94　6　22

行政院國家科學委員會補助專題研究計畫成果報告

計畫名稱：

語意範疇與詞彙分類：英文與中文常用詞彙對比分析
(Semantic Categories and Word Classification:
A Contrastive Analysis of the English and Chinese Lexicons)

成果報告類型(依經費核定清單規定繳交)：精簡報告

本成果報告包括以下應繳交之附件：無

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究
計畫、列管計畫及下列情形者外，得立即公開查詢

執行單位：國立台灣大學外國語文學系暨研究所

中華民國 93 年 12 月 31 日

（一）　　計畫中文摘要

關鍵詞：語意範疇、詞彙分類、詞庫、詞彙語意學、知識本體

　　英文詞彙之語意範疇研究，應以 1852 年出版之 *Roget's Thesaurus of English Words and Phrases* 為最早。這本劃時代的詞庫的語意範疇是十九世紀的一些世界觀，共分六類：抽象關係(Abstract Relations)、空間(Space)、物質(Matter)、知識(Intellect)、意志(Volition)、情緒(Affections)。每一大類之下又細分許多次類，每次類之下又細分許多小類，在 P. M. Roget 之原版中共有一千個小類。這是根據英文語意範疇所建立的第一個知識本體(ontology)。

　　自從 1970 年以來，研究詞彙語言學(lexical semantics)之學者及認知語言學學者也開始注意知識本體與詞彙分類的關係。WordNet、SUMO 及中研院之 BOW 均是此種研究的與料庫。

　　本計畫因為主要目的是將《大考中心高中參考詞彙表》之英文詞彙與其相對應中文詞彙加以分析，以期建立一套語義範疇系統，反應《大考中心高中參考詞彙表》約六千多英文詞彙與其相對應中文詞彙之知識本體。

　　本研究計畫之成果是把上述中英文詞彙分成五十大類之語義範疇，其下再依每一語義範疇之特性分若干結構層次(hierarchical levels)。此種語義範疇及其結構層次對於英文及中文之詞彙分類極有助益，同時對於了解中英文之知識結構頗有貢獻。

　　本研究計畫也以上述五十大類之語義範疇及結構層次來檢驗 Rosch 及 Lakoff 所提出之原型效應(prototype effects)及基本詞彙效應(basic-level effects)，發現他們的這兩個觀念對於分類結構層次(taxonomic hierarchy)之語義範疇頗有啟發性，但是對於部分與整體關係之結構層次(meronymic hierarchy)則不能適用。此項發現對於認知語言學頗有貢獻。

　　本研究計畫所建立之五十個中英文語義範疇及其結構層次可以應用於中英文教學，因為中英文之詞彙雖然大致對應，但是有的英文詞彙中文沒有，有的中文詞彙英文沒有。本研究計畫所指出之中英文詞彙之異同是中英文語言教學必須注意之重點，有助於提升中英文教學之效果。

(二)計畫英文摘要

Semantic Categories and Word Classification:
A Contrastive Analysis of the English and Chinese Lexicons

Key words: semantic categories, word classification, thesaurus, lexical semantics, ontology

The first English thesaurus was *Roget's Thesaurus of English Words and Phrases*, which was published by Peter Mark Roget (1779-1869) in 1852. Different from the traditional English dictionaries, which arranged words alphabetically, this thesaurus arranged English words according to their semantic categories in terms of synonyms and antonyms. *Roget's Thesaurus of English Words and Phrases* has six major semantic categories: Abstract Relations, Space, Matter, Intellect, Volition, and Affections. Under each major semantic categories, there are many sub-categories, amounting to a total of 1,000 sub-sub-categories. This was the first ontology constructed for the English language.

Since the 1970s, many scholars in the field of lexical semantics and cognitive science have also attempted to construct models of ontology, and WordNet, SUMO and BOW of Academia Sinica of Taiwan are the most notable achievements.

This project aims to categorize the approximately 6,000 English words in *College Entrance Examination High School English Word List* and their Chinese counterparts into semantic categories so as to construct an ontological model for these two languages for pedagogical purposes.

This project has now classified these English and Chinese words into 50 superordinate semantic categories, and under these superordinate semantic categories various hierarchical levels have also been postulated. These 50 superordinate semantic categories together with their various hierarchical levels are helpful to understanding the ontological structures of these two languages.

This project has also attempted to verify the cognitive concepts of prototype effects and basic-level effects put forward by Rosch and Lakoff with the above findings of semantic categories and their various hierarchical levels. It has been found that their cognitive concepts do apply to semantic categories with taxonomic hierarchies, but not to those with meronymic hierarchies.

The 50 semantic categories together with their hierarchical levels established by this project will also be useful for teaching English and Chinese, because some lexical gaps in English and Chinese found in this project will call for greater attention in learning either language.

Semantic Categories and Word Classification:
A Contrastive Analysis of the English and Chinese Lexicons

Hengsyung Jeng
鄭恆雄

台灣大學外文系

1 Introduction

Ferdinand de Saussure in his posthumous book *Cours de linguistique generale* (1916) first pointed out that the lexical items in a language enter into both paradigmatic and syntagmatic relationships in its linguistic network. This concept of his has since been the guideline for investigating the semantic relations in the lexicon of a language. But during the booming period of Chomsky's generative enterprise in the 1960s and 1970s, syntagmatic relationships were spotlighted, relegating paradigmatic semantic relations of the lexicon to the dark background. However, since the 1970s, more and more linguists (Cruse 1986/2004; Palmer 1976; Lyons 1977; Jackendoff 1990; Ravin 1990; Nirenburg and Levin 1992; Pustejovsky 1995; Fellbaum 1999; Miller 1999; Huang et al. 2000; Liu et al. 2000; Ahrens et al. 2003; Cruse 2004; Fillmore 2004; Nirenburg and Raskin 2004) and cognitive scientists (Berlin and Kay 1969; Berlin et al. 1974; Ekman 1971; Jackendoff 1983; Rosch 1973/1975/1877/1978/1981; Kay and McDaniel 1978, Lakoff 1980/1987/1999) have revisited paradigmatic relationships in the lexicon in terms of the concepts of categorization, prototypes and hierarchy and discovered that the network of the lexicon is closely linked to human conceptual structures which represent our knowledge of the world.

Current studies on the network relationships of the lexicon, or lexical semantics, still follow the two approaches put forward by Saussure: (1) syntax-driven lexical semantics (Cruse 1986; Jackendoff 1990; Ravin 1990; Nirenburg and Levin 1992; Pustejovsky 1995; Huang et al. 2000; Liu et al. 2000; Ahrens et al. 2003; Cruse 2004; Fillmore 2004;); (2) ontology-driven lexical semantics (Cruse 2004; Fellbaum 1999;

Huang et al. 2003; Miller 1999; Murphy 2003; Nirenburg and Levin 1992; Nirenburg and Raskin 2004). The former is mainly concerned with such lexical information as argument structures of predicates and their mappings onto syntactic structures (Nirenburg and Levin 1992: 7), whereas ontology-driven lexical semantics focuses on constructing models of ontology (world views/conceptual knowledge of the world) and semantic categories in terms of the network of lexical items. (Nirenburg and Levin 1992: 9-10) But these two approaches need not be mutually exclusive. As a matter of fact, they can be complementary. (Nirenburg and Levin 1992: 20)

This study on semantic categories and word classifications in English and Chinese follows the framework of ontology-driven lexical semantics. It will examine 6,315 English content words (nouns, verbs, adjectives and adverbs)[1] of the *College Entrance Examination Center High School English Word List* (Jeng et al. 2002, henceforth "*CEECEWL*") (excluding 165 function words such as articles, pronouns, prepositions and conjunctions)[2] and their Chinese counterparts to find out the superordinate semantic categories for classifying these words, and then compare the Chinese counterparts of these English words to ascertain whether such superordinate semantic categories are also suitable for classifying these Chinese words or not. The similarities and differences found in this contrastive study of the English and Chinese word lists will serve as the basis for constructing an ontology applicable to the word lists of these two languages. The cognitive model of prototype as initiated by Rosch and expounded by Lakoff will be tentatively used to categorize the English and Chinese words under study, and semantic relations such as hyperonymy, hyponymy, synonymy, contrast, antonymy, meronymy, polysemy, metaphor and metonymy will be the devices to verify the various levels of semantic categories.

2. Previous Ontological Models:

2.1 *Roget's Thesaurus of English words and phrases*

Roget's Thesaurus of English words and phrases (1852) was perhaps the first

---

[1] The adverbs derived from the adjective stems plus "-ly" are not included.

[2] The *CEECEWL* contains altogether 6,480 words, which, according to my statistics, can cover at least 98% of all the words used in the famous newspapers such as *New York Times*, *Washington Post*, news magazines such as *Time* and *Newsweek*, and broadcast materials on *BBC*, *CNN* and *VOA*.

attempt to categorize English vocabulary according to an ontological model or conceptual system of the knowledge of the world at that time. Owing to the limitations of world knowledge and philosophical frameworks, his 6 superordinate categories of "Abstract relations," "Space," "Matter," "Intellect," "Volition," and "Emotion, religion and morality" were mainly based upon the world views of Francis Bacon, Descartes and Leibniz. (Lyons 1977 vol. 1:300) Over the past 150 years, his ontological model has become obsolete and hence it has almost completely been revamped in *Roget's International Thesaurus* (Kipfer 2001). Kipfer (2001) postulates 15 superordinate categories: (1) The Body and the Senses; (2) Feelings; (3) Place and Change of Place; (4) Measure and Shape; (5) Living Things; (6) Natural Phenomena; (7) Behavior and the Will; (8) Language; (9) Human Society and Institutions; (10) Values and Ideas; (11) Arts; (12) Occupations and Crafts; (13) Sports and Amusements; (14) The Mind and Ideas; (15) Science and Technology. These 15 superordinate categories, being more in keeping with the Western world view of the 20th and 21st centuries, are assuredly a great improvement over the old-fashioned 1852 version of 6 superordinate categories.

