

# Building a Database of Related Concepts of Mandarin Metaphors based on WordNet and SUMO

Siaw-Fong Chung  
Department of English  
National Chengchi University  
Taipei, Taiwan  
e-mail: sfchung@nccu.edu.tw

Chun-Hung Chen  
Department of English  
National Chengchi University  
Taipei, Taiwan  
e-mail: 97551020@nccu.edu.tw

F.Y. August Chao  
Department of Management  
Information Systems  
National Chengchi University  
Taipei, Taiwan  
e-mail: fychao.tw@gmail.com

**Abstract**—The ontological concepts for 1,256 Mandarin metaphor types (such as *cheng2zhang3* ‘grow/growth’ and *jian4she4* ‘construct/construction’) were stored in a database in the form of metaphor-concept pairings. In each pair, a Mandarin metaphor type is paired with ontological concepts in English. Based on the relations given in these pairings, a database of relational concept was created. Calculation of co-appearance of concepts can be undertaken using the database. The work will contribute to analysis of conceptual relatedness, semantic networking and relatedness of Mandarin metaphors.

**Keywords**- *metaphor; concept; English; Chinese; SUMO; WordNet; database*

## I. INTRODUCTION

Since the proposal of the Conceptual Metaphor Theory [1] [2], the theory has been applied and investigated in various fields such as cognitive linguistics, corpus linguistics and computational linguistics. The strength of the theory lies in its ability to explain the relationship between two related domains – the target and source domains. The following example (1) illustrates a linguistic expression for LOVE IS WAR in which LOVE is the target domain and WAR is the source domain. The words in italics fall under the domain of WAR but are used to describe LOVE.

(1) He is known for his many rapid *conquests*. She *fought* for him, but his mistress *won out*. (Lakoff and Johnson [1], p. 49)

The theory emphasizes a source-domain mapping in which the source domain (WAR) is usually more concrete than the target domain (LOVE). Linguistic examples such as (1) above have been collected under the Conceptual Metaphor Homepage (<http://cogsci.berkeley.edu/lakoff/>).

In addition to the Conceptual Metaphor Homepage, other work on metaphor database includes Metalude at the Lingnan University (<http://www.ln.edu.hk/lle/cwd/project01/web/home.html>) and Fillmore’s [3] Framenet at University of California, Berkeley ([\[berkeley.edu/\]\(http://framenet.icsi.berkeley.edu/\)\). The Metalude database has its own search engine for the target and source domains. The system provides lexical items within a target domain and the examples of the sentences come from dictionaries and the Conceptual Metaphor Homepage mentioned earlier. Framenet, on the other hand, provides the cores and relations within each lexical unit. For instance, the lexical unit ‘birth’ has the cores of Child \[Child\], Egg \[Egg\], Father \[Pop\], Mother \[Mom\] and Parents \[Par\]. These cores, however, are not designed with the purpose of defining the metaphorical meanings of ‘birth,’ which is one of the concerns in the Conceptual Metaphor Theory.](http://framenet.icsi.</a></p></div><div data-bbox=)

In addition, other systems such as Martin’s MIDAS (Metaphor Interpretation, Denotation, Acquisition System) [4], Wang’s Chinese Noun-Noun metaphor database [5], Fass and Wilks’ Preference Semantics [6] and Mason’s CorMet [7] are among the databases which aim to extract or interpret metaphors automatically. This work, however, is different in that it aims to provide conceptual links between lexical items (which are metaphors) through using WordNet [8] and Suggested Upper Merged Ontology (SUMO) [9]. SUMO is used because it has been widely tested by many scholars and it is also an upper ontology that can complement the more specific concepts of WordNet.

The metaphor extraction process (i.e., to extract metaphorical use such that in (1) previously) follows those used in Chung [10] in which metaphor types (such as *conquests*, *fought for* and *won out* in (1)) in Mandarin metaphors were first manually extracted. Metaphor types are defined as metaphorical expressions which denote a mapping between domains such that shown in (1) previously in which expressions from WAR (in italics) are used to describe LOVE. (They are called metaphor “types” because their repeated occurrences or frequencies in corpora – the “tokens” – are not considered in this current work.) After the extraction process, these metaphor types will serve as keywords to extract their conceptual information through WordNet and SUMO with the assistance of a bilingual database. In this work, we report the building of MetCon (abbreviated from Metaphor Concept), a database created based on the mapping results between Mandarin metaphor types and their ontological concepts.

## II. EXTRACTION OF MANDARIN METAPHOR TYPES

As reported in Chung [10], at the very first step, Mandarin metaphor types were manually identified from the Chinese Gigaword Corpus 1.0 [11] through the Chinese Sketch Engine [12], a grammatical relation platform which allows the display of collocations according to ‘subject,’ ‘object,’ ‘modifiers,’ etc. The Chinese Gigaword corpus produces a huge database encompassing approximately one billion (1,000,000,000) characters. It comprises two sub-corpora – newspapers from Central News Agency of Taiwan (CNA) and Xinhua News Agency of Beijing (XIN). For Chung [10], the data from CNA was collected between January, 1991 and December, 2002 while the data for XIN was collected between December, 1990 and September, 2002.

Five (target domain) keywords – ECONOMY, EDUCATION, POLITICS, FINANCE and FOREIGN AFFAIRS – were searched in the corpus in order to locate metaphor types used to describe these five terms. A total of 106,465 instances was selected from these five target domains with Taiwan data constituting 54% and China data constituting 46% of the total instances.

Sample sentences can be found in (2) below in which potential metaphor types (underlined) identified for these five keywords (boxed) were manually extracted.

- (2) (a) *de2guo2*                      *wai4jiao1bu4*  
 Germany              department.of.foreign.affairs
- tong2shi2*                      *you4 fa1dong4*  
 at.the.same.time              again              start
- wai4jiao1*                      *gong1shi4* (XIN)  
 foreign.affairs                      attack
- “At the same time, the Department of Foreign Affairs in Germany has again started their attack”
- (b) *jiu4*                      *guo2jia1* *yu4suan4* *cai2zheng4*  
 with.regard.to              country              budget              finance
- yu3* *jing1ji4* *cheng2zhang3* *biao3xian4*  
 and economy              growth              perform
- jin4xing2* *bao4gao4* (CNA)  
 proceed              report
- “To make report with regard to the performance of national budget, finance and economic growth”

Based on the selected hits for these five target domains, 1,256 metaphor types (underlined) were identified from both datasets. (Some metaphor types were not distinguishable in terms of their use in Taiwan-only or China-only because the same terms were found. The total 1,256 shows the combination of the metaphor types from both datasets.)

## III. MAPPINGS TO ONTOLOGICAL CONCEPTS

The manually extracted metaphor types were later mapped to their respective ontological concepts. Since WordNet and SUMO are both presented in English, we first mapped all our Mandarin expressions to English. We used a bilingual dictionary database called the Academia Sinica’s in-house database of the Chinese-English Merged Word List which contains translated Chinese-English words from a collection of bilingual dictionaries. This database consists of 527,132 Chinese-English translations (151,305 English words and 283,572 Chinese words). Since one Mandarin word may be translated in different ways, a multiple-mapping of Chinese-English translation was produced based on this step.

Next, each corresponding English translation was mapped to their different senses through WordNet 1.6. As a result, a huge network-like database was created based on the mapping to different English translations with their respective WordNet senses (i.e., one English translation may be mapped to one or more WordNet sense). Then, a WordNet sense will be mapped automatically to a SUMO concept (cf. [9]). Based on the mappings to SUMO, every metaphor type has one or more corresponding SUMO concept. One example of metaphor type-SUMO pairings is *cheng2zhang3* ‘growth’ -‘Growth.’

However, the mapping at this stage (metaphor type-SUMO concept) was found not able to state the relationships between the metaphor types in a specific way (because if the process stopped here, the final conceptual structure would follow the SUMO hierarchy without being different from it). In order to overcome this problem, Chung [10] utilized keywords in the SUMO definitions. One example is given in (3) below for ‘Growth’ (bold added). (Words in capitals are ontological concepts themselves. Other documentation for SUMO can be found here: [http://sigmakee.cvs.sourceforge.net/sigmakee/KBs/.](http://sigmakee.cvs.sourceforge.net/sigmakee/KBs/))

- (3) The **Process** of **biological development** in which an **Organism** or **part** of an **Organism** **changes** its **form** or its **size**.

Since also a definition like (3) appears in sentence form, it is hard to relate a Mandarin type to a whole string of words. A wordlist was then created based on the keywords found in all SUMO definitions. For example, in the definition for the node of ‘Growth’ above, the bolded words are lexical words which were kept for concept analyses. All non-content words (prepositions and function words) were removed automatically.

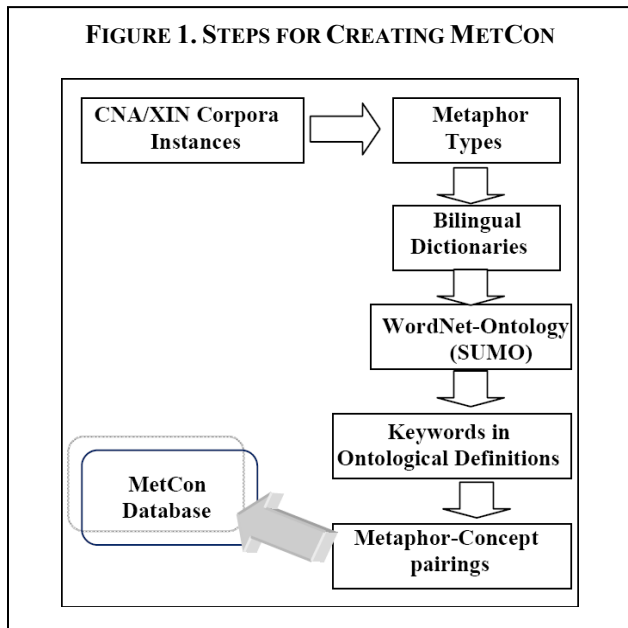
This step by Chung [10] has provided us a chance to create a huge database which contains the link between a metaphor type and all the content words in its SUMO definitions. The assumption behind is that all SUMO definitions contain key concepts used to defined a particular node. If some concepts are repeatedly found in several different definitions, these concepts may be important

concepts and their repeated use can show relatedness between nodes. Sharing this assumption with Chung [10], we created a database which contains a total of 30,683 metaphor type-concept keyword pairings (hereafter ‘metaphor-concept’ pairings). Examples of pairings are given in (4) below (‘Organism’ appears twice in (3) but it will only be recorded once in our mapping).

- (4)
- cheng2zhang3* ‘grow/growth’-Process
  - cheng2zhang3* ‘grow/growth’-biological development
  - cheng2zhang3* ‘grow/growth’-Organism
  - cheng2zhang3* ‘grow/growth’-part
  - cheng2zhang3* ‘grow/growth’-changes
  - cheng2zhang3* ‘grow/growth’-forms
  - cheng2zhang3* ‘grow/growth’-size

From (4), one can see one metaphor type can be mapped to several concept keywords. In MetCon, repeated concept keywords will be used to link related metaphor types. For example, for another Mandarin metaphor type also contains ‘Process,’ it will be linked to *cheng2zhang3* ‘grow/growth’ in (4) above by our system. Similarly, if two ontological concepts share similar metaphor type, these two concepts will also be related in our system.

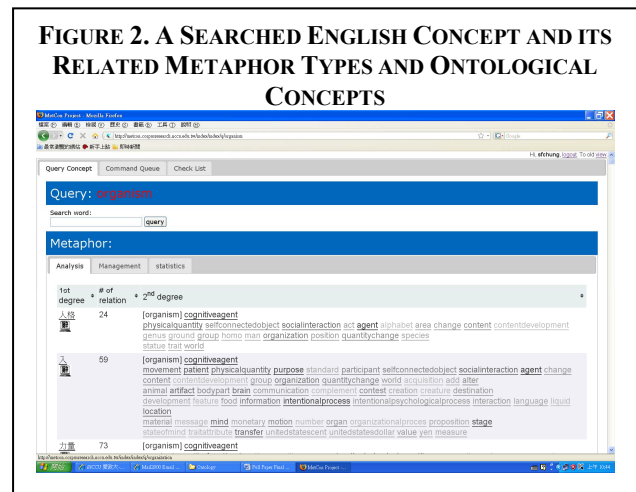
The steps taken to create the metaphor-concept pairings can be summarized in Figure 1 below.



Based on the metaphor-concept pairings collected, MetCon was created. In the next section, some features of MetCon will be discussed.

#### IV. METCON DATABASE

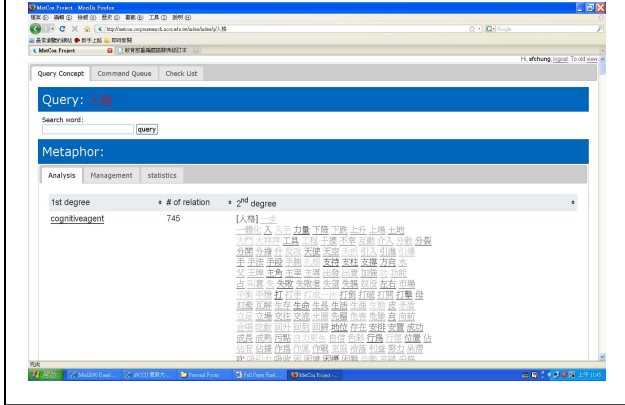
The MetCon database (<http://metcon.corpusresearch.nccu.edu.tw>; registration required) was formed based on the pairings of metaphor-concept described above. Based on these pairings, a network indicating conceptual relatedness between Mandarin metaphors was created, as shown in Figure 2 below. When an English word is searched, all its Chinese corresponding metaphor types (i.e., all metaphor types that also possess this concept) will be listed in the first column. These Chinese terms are the first degree concepts that are related to the searched English word. The first one listed on Figure 2 is *ren2ge2* ‘personality.’ In the second column, the number of English concepts (24) related to this metaphor type (thus, a second degree relationship) is displayed. In the third column, all these 24 English concepts are listed. Each of the words has hyperlink to open a new search result based on the selected word. All the words in the third column are displayed using different degrees of grayness. The darker the color, the more prominent a concept is. MetCon also provides a calculation of conceptual relatedness based on the prominence of a concept to a searched word (to be explained in the next section).



In Figure 2, one can see that all concepts are linked to one another. For example, if we click on the first Chinese word *ren2ge2* ‘personality,’ it will become a new searched term (Figure 3) and its related English concepts (first column) and the number (second column) and listing of related metaphors types (third column) can be seen.

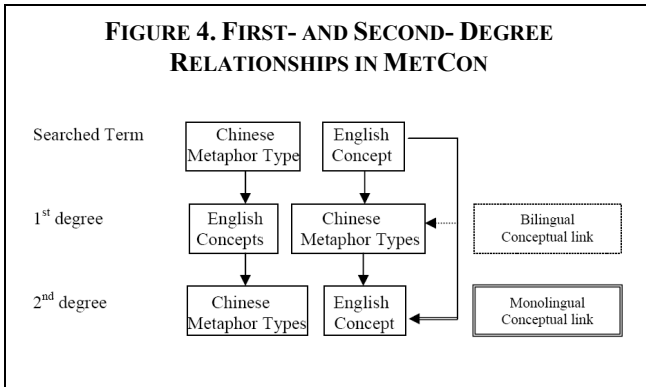
Therefore, based on MetCon, one can figure out the first- and second- degree relationships of a searched English or Chinese word. A first degree relationship allows the comparisons of related bilingual concepts (Chinese-English or English-Chinese). A second degree relationship, on the other hand, enables the comparisons of Chinese-Chinese and English-English relations. In other words, the Chinese-Chinese and English-English relationships can be established through using information from the sharing of certain concepts in another language.

**FIGURE 3. A SEARCHED CHINESE CONCEPT AND ITS RELATED METAPHOR TYPES AND ONTOLOGICAL CONCEPTS**



The relationship between first- and second-degree relationships can be seen in Figure 4 below.

**FIGURE 4. FIRST- AND SECOND- DEGREE RELATIONSHIPS IN METCON**

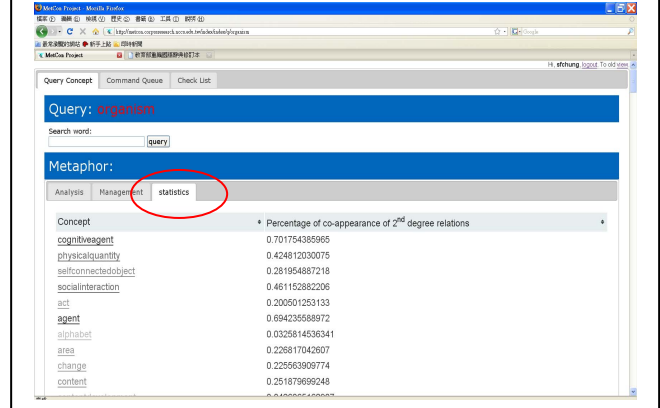


From Figure 4, one can see that a searched term can be examined in terms of its bilingual conceptual links from metaphor-to-concept and concept-to-metaphor links (moving down from the searched boxes) and its monolingual conceptual links from metaphor-to-concept-to-metaphor and concept-to-metaphor-to-concept links. Through inter-relatedness as such, MetCon is able to display conceptual information based on monolingual or bilingual related concepts. In the section below, we provide a calculation of related second-degree concepts in MetCon based on these inter-relationships.

### V. METCON AND CALCULATION OF RELATED CONCEPTS

In MetCon, some statistics about the relationships between words can be found. This can be found under one of the tabs provided (circled below) on the interface.

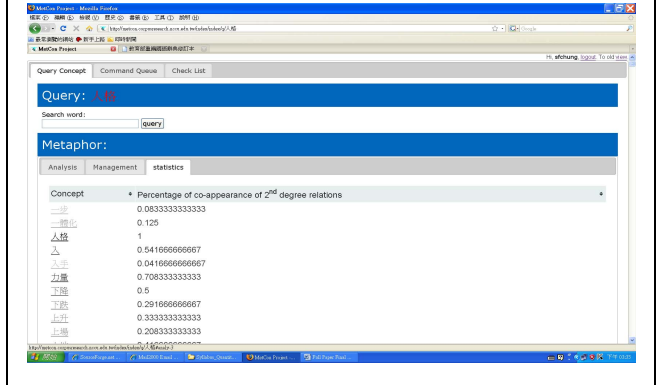
**FIGURE 5. RELATED ENGLISH-ENGLISH CONCEPTS FOR A SEARCHED ENGLISH TERM**



In order to analyze the relationship within the same language, percentages of co-appearing second-degree relations between two concepts in the similar language are presented. Under the title “Percentage of co-appearance of 2<sup>nd</sup> degree relations” in Figure 5, statistics about the co-appearances of two concepts are reported. For ‘Organism’ (the searched concept in Figure 5), it co-occurs 70% (or 0.70) with ‘Cognitive Agent,’ followed by ‘Physical Quantity’ (42%) and ‘Self Connected Object’ (28%). (The agglutinated words in the database are due to their original presentation in SUMO definitions. Two or more agglutinated words form a single concept.)

Similarly, when one searches for a Chinese metaphor type such as *ren2ge2* ‘personality,’ its second-degree relationship can also be found, as shown in Figure 6 below.

**FIGURE 6. RELATED CHINESE-CHINESE CONCEPTS FOR A SEARCHED CHINESE TERM**

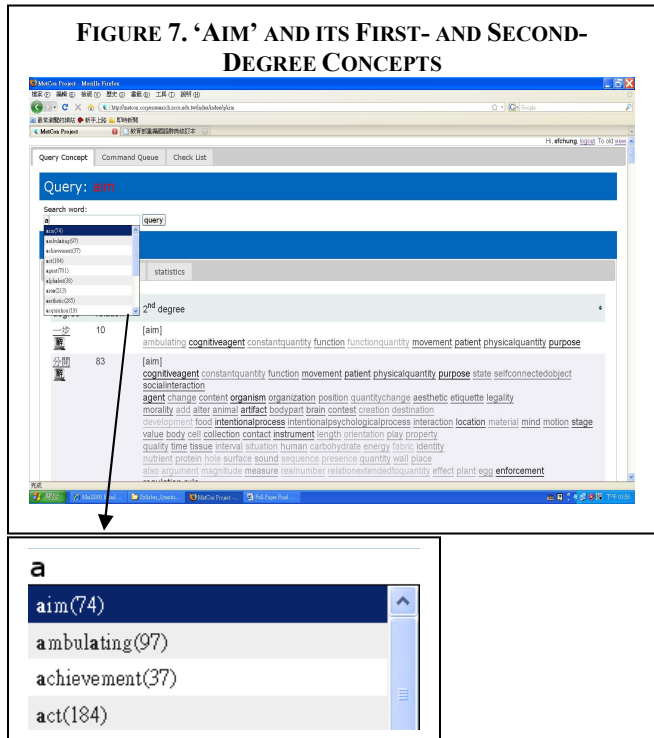


The percentages shown in Figures 5 and 6 above are calculated based on the computation of the shared concepts, shown in the following algorithm.

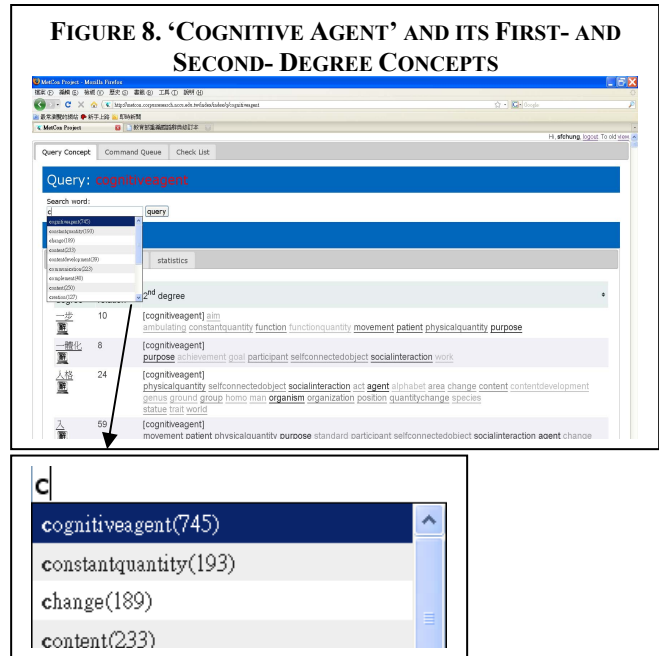
$$CoAP_x^2(y) = \frac{Y_y^1 \cap Y_x^1}{Y_x^1} \quad (5)$$

In the MetCon database, each ontological English concept (e.g., ‘Organism’) maybe linked to several Mandarin metaphor types. Let  $x$  be the queried word, and  $Y_x^n = (y_i | i = 1..k)$  is the word set of  $n$  degree relationship of  $x$ ; therefore,  $Y_x^1$  is the first-degree relationship of  $x$ , and  $Y_x^2$  is the second-degree relationship of  $x$ . In our study, conceptual relationships are built by using Chinese and English words. If the queried word  $x$  is English, then  $Y_x^1$  is Chinese word set and  $Y_x^2$  is English word set and vice versa.

Take an English concept ‘aim’ as an example (Figure 7),  $Y_{aim}^1$  contains Chinese word set with *yilbu4* ‘one step,’ *qinglhai4* ‘to infringe on,’ *ce4lu4* ‘tactics,’ etc. (there are altogether 74 first-degree concepts); and  $Y_{aim}^2$  contains English word set with ‘Cognitive Agent,’ ‘Purpose,’ ‘Organism,’ etc. (with darker degree of grayness). (The total number of the first-degree concepts is given in the drop-down menu in Figure 7.)



For ‘Cognitive Agent’ for example (Figure 8), its  $Y_{cognitiveagent}^1$  contains 745 first-degree Chinese words. Among these 745 words, only 66 words co-exist with the first-degree Chinese word of ‘aim.’ (The number of shared concepts is not displayed in MetCon.) Therefore, the calculation can be seen in (6) to follow.



$$CoAp_{aim}^2(cognitiveagent) = \frac{Y_{cognitiveagent}^1 \cap Y_{aim}^1}{Y_{aim}^1} \quad (6)$$

$$= \frac{66}{74} \approx 89.19\%$$

The result in (6) shows that, among all the 74 concepts of ‘Aim,’ 66 concepts co-occur with ‘Cognitive Agent,’ therefore, the percentage is 89.19%. In Figure 9a below, one can see the percentages for all other second-degree concepts of ‘aim’ presented in the second column. In Figure 9b, when we compare the results of two searched words, ‘Aim’ and ‘Organism,’ we find that ‘Cognitive Agent’ has higher percentage (89.19%) with ‘Aim’ than with ‘Organism’ (70.21%).

**FIGURE 9A. ‘AIM’ AND ITS SECOND- DEGREE CONCEPTS**

Concept	Percentage of co-appearance of 2 <sup>nd</sup> degree relations
<u>aim</u>	1
<u>ambulating</u>	0.216216216216
<u>cognitiveagent</u>	0.891891891892
<u>constantquantity</u>	0.324324324324
<u>function</u>	0.472972972973
<u>functionquantity</u>	0.108108108108
<u>movement</u>	0.635135135135
<u>patient</u>	0.743243243243
<u>physicalquantity</u>	0.594594594595

**FIGURE 9B. COMPARISONS OF TWO ENGLISH CONCEPTS ‘AIM’ AND ‘ORGANISM’ USING  $CoAP_x^2(y)$**

‘Aim’	‘Organism’
$Y_{aim}^1 = 74$	$Y_{organism}^1 = 799$
<ul style="list-style-type: none"> <li>• cognitiveagent 89.19%</li> <li>• purpose 87.84%</li> <li>• organism 85.14%</li> <li>• agent 82.43%</li> <li>• rule 79.73%</li> <li>• regulation 79.73%</li> <li>• enforcement 79.73%</li> <li>• artifact 74.32%</li> <li>• patient 74.32%</li> <li>• intentionalprocess 68.92%</li> <li>• movement 63.51%</li> <li>• instrument 62.16%</li> <li>• device 62.16%</li> <li>• socialinteraction 62.16%</li> <li>• transfer 59.46%</li> <li>• stage 59.46%</li> <li>• location 59.46%</li> <li>• physicalquantity 59.46%</li> </ul>	<ul style="list-style-type: none"> <li>• cognitiveagent 70.21%</li> <li>• agent 69.46%</li> <li>• purpose 65.21%</li> <li>• intentionalprocess 60.58%</li> <li>• patient 57.07%</li> <li>• stage 51.81%</li> <li>• movement 49.44%</li> <li>• location 48.56%</li> <li>• mind 48.06%</li> <li>• brain 48.06%</li> <li>• socialinteraction 46.18%</li> <li>• artifact 46.06%</li> <li>• transfer 45.18%</li> <li>• physicalquantity 42.55%</li> <li>• information 39.30%</li> <li>• state 36.92%</li> <li>• tissue 35.17%</li> </ul>

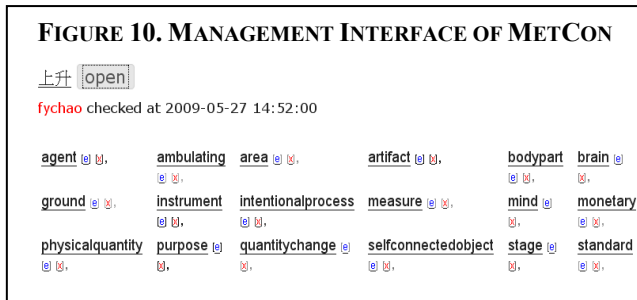
Therefore, based on our methodology, MetCon allows the comparison among all second-degree concepts of a searched word (as in Figure 9a) or the percentages of concepts for two compared searched words (as in Figure 9b). From analyzing the number of related concepts, too, one can examine the generality of a concept. For example, in Figure 5b above, the number of related concepts for ‘aim’ is fewer than those of ‘organism’ ( $Y_{aim}^1 > Y_{organism}^1$ ) and this may indicate that ‘organism’ is more general than ‘aim.’

From the above description, MetCon is found resourceful not only for linguistic research in Mandarin but also for research on conceptual relatedness.

## VI. CONCLUSIONS AND IMPLICATIONS

This paper reports a newly created MetCon database which provides information about conceptual relationships between Chinