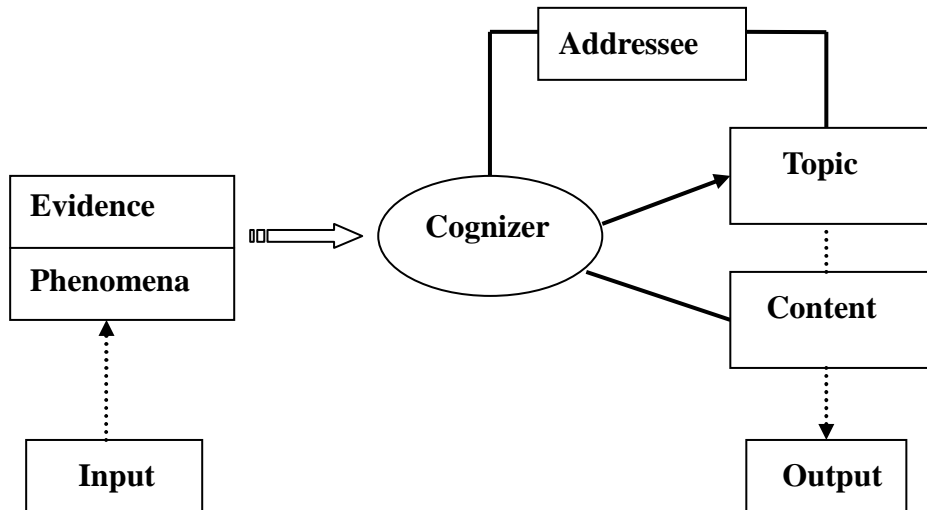
Based on this model, Mandarin communication verbs (or verbs in the Communication Domain) were classified into 14 Frames, which were further grouped into subframes with Near-synonym sets, and finally individual lexical items are presented (Liu and Huang 2004a, 2004b).

1.3 Research Method

The English lexical network FrameNet is taken as the source of lemma extraction. The English verbs in each frame are put into Sinica Bow (<http://bow.sinica.edu.tw/wn/>) and Chinese equivalents are the basis of the analysis. Based on the semantic and syntactic properties of the verbs in each frame, they are classified into subframes and finer layers such as near-synonym pairs.

2 The Cognition Domain

2.1 Schematic Representation



2.2 Domain Information

There are five frames in the Cognition Domain: 1) *Becoming_aware*, 2) *Cogitation*, 3) *Becoming_to_believe*, 4) *Awareness*, 5) *Reasoning* and 6) *Certainty*. The description of these frames is as follows.

3 Frames in the Cognition Domain

3.1 *Becoming_aware* Frame

3.1.1 Definition from FrameNet

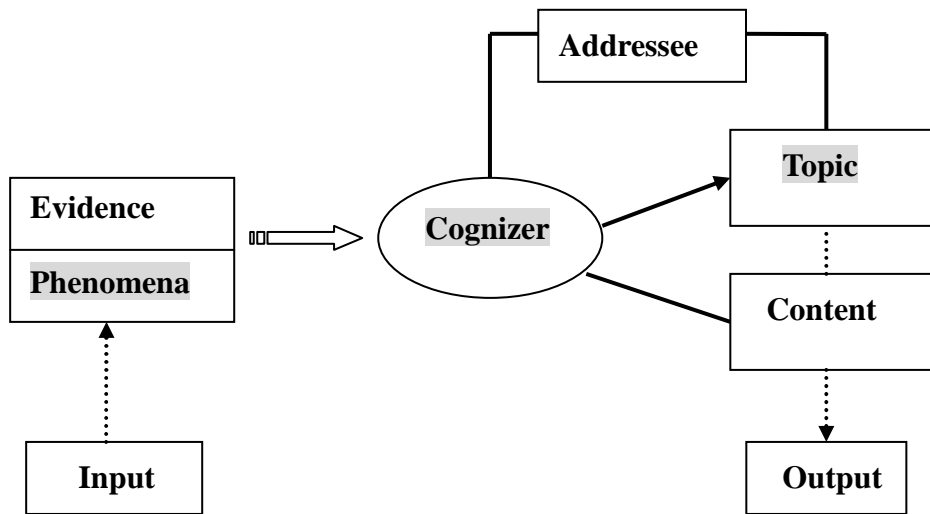
Words in this frame have to do with a Cognizer adding some Phenomenon to their model of the world. They are similar to *Coming-to-believe* words, except the latter generally involve reasoning from Evidence. The words in this frame take direct objects that denote entities in the world, and indicate awareness of those entities, without necessarily giving any information about the content of the Cognizer's belief or knowledge. These words also resemble perception words, since creatures often become aware of things by perceiving them.

3.1.2 Core Frame Elements

The core frame elements in this frames are: Cognizer, Phenomenon, Topic, Instrument and Means.

3.1.3 Schematic Representation

In this frame, some phenomena are awarded and these phenomena are not production of the event i.e. incremental theme but an existed real entity in the world.



3.1.4 Chinese lemma

The lexical items in this frame are 辨認、認出、認得、注意、注意到、發現、發覺、察覺、意識到、想到、想起。

3.1.5 Distribution of Grammatical Function

The Subframe-Come_(upon) does not have representation of nominalization and only have the function of predicate.

	Predicate	[+nom]	modifier
認得/辨認 (VJ 狀態及物)	89.66%	8.05%	2.3%
發現/注意 (VE 動作句賓)/(VK 狀態句賓)	92%	5.25%	2.75%
想到/想起 (VE 動作句賓)	100%	0%	0%

3.1.6 Distribution of Core Frame Element

	Cognizer	Phenomenon [NP]	Phenomenon [CL]	Phenomenon as human	Instrument	Means	Topic	[+nom]
認得 / 辨認	78.16%	70.11%	0%	44.26%	1.15%	1.15%	2.3%	10.34%
發現 / 注意	98.37%	20.5%	49%	1.8%	0%	1.5%	20.75%	8%
想到 / 想起	100%	25.75%	51.5%	12.3%	0%	0%	22.25%	0%

3.2 Cogitation Frame

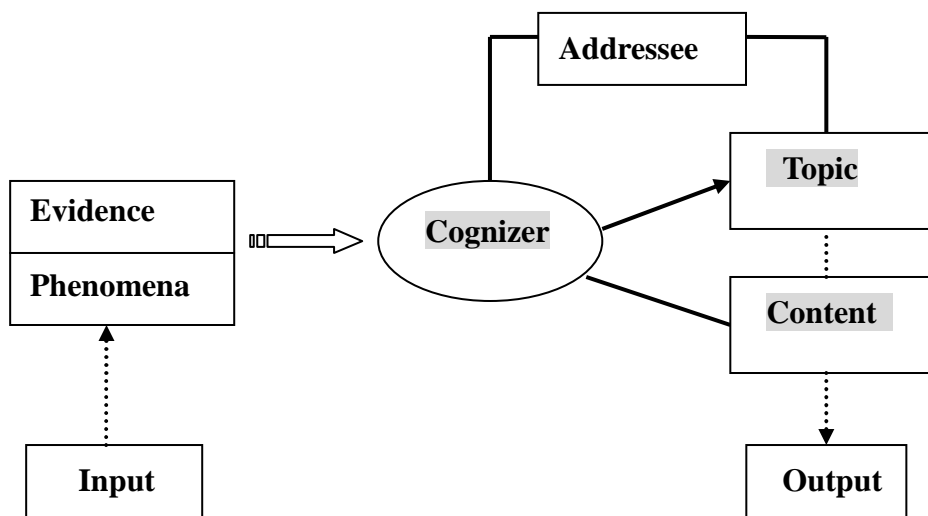
3.2.1 Definition from FrameNet

A person, the Cognizer, thinks about a Topic over a period of time. What is thought about may be a course of action that the person might take, or something more general.

3.2.2 Core Frame Elements

The core frame elements in this frames are: Cognizer, Content and Topic.

3.2.3 Schematic Representation



3.2.4 Chinese lemma

The lexical items in this frame are 思考、思索、思忖、思量、沉思、深思熟慮、考慮、推敲、斟酌、權衡、打算、盤算、計畫、想、覺得、認為。

3.2.5 Distribution of Core Frame Element

The distribution of the core frame elements in this frame is provided in the following table with examples extracted from the Sinica Corpus.

	Cognizer	Content	Topic
考慮	68/200 (34%)	35/200 (17.5%)	86/200 (43%)
思考	90/200 (45%)	11/200 (5.5%)	28/200 (14%)
想	57/200 (28.5%)	41/200 (20.5%)	18/200 (9%)
覺得	118/200 (59%)	139/200 (69.5%)	4/200 (2%)
認為	137/200 (68.5%)	179/200 (89.5%)	2/200 (1%)
計劃	11/200 (5.5%)	14/200 (7%)	2/200 (1%)

3.3 Becoming_to_believe Frame

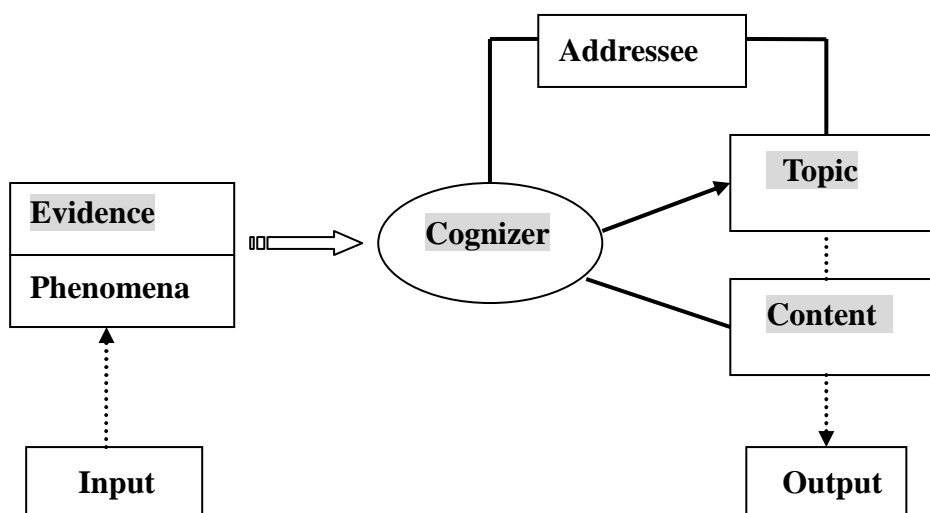
3.3.1 Definition from FrameNet

A person (the Cognizer) comes to believe something (the Content), sometimes after a process of reasoning. This change in belief is usually initiated by a person or piece of Evidence. Occasionally words in this domain are accompanied by phrases expressing Topic, i.e. that which the mental Content is about.

3.3.2 Core Frame Elements

The core frame elements in this frames are: Cognizer, Content, Topic, Evidence, Incremental Theme.

3.3.3 Schematic Representation



3.3.4 Chinese lemma

The lexical items in this frame are 推測、推論、推斷、推導、查明、確定、確認.

3.3.5 Distribution of Grammatical Function

	Predicate	Head Noun	Modifier	Total
推測/推論	63%	26%	11%	100%
確認/確定	60%	31%	9%	100%

3.3.6 Distribution of Core Frame Element

	Cognizer	Evidence	Topic	Incremental theme	Content
推測/推論	100%	62%	20%	5%	88%
確認/確定	95%	13%	25%	0%	45%

3.4 Awareness Frame

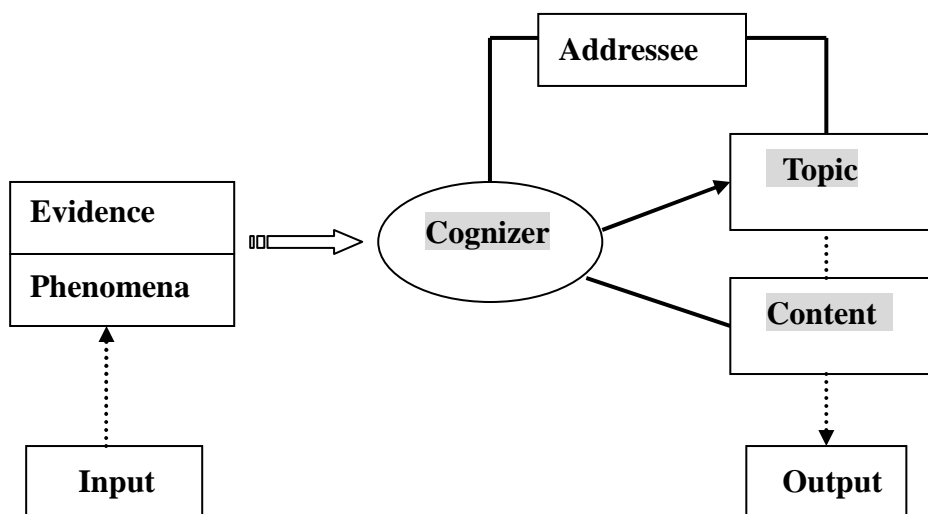
3.4.1 Definition from FrameNet

The words in this frame have to do with the presence of some Content in the awareness, knowledge or beliefs of a Cognizer. Verbs in this frame profile the process that something, normally content occurs to the Cognizer.

3.4.2 Core Frame Elements

The core frame elements in this frame are Cognizer, Content and Topic.

3.4.3 Schematic Representation



3.4.4 Chinese lemma

The lexical items in this frame are 回憶、回想、回顧、記起、知道、了解、瞭解、懂、諒解、覺得、認為、感覺、以為、想。

3.4.5 Distribution of Grammatical Function

Subframe / Function	Predicate 謂語	Head Noun 名詞組中心語	Modifier 定語	Total
記憶/記起	68.8%	29.2%	2%	100%
懷疑/知道	93.5%	4%	2.5%	100%
覺得/認為	100%	0%	0%	100%

3.4.6 Distribution of Core Frame Elements

Frame Elements / Lemma	Cognizer	Content[CL]	Human Topic[NP]	Topic
記憶/記起	100%	63.1%	0%	37.4%
知道/了解	94.1%	77.5%	0.5%	2.5%
覺得/認為	99.8%	99%	0%	0%

3.5 Reasoning Frame

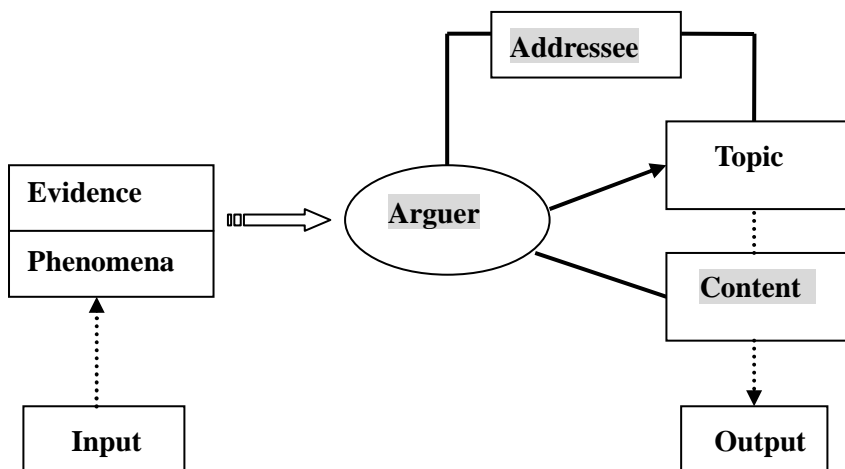
3.5.1 Definition from FrameNet

An Arguer presents a Content, along with Support, to an Addressee. The Content may refer elliptically to a course of action or it may refer to a proposition that the Addressee is to believe. Some lexical units (e.g. "prove") indicate the speaker's belief about the Content.

3.5.2 Core Frame Elements

The core frame elements in this frames are: Arguer, Content, and Addressee.

3.5.3 Schematic Representation



3.5.4 Chinese lemma

The lexical items in this frame are 爭辯、理論、辯論、論證、證明、駁斥、爭論.

3.6 Certainty Frame

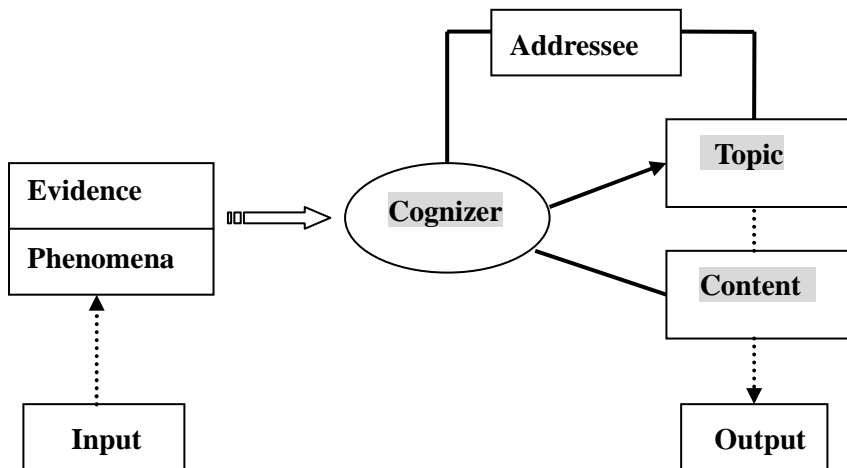
3.6.1 Definition from FrameNet

This frame concerns a Cognizer's certainty about the correctness of beliefs or expectations. It only includes uses where a Cognizer is expressed.

3.6.2 Core Frame Elements

The core frame elements in this frames are: Cognizer, Content, and Topic.

3.6.3 Schematic Representation



3.6.4 Chinese lemma

The lexical items in this frame are 相信、懷疑、疑惑、信任、信賴、確信、確定.

3.6.5 Distribution of Grammatical Function

Function Lemma	Predicate 謂語	Head Noun 名詞組中心語	Modifier 定語	Total
相信/懷疑	95.44%	3.61 %	0.95%	100 %
確定/確信	97.37%	1.48 %	1.14 %	100 %
信任/信賴	65.05 %	31.55 %	3.4 %	100 %

3.6.6 Distribution of Core Frame Elements

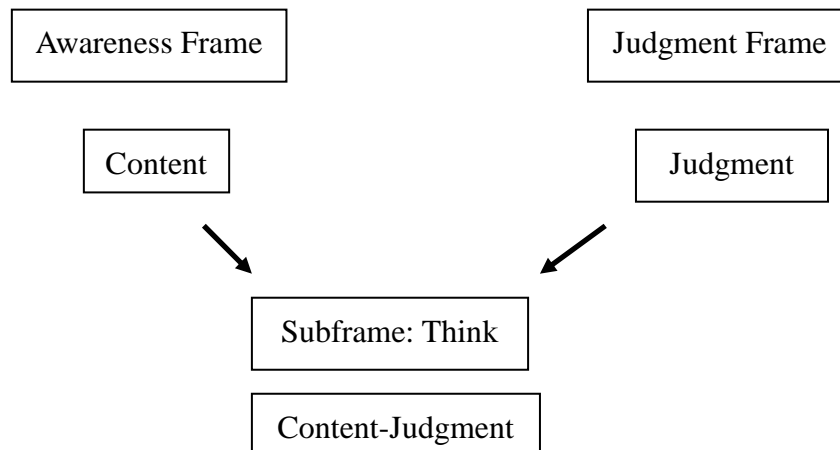
Frame Elements Lemma	Overt Cognizer	Content	Content (as human)	Topic
相信/懷疑	70.52 %	88.71 %	3.9 %	4 %
信任/信賴	62.48 %	71.61 %	81.1%	5.59 %
確定/確信	59.92 %	85.28 %	0	1.71 %

4 Conclusion and Discussion

One of the intriguing phenomena in our research project is “multiple inheritance”, where verbs of a certain subframe “inherit” semantic and syntactic characteristics from two frames. This notion is also discussed in our previous studies on the communication verbs, where the Quarreling Subframe presents semantic and grammatical features of the Hostile_encounter Frame and Conversation Frame.

In terms of cognition verbs, the Think Subframe in the Awareness Frame also shows multiple inheritance. That is, the verbs of thinking inherit “Content” from the Awareness

Frame and “Judgment” from the Judgment Frame. As a result, In the Think subframe, the FE ‘Content-Judgment’ is the **evaluation** occurs to the Cognizer which is usually realized as 1) NP + 為/是/就是+ NP/AdjP; 2) NP + AdjP. The multiple inheritance is illustrated below.



In sum, a detailed analysis of Mandarin cognition verbs serves as a valuable basis for the construction of Mandarin VerbNet, a resource providing rich information with regard to the fine-grained semantic and syntactic interdependencies of Mandarin verbs.

Reference

- Baker, Collin F., Charles J. Fillmore and Beau Cronin 2003. The Structure of the Framenet Database. *International Journal of Lexicography* 16(3).281-296.
- Das, Subrata, Kurt Shuster, and Curt Wu. 2002. Ontologies for Agent-Based Information Retrieval and Sequence Mining. *Proceedings of the Workshop on Ontologies in Agent Systems (OAS02)*, held at the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems, Italy, July 15-19.
- 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.
- Kennedy, C. and Levin, B. 2002. “Telicity Corresponds to Degree of Change” Handout to Speech at Georgetown University.
- Liu, Mei-Chun and Yiching Wu. 2003. Beyond Frame Semantics: Insight from Mandarin Verbs of Communication. Paper presented at the 4th Chinese Lexical Semantics Workshop. City University of Hong Kong, Hong Kong. June22-July 11. (<http://icl.cityu.edu.hk/conference/4CLSW/BIG5/home.htm>)
- Miller, A., R. Beckwith, C. Fellbaum, D. Gross, K.J. Miller. 1990. Introduction to WordNet: An on-line Lexical Database. *International Journal of Lexicography*. 3.235-244.

Niles, Ian and Adam Pease. Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. Proceedings of the 2003 International Conference on Information and Knowledge Engineering (IKE'03), Las Vegas, Nevada, June 23-26, 2003.

Pease, A., Niles, I., and Li, J. 2002. The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications. Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web, Edmonton, Canada, July 28-August 1, 2002.

劉美君、黃居仁，2004a，【意義與詞意】系列之五－中文動詞詞網的建構(2)談話類動詞的框架與次框架，交通大學語言與文化研究所知識化語料庫研究室、中央研究院語言學研究所文獻語料庫研究室。

劉美君、黃居仁，2004b，【意義與詞意】系列之五－中文動詞詞網的建構(1)總論及溝通動詞的框架與階層，交通大學語言與文化研究所知識化語料庫研究室、中央研究院語言學研究所文獻語料庫研究室。

Website resources:

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

HowNet: http://www.keenage.com/html/c_index.html/

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

SUMO Ontology: <http://ontology.teknowledge.com/>

WordNet: <http://www.cogsci.princeton.edu/~wn/>

Sinica BOW: <http://bow.sinica.edu.tw/>

Evaluation:

There are a total of 6 conference papers related to the current project:

(A) 14th Annual Conference of the International Association of Chinese Linguistics & 10th International Symposium on Chinese Languages and Linguistics (IACL-14 & IsCLL-10): 4 papers and posters

- Liu, Meichun and Tzi-chun Lin. 2006. Functional Distinction Between Zai, Zhengzai, and Zheng: Evidence from Collocational Association with Adverbs. Paper presented in the IACL-14 & IsCLL-10 Joint Conference, Academia Sinica, Taipei, 5/25-29/2006.
- Liu, Meichun and Tingyi Chiang. 2006. Perception for Cognition: interaction of subjectivity and intersubjectivity in spoken discourse. Paper posted in the IACL-14 & IsCLL-10 Joint Conference, Academia Sinica, Taipei, 5/25-29/2006.
- Liu, Meichun and Yaling Hsu. 2006. A Hybrid Model for Polysemy: insight from Mandarin verbs. Paper posted in the IACL-14 & IsCLL-10 Joint Conference, Academia Sinica, Taipei, 5/25-29/2006.
- Chou, Ming-hui and Mei-chung Liu. 2006. Semantic Extension and Polysemous Near-Synonyms: Two Case Studies in Mandarin. Paper posted in the IACL-14 &

IsCLL-10 Joint Conference, Academia Sinica, Taipei, 5/25-29/2006.

(B) The 18th North America Conference on Chinese Linguistics (NACCL-18): **1** paper presentation

- Liu, Meichun and Tingyi Chiang. 2006. Paper presentation at NACCL-18.

(C) Lexis and Grammar Conference 2006: **1** abstract submission

- Chia-yin Hu. 2006. Constructional Analysis of the ABAC lexical-grammatical pattern. Paper submitted to Lexis and Grammar Conference 2006.