One characteristic of *Roget's Thesaurus* (1982) is that it has a four-level hierarchy: the 6 superordinate categories (classes) at the top, below them are the subdivisions, which in turn contain heads, which further include their respective basic-level categories as hyponyms in the initial positions of each group of words, followed by clusters of synonymous words set off with semicolons. For example, the superordinate category "Space" has "Motion" as one of its subdivision, which contains among other heads "274 Vehicle," which includes the basic level category "bus" in italics in the initial position of a group of words, and it is followed by such synonymous words as "horsebus" and "motorbus" in one cluster, another set of synonymous words "omnibus," "doubledecker," "single-d." in another cluster, and "autobus," "trolleybus," "motor coach," "coach," "postbus," "minibus" in still another cluster. But many of these synonymous words are subordinate categories. The heads in this version amount to 990, and the number of words and phrases adds up to more than 200,000. However, in *Roget's International Thesaurus* (Kipfer 2001), the hierarchy is reduced to three levels. For example, under the superordinate category

"Place and Change of Place," there is the head "179 Vehicle" among other heads, and under this head there is the basic level category "train" in boldface type in the initial position of a group of words, and it is followed by a synonym "railroad train," which is set off with a semicolon from another synonym cluster "passenger train" and "Amtrak". In addition, there are many synonymous clusters under this head: "shuttle train," "shuttle"; "express train," "express"; "subway," "metro" <Fr>, "tube," "underground" <Brit>, etc. The heads in this version amount to 1,075, and the number of words and phrases totals 330,000.

Another feature of *Roget's Thesaurus* is its tradition of including under a head nouns, adjectives, verbs, sometimes adverbs, and occasionally prepositions, conjunctions and interjections. Grouping these different parts of speech together under one single head certainly facilitates a writer's search for the most precise words and the most appropriate syntactic structures.

2.2 WordNet

WordNet, constructed by Princeton University, is an electronic lexical database, which contained 91,000 synsets (synonym sets) in 1999. These synset are all content words arranged according to their parts of speech: nouns, verbs, adjectives and adverbs. In having synsets as   building blocks, WordNet is similar to a thesaurus. Each synset contains all the words expressing a certain concept. In addition, the synsets in WordNet also provide such semantic relations as hyponomy, meronymy, antonymy and entailment. It distinguishes between the conceptual and lexical levels, because certain antonyms involve only the lexical level instead of the conceptual level, such as "big/little" and "large/small": "big brother" is an antonym to "little brother," but not "small brother." However, it does not contain information of cognitive domains, which are essential for establishing an ontological model.

2.3 SUMO (Suggested Upper Merged Ontology)

According to Sevcenko (2003), while WordNet mainly maps conceptualizations of our world at the lexical level, SUMO, created at Teknowledge Corporation, aims to organize them into a logical structure. In 2003, SUMO already contained 1,000

concepts, forming a semantic network and supplemented by some axioms[3]. These concepts may be organized into a many-leveled hierarchy from the topmost concept "Entity" to the second level "Abstract" or "Physical." The concept "Physical" may also have the lower level concepts "Object" or "Process." And so on and so forth. An example of the ontology of the word "blues" may have as many as 12 levels: from "Entity" to "Physical" to "Object" to "Artifact" to "Creation" to "Art" to "Music" to "Music Genre" to "Popular music" to "Folk Music" to "Folk song" and finally to "blues." All the subclasses under "Entity" are supposed to be mutually exclusive so that they do not overlap. Since SUMO is a purely logical structure of ontology, it has to be linked to WordNet to have its substantial representations.

2.4 BOW (The Academia Sinica Bilingual Ontological Wordnet)

BOW, constructed by Academia Sinica, Taiwan, is a bilingual (Chinese and English) ontological network merging the different electronic databases of WordNet, SUMO, Sinica Corpus and others. It is intended to provide lexical information about ontological domains based on SUMO, semantic relations offered by WordNet, and English translations of Chinese expressions. Its ontological structure is essentially the same as that of SUMO.

2.5 *Thesaurus of Chinese Synonyms* (《同義詞詞林》Mei et al. 1996)

This is the first thesaurus of Chinese synonyms containing 70,000 words and phrases. There are 12 superordinate categories, which in turn consist of 94 subdivisions, which are further divided into 1,428 categories. The 12 superordinate categories are: (1) People (人); (2) Things, Creatures, Plants, Architecture, Materials, Instruments, Clothing, Food, Drugs and Poison (物); (3) Time and Space (時間與空間); (4) Abstract entities (抽象事物); (5) Attributes (特徵); (6) Actions (動作); (7) Feelings (心理活動); (8) Activities (活動); (9) Natural, physiological, physical and personal phenomena (現象與狀態); (10) Relations (關聯); (11) Function words (助語); (12) Greetings (敬語). These 12 superordinate categories present a Chinese

---

[3] An axiom may be stated as "If *c* is an instance of combustion, then there exist heating *h* and radiating light *l* so that both h and l are subprocesses of c." Sevcenko (2003: 2)

ontology very different from those of the above Western ontological models. For example, the second superordinate category "物" (things) in Chinese embraces such heterogeneous entities as animate beings, inanimate beings, architecture, materials, instruments, clothing, food, drugs and poison because they all share the Chinese morpheme "物". This kind of categorization may appear to Westerners as inscrutable as the Australian Dyirbal category of women, fire and dangerous things (Lakoff 1987: 92-104).

3. Constructing an Ontological Model in Terms of Prototypes

This study attempts to categorize the 6,315 English words in *CEECEWL* and their Chinese counterparts on the basis of the concept of prototype effects and basic-level effects proposed by Rosch (1973) and Rosch and Mervis (1975) and expounded by Lakoff (1987:58): "Linguistic categories should be of the same type as other categories in our conceptual system. In particular, they should show prototype and basic-level effects." The prototype effects are characterized by the asymmetrical phenomenon of having some members in a category as more ideal or central members than others. (Lakoff 1987:40) That is why as far as the category of BIRD is concerned, robins are considered more prototypical than chickens, penguins and ostriches, and in the category of CHAIR, desk chairs are more prototypical than rocking chairs, barber chairs, and electric chairs. As for he basic-level effects, they equally demonstrate the asymmetrical phenomenon of having a certain level of words more basic than the words of other levels. For example, a superordinate category word such as "furniture" and subordinate category word "rocker" are psychologically less basic than the middle level word "chair." Therefore, the basic-level words primarily constitute our knowledge representation of the world.

This research project will try to find out whether prototype effects coupled with the basic-level effects can satisfactorily classifying the 6,315 English content words of *CEECEWL* and their Chinese counterparts so as to construct an ontological model for these words. As the number of words in *CEECEWL* is not very large, and this ontological model to be constructed is mainly pedagogy-oriented, the levels of categories will be kept to the minimum. This ontological model is more like that of *Roget's International Thesaurus* (Kipfer 2001) than those of SUMO and BOW, whose many-leveled logical structure is far too complex for pedagogical purposes. It is not like that of *Thesaurus of Chinese Synonyms* because it is too incompatible with

Western ontologies. The superordinate level of this ontology consists of 50 semantic categories and each superordinate category will have different levels of categories depending on its ontological structure. These semantic categories of different levels are supposed to be mutually exclusive, but there is no denying that some of them may overlap. Furthermore, if necessary, the semantic relations such as hyperonymy, hyponymy, synonymy, contrast, antonymy, meronymy, polysemy, metaphor, and metonymy will be used to verify the categorization.

The 50 superordinate categories for CEECEWL are postulated as follows:

(1) Body parts (人身部位); (2) Home life (家居生活與相關詞彙); (3) Food from plants (蔬菜、五穀、豆類、硬果等食物); (4) Food from animals (魚鮮與肉類); (5) Meals and snacks (餐點與相關詞彙); (6) Verbs about metabolism (與新陳代謝有關的動詞); (7) Senses and actions of body parts (感官與身體動作詞彙); (8) Beverages (飲料); (9) Cooking and ingredients (烹飪及調味料); (10) Fruits (水果); (11) Kinship terms (親屬名稱); (12) Clothing, accessories and cosmetics (衣物、修飾、飾物、化妝與化妝品); (13) Mathematical concepts and assessment (數量與評量); (14) The natural environment (自然與環境); (15) Architecture and public places (建築、公共場所與相關詞彙); (16) Non-food animals (非食用之動物與相關詞彙); (17) Non-food plants (非食用之植物與相關詞彙); (18) Temporal concepts (時間觀念); (19) Colors (顏色); (20) Educational terms (教育與相關詞彙); (21) Transportation and communication (交通與溝通之方式); (22) The humanities (人文與歷史); (23) The arts (藝術與手工藝); (24) Science and technology (科技); (25) The law and public affairs (公務與法律); (26) Economy and business (經濟與商務); (27) Instruments not used at home (家居生活之外的工具、方法與方式); (28) Recreation (休閒); (29) Health and medicine (健康與醫療); (30) Conflicts and weapons (抗爭與武器); (31) Materials and energy (物料與能源); (32) Emotions and personality traits (情緒、個性與身體狀況); (33) Agriculture and fishing (農牧業與漁業); (34) People and occupations (各種人物與職業); (35) Attributes of people, animals and things (adjectives, adverbs and verbs) (有關人物、動物或事物特性及狀態之形容詞、副詞及名詞);(36) Actions of people, animals and things (verbs) (人物、動物或事物的動作(動詞)); (37) Light and fire (有關

光、火與爆炸的詞彙); (38) Address forms (稱呼); (39) Spatial concepts (空間觀念); (40) Measurements (計量標準與單位); (41) Possession, management and planning (擁有、管理與計畫); (42) Different stages of life (人生不同階段); (43) Weather and climate (天氣與氣候); (44) Fortune and risks (機運與考驗); (45) Social relations and services (社會關係與服務); (46) Language structure (語言結構與相關詞彙); (47) Cultural values (禮儀與德行); (48) Weaknesses and evils (缺失與罪惡); (49) Observation and reasoning (觀察與推理); (50) Urban and suburban environments (城市與鄉村).

To illustrate the semantic hierarchies of these 50 superordinate categories, 3 of them will be analyzed in the following sections.

3.1 Body parts (人身部位)

There are altogether 80 body parts in *CEECEWL*. They can be analyzed into a meronymic hierarchy of six levels as follows:

A. The superordinate level: Body parts (人身部位)

B. The 2nd-level categories: head (n.)頭; neck (n.)頸; trunk (n.)軀幹; limbs (n.)四肢; skin (n.)皮膚; artery (n.) 動脈; vein (n.)靜脈; blood (n.)血; skeleton (n.)骨架; flesh (n.)肉; muscle (n.)肌肉; hair (n.)毛髮; nerve (n.)神經;

C. The 3rd-level categories:

1. head (n.)頭> skull (n.)頭顱骨; hair (n.)頭髮; brain (n.)腦; face (n.)臉/面; ear (n.)耳朵

2. neck (n.)脖子> throat (n.)喉部

3. trunk (n.)軀幹> shoulders (n.)肩膀; bosom/breast (n.)胸部/女人乳房; chest (n.)胸部; organs (n.)內臟; abdomen/stomach/belly/tummy (n.)肚子; navel (n.)肚臍; rib (n.)肋骨; backbone/spine (n.)脊椎; waist (n.)腰; hip (n.)臀部

4. limbs (n.)四肢> arms (n.)手臂; legs (n.)腿

5. skeleton (n.)骨架> bones (n.)骨頭;

D. The 4th-level categories:

    1. face (n.)臉/面> forehead/brows(n.)前額; eye (n.)眼睛; eyebrows/brows (n.)
眉毛; eyelash/lash (n.)睫毛; eyelid (n.)眼皮; cheek (n.)臉頰;
nose (n.)鼻子; mouth (n.)嘴 (puff (n.)呼/吹); mustache (n.)鬍
子; chin (n.)下巴; jaw (n.)頜; lips (n.)嘴唇(kiss (v.)/(n.)親吻);
beard (n.)鬍鬚; wrinkle (n.)皺紋

    2. arm (n.)手臂> elbow (n.)肘; joint (n.)關節; hand (n.)手; fist (n.)拳頭;

    3. leg (n.)腿> thigh (n.)大腿 knee (n.)膝蓋; joint (n.)關節; calf (n.)腿肚;
foot (n.)腳; lap (n.)大腿上

    4. organs (n.)內臟> lungs (n.)肺; heart (n.)心臟; liver (n.)肝; kidney (n.)腎臟;
bowels (n.)大腸; intestines (小腸)

E. The 5th-level categories:

    1. nose (n.)鼻子> nostrils (n.)鼻孔.

    2. mouth (n.)嘴> tongue (n.)舌頭; tooth (n.)牙齒; throat (n.)咽喉

    3. hand (n.)手> wrist (n.)手腕; palm (n.)掌; finger (n.)手指; thumb (n.)拇指;
nail (n.)指甲

    4. foot (n.)腳> ankle (n.)足踝; heel (n.)足跟; toe (n.)腳趾; nail (n.)指甲

    5. heart (n.)心臟>pulse (n.)脈搏

F. The 6th-level categories:

    1. finger (n.)手指> knuckles (n.)指節

Six levels of categories have been postulated for the above meronymic hierarchy of the human body parts. Even though the meronymic hierarchy is generally quite straightforward, that is, it is constructed according to the principle of "A is a meronym/part of B," overlap still can be found at some levels. For example, "hair" in English is polysemous, because it refers to both "body hair" at the second level and "hair on the head" at the third level. But in Chinese, the two senses of "hair" are lexicalized into two words, "毛髮" and "頭髮" respectively, so there can be no ambiguity in Chinese. The same with "throat" in English: it means either "the front part of a person' neck" (喉部) or "the passage leading from the back of the mouth of a person" (咽喉), but in Chinese they are again two different words. English "bosom"

(胸部/女人乳房) and "breast" (胸部/女人乳房) are also polysemous, referring to either the chest of both sexes and also "either of the two soft, protruding organs on the upper front of a woman's body that secrete milk after pregnancy" (*The New Oxford American Dictionary* 2001), but in Chinese these two senses are lexicalized differently. The cumbersome definition given by *The New Oxford American Dictionary* shows that English does not have a word for it. But Chinese has exactly the right word for it. The English word "chest" is more neutral and is equivalent to the Chinese word "胸部".

Some body part words such as "flesh" and "muscle" may seem to be synonymous, but in fact they are not. Observe the following two sentences with adjectives derived from these two nouns respectively "fleshy" (肥胖) and "muscular" (肌肉發達).

(1) *He is a fleshy man.* (他有點胖。)

(2) *He is a muscular man.* (他肌肉發達。)

The first sentence is apparently not complimentary, meaning that this person is "a bit too fat." But the second sentence is certainly approbatory, meaning that this person is physically fit and strong. The same for the Chinese translations of these two sentences.

For the human body parts, there are such synonyms as "abdomen," "stomach," "belly" and "tummy," but with quite different registers. "Abdomen" is by far the most formal, because it is derived from Latin, and "tummy" is the least formal, because it is an abbreviation of "stomach" used by children. In between are "stomach" and "belly": the former, being derived from Latin, is more formal than the latter, which is derived from Old English.

Many of these body parts are often used in idioms to achieve metaphorical or metonymical senses. In this regard, English embodied idioms do not always correspond to their Chinese counterparts in meaning. For example, the idiom "go in one ear and out the other" and its Chinese counterpart "從一隻耳朵進去另一隻耳朵出來" have exactly the same meaning of being not attentive. But the other idiom "wash one's hands of someone or something" (refuse to take responsibility for someone or something) is not the same as "洗手不幹", which means to free oneself from evil doing. And frequently the body parts used metonymically are quite different

in English and Chinese. For example, in English, if you "count noses," you count the number of people, but in Chinese, you "count heads" instead.

According to Rosch (1973/1975) and Lakoff (1987), in any hierarchy of categories there should be the basic-level categories which are presumably more salient and constitute the major parts of our knowledge representation. These basic-level categories should also be acquired earlier by children. But the above meronymic hierarchy of body parts has 6 levels and it is not clear which level of categories should be the basic-level categories. The word "knuckles/指節" at level 6 is certainly more specific than "finger" at level 5, and hence it is more difficult to learn. And the level-5words "nostrils/鼻孔," "wrist/手腕," "ankle /足踝" and "heel/足跟" are more specific than the level-4 words "nose/鼻子," "hand/手" and "foot/腳". But the same level-5 words such as "tongue/舌頭," "tooth/牙齒" and "finger/手指" may not be more specific and difficult than such level-4 words as "mustache/鬍子," "wrinkle/皺紋" and "calf/腿肚". Moreover, there are some level-2 words such as "trunk/軀幹," "limbs/四肢," "artery/動脈" and "vein/靜脈," which appear to be more difficult than some lower level words such as "face/臉" and "leg/腿" (both level 3), and also "eye/眼睛" "hand/手" (both level 4). Therefore it is quite difficult to determine which level is the basic level. Cruse (2004:180-81) observes that "The major formal difference between a taxonomy and a meronymy is the lack of clear generalized levels in the latter…For this reason, there seems to be no equivalent to the basic level of a taxonomy, no unmarked levels of specificity independent of context."

Another problem with the prototype model is that for the human body parts, prototype effects do not seem appropriate, because we cannot identify which body part is more prototypical. Is "nose" more prototypical than "mouth"? This question is definitely irrelevant.

These problems about the prototype effects and basic-level effects seem to suggest that the semantic categories of the human body parts do not conform to the concept of prototype. And perhaps it can be extrapolated that the prototype and basic-level effects do not apply to all other meronymic hierarchies.

3.2 Food from plants (蔬菜、五穀、豆類、硬果等食物)

There are altogether 33 English words about food from plants in *CEECEWL*. They can be categorized as follows:

  A. The superordinate level: Food from plants

  B. The 2nd-level categories: cereals (n.)穀類; vegetables (n.)蔬菜;. nuts (n.)堅果; legumes (n.)豆類; roots (n.)根莖類; fungi (n.)菌類

  C. The 3rd-level categories:

    1. cereals (n.)穀類> corn (n.)玉米; grain (n.)穀物; oatmeal (n.)燕麥片; rice (n.)米/飯; wheat (n.)小麥

    2. vegetables (n.)蔬菜> cabbage (n.)包心菜; celery (n.)芹菜; cucumber (n.)黃瓜; lettuce (n.)萵苣; onion (n.)洋蔥; pumpkin (n.)南瓜; salad (n.)沙拉; spinach (n.)菠菜; squash (n.)美國南瓜; tomato (n.)蕃茄

    3. nuts (n.)堅果> chestnut (n.)栗子; kernel (n.)果仁; walnut (n.)胡桃

    4. legumes (n.)豆類> bean (n.)豆子; pea (n.)豌豆; peanut (n.)花生

    5. roots (n.)根莖類> carrot (n.)胡蘿蔔; potato (n.)馬鈴薯; radish (n.)蘿蔔; yam/sweet potato (n.)甘藷

    6. fungi (n.)菌類> mushroom (n.)蘑菇

    7. others> pickle (n.)泡菜

In this category of vegetables, it is quite straightforward that the lower level words are the hyponyms of the upper level words. Because this is a taxonomic hierarchy (Cruse 2004:176-80), there is a distinct basic level, that is, level 3 words: "corn/玉米," "oatmeal/燕麥片," "rice 米/飯," "wheat/小麥," "cabbage/包心菜," "celery/芹菜," "chestnut/栗子," "walnut/胡桃," "pea/豌豆," "peanut/花生," "carrot/胡蘿蔔," "potato/馬鈴薯," and so on. Even though this basic-level is not the middle level between the superordinate and subordinate levels, it does contain most of the common food from plants. This category of English words and their Chinese counterparts does support the basic-level effects put forward by Rosch and Lakoff.

But "tomato" is generally considered fruit in Taiwan instead of a vegetable.

3.3 Verbs about Metabolism (與新陳代謝有關的動詞)

There are 18 verbs about metabolism in *CEECEWL*. These verbs can be categorized into verbs of ingestion and excretion.

A. The superordinate level: Verbs of metabolism

B. The 2nd-level categories: ingest (v.)飲食; excrete (n.)排泄

C. The 3rd-level categories:

    1. ingest (v.)飲食> drink (v.).喝; eat (v.)吃;

    2. excrete (n.)排泄> piss (v.)撒尿; sweat (v.)流汗; urinate (v.)尿

D. The 4th -level categories:

    1. eat (v.)吃> chew (v.)/咀嚼; devour (v.)吞食; dine (v.)進餐; feed(v.)餵養; gnaw (v.)啃咬; gobble (v.)狼吞虎嚥; gorge (v.)貪婪地吃; nibble (v.)輕咬; overeat (v.)吃得過多;

    2. drink (v.)> quench (v.)解渴; sip(v.)小口喝;

    3. eat (v.)吃/drink (v.)> swallow (v.)嚥下; gulp (v.)貪婪地吞嚥;

In this category of verbs about metabolism, the superordinate category is just an abstract term, not a real verb "metabolize," because this verb is not found in *CEECEWL*. The same with the 2nd-level verbs "ingest/飲食" and "excrete/排泄," which also serve as abstract terms in this hierarchy of verbs. The 3rd-level verbs seem to be the basic-level verbs, because they are the default verbs frequently used in daily life. As for the 4th-level verbs, they are more specific and less basic than the 3rd-level verbs.

4. Conclusions

1. The ontological model of 50 superordinate categories constructed in this study is pedagogy-oriented, and it is more like that of *Roget's International Thesaurus* (Kipfer 2001) than the many-leveled logical structure of SUMO and BOW.

2. The hierarchical structure of the lexical items under each superordiante category facilitates the learning of English vocabulary if it is compatible with that of Chinese. The small number of English lexical items which are not compatible with their Chinese counterparts calls for greater attention on the part of the learner of English.

3. It has been found that the lexical items and the conceptual structures underlying them are generally shared by English and Chinese. This may provide evidence for the rather universal cognitive systems of human beings. However, English sometimes does not lexicalize some words which exist in Chinese, such as "乳

房" and "頭髮" vs. "毛髮".

4. The cognitive concepts of prototype and basic-level effects put forward by Rosch and Lakoff have been found to be useful for taxonomic hierarchies of words, such as vegetables and verbs of metabolism, but not appropriate for meronymic hierarchies such as the body parts and other part-whole relationships.

參考書目

Ahrens, K. & C. R. Huang (2001) "A Comparative Study of English and Chinese Synonym Pairs: An Approach based on The Module-Attribute Representation of Verbal Semantics," *Proceedings of the 15th Pacific Asia conference on Language, Information and Computation*. Hong Kong: City University of Hong Kong.

Ahrens, K., L. L. Chang, K. J. Chen, and C. R. Huang (1998). "Meaning Representation and Meaning Instantiation for Chinese Nominals," *Computational Linguistics and Chinese Language Processing*. Vol.3.1.45-60.

Ahrens, K., and C. R. Huang (1996) "Classifiers and Semantic Type Coercion: Motivating a New Classification of Classifiers," *Proceedings of PACLIC11. 1-10. Seoul: Kyung Hee University.*

American Heritage Dictionary(1980) *Roget's II: The New Thesaurus*. Boston: Houghton Mifflin Co.

Biq, Y. O. (2000) (ed.) "Special Issues on Chinese Verbal Semantics," *Computational Linguistics and Chinese Language Processing*. Vol. 5 No. 1.

Brady, J. (1991) "Toward Automatic Categorization of Concordances Using Roget's International Thesaurus," *Proceedings of the Third Annual Midwest Artificial Intelligence and Cognitive Science Society Conference*. Southern Illinois University, Carbondale, pp. 93-97.

Cheng, C. C. (1998) "Learning Words with Many Texts", in *Proceedings of the 1st International Conference on Multimedia Language Education* (1-12). Taipei: The Crane Book Store.

Cruse, D. A. (1986) *Lexical Semantics*. Cambridge: Cambridge University Press.

Cruse, D. A. (2004) *Meaning in Language*. Oxford: Oxford University Press.

Chapman, R. L. (1992) *Roget's International Thesaurus* (5th edition). New York: HarperCollins.

Ellman, J. and J. Tait (2000) "On the Generality of Thesuarally Derived Lexical Links," *Proceedings of the 5th International Conference on the Statistical Analysis of Textual data (JADT 2000)*. Lauzanne, Switzerland, May, 147-54.

Fellbaum, C. (1999) *WordNet*. Cambridge: MIT Press.

Glazier, S. (1997) *Random House Word Menu*. New York: Random House.

Goddard, C. (1998). *Semantic analysis*. Oxford: Oxford University Press.

Huang, C. R., K. Ahrens, and K. J. Chen (1998) "A Data-driven Approach to the Mental Lexicon: Two Studies on Chinese Corpus Linguistics," *Bulletin of the Institute of History and Philology*. 69.1.151-179.

Hudson, R. (1995). *Word Meaning*. London: Routledge.

Ide, N. and J. Veronis (1998) "Introduction to the Special Issue on Word Sense Disambiguation: The State of the Art," *Computational Linguistics*: Special Issue on Word Sense Disambiguation, 24 (1), 1-40.

Jarmasz, M. and S. Szpakowicz (2001) "*Roget's Thesaurus* as an Electronic Lexical Database," In W. Gruszczynski and D. Kopcinska (eds.) to appear. http://www.site.uottawa.ca/~mjarmasz/pubs/TR-2000-02.pdf.

Jarmasz, M. and S. Szpakowicz (2001) "*Roget's Thesaurus*: A Lexical Resource to Treasure," Proceedings of the NAACL Wordnet and Other Resources Workshop, Pittsburgh, June, 186-88.

Kipfer, B. A. (2001) *Roget's International Thesaurus* (6th edition). New York: HarperCollins.

Kirkpatrick, B. (1998) *Roget's Thesaurus of English Words and Phrases*. UK: Penguin.

Lakoff, G. (1987) *Women, Fire and Dangerous Things*. Chicago: University of Chicago Press.

Lehmann, F. (1995) "Combining Ontologies, Thesauri, and Standards," Proceedings of the IJCAI Workshop on Basic Ontological Issues in Knowledge Sharing. Montreal, Canada, August.

Levin, B. & S. Pinker (eds.) (1991). *Lexical and conceptual semantics*. Oxford: Blackwell.

Lloyd, S. M. (1982) *Roget's Thesaurus of English Words and Phrases*. UK: Longman Group Ltd.

Lyons, J. (1977) *Semantics*. Cambridge: Cambridge University Press.

Merriam Company (1976) *Webster's Collegiate Thesaurus*. Taipei: Meiya Publications.

Murphy, M. L. (2003) *Semantic Relations and the Lexicon*. Cambridge: Cambridge University Press.

Natase, V. and S. Szpakowicz (2001) "Word Sense Disambiguation in Roget's Thesaurus Using Wordnet," *Proceedings of the NAACL Wordnet and Other Lexical Resources Workshop,"* Pittsburgh, June, 17-22. http://www.seas.smu.edu/~rada/mwnw/papers/WNW-NAACL-220.pdf.

Nirenurg, S. and V. Raskin (2004) *Ontological Semantic*s. Cambridge: MIT Press.

Old, J. (1991) "Analysis of Polysemy and Homography of the Word "Lead" in *Roget's International Thesaurus*," *Proceedings of the Third Annual Midwest Artificial Intelligence and Cognitive Science Society Conference*. Southern Illinois University, Carbondale, pp. 98-102.

Oxford University Press (1998) *Oxford Essential Thesaurus*. New York: Berkley Publishing Group.

Palmer, F. R. (1981) *Semantics*. Cambridge: Cambridge University Press.

Patrick, A. B. (1985) *An Exploration of Abstract Thesaurus Instantiation*. M. Sc. Thesis, University of Kansas, Lawrence, KS.

Pustejovsky, J. (1992) *Lexical Semantics and Knowledge Representation*. Berlin: Springer-Verlag.

Pustejovsky, J. (ed.) (1993). *Semantics and the lexicon*.   Academic Press.

Pustejovsky, J. (ed.) (1995). *Generative Lexicon*. Cambridge: MIT Press.

Pustejovsky, J. (ed.) (1997) *Lexical Semantics: The Problems of Polysemy*. Oxford: Oxford University Press.

Random House (1998) *Random House Webster's College Thesaurus*. New York: Random House.

Rosch, E. H. (1973) "Natural Categories," *Cognitive Psychology*, 4, 328-350.

Rosch, E. H. and C. B. Mervis (1975) "Family Resemblances: Studies in the Internal Structure of Categories," *Cognitive Psychology*, 7, 573-605.

Sedelow S. and W. Sedelow (1986) "Thesaural Knowledge Representation," *Proceedings of the 2nd Annual Conference of the University of Waterloo Centre for the New Oxford English Dictionary: Advances in Lexicology.* University of Waterloo.

Sowa, J. F. (2000) *Knowledge Representation: Logical, Philosophical and Computational Foundations*. CA: Brooks Publishing Co.

Sparck, J. K. (1986) *Synonymy and Semantic Classification*. Edinburgh: Edinburgh University Press.

Talburt, J. R. and D. M. Mooney (1989) The Decomposition of *Roget's International Thesauru* into Type-10 Semantically Strong Components," *Proceedings of the 1989 ACM South Regional Conference*, 78-83. Tulsa, Oklahoma.

Wilks, Y. (1998) "Language processing and the thesaurus," *Proceedings of the National Language Research Institute*. Tokyo, Japan.
http://www.dcs.shef.ac.uk/~yorick/papers/cs-97-13.
梅家駒等(1997)《同義詞詞林》台北:東華書局。

計畫成果自評

一、本研究計畫所研究之《大考中心高中參考詞彙表》之英文詞彙與其相對應中文詞彙之語義範疇，共分五十大類，其下再依每一語義範疇之特性分若干結構層次(hierarchical levels)。此種語義範疇及其結構層次對於英文及中文之詞彙分類極有助益，同時對於了解中英文之知識本體(ontology)頗有貢獻。

二、本研究計畫以上述五十大類之語義範疇及結構層次來檢驗 Rosch 及 Lakoff 所提出之原型效應(prototype effects)及基本詞彙效應(basic-level effects)，發現他們的這兩個觀念對於分類結構層次(taxonomic hierarchy)之語義範疇頗有啟發性，但是對於部分與整體關係之結構層次(meronymic hierarchy)則不能適用。此項發現對於認知語言學頗有貢獻。

三、本研究計畫所建立之五十個中英文語義範疇及其結構層次可以應用於中英文教學，因為中英文之詞彙雖然大致對應，但是有的英文詞彙中文沒有，有的中文詞彙英文沒有。本研究計畫所指出之中英文詞彙之異同是中英文語言教學必須注意之重點，有助於提升中英文教學之效果。

可供推廣之研發成果資料表

附件二

☐ 可申請專利　　☐ 可技術移轉　　　　　　　　　日期：＿＿年 月＿＿日

國科會補助計畫	計畫名稱： 計畫主持人： 計畫編號：　　　　　　　　　　學門領域：
技術/創作名稱	
發明人/創作人	
技術說明	中文： （100~500 字） 英文：
可利用之產業 及 可開發之產品	
技術特點	

推廣及運用的價值	24

※ 1.每項研發成果請填寫一式二份，一份隨成果報告送繳本會，一份送 貴
　　單位研發成果推廣單位（如技術移轉中心）。

※ **2.本項研發成果若尚未申請專利，請勿揭露可申請專利之主要內容。**

※ 3.本表若不敷使用，請自行影印使用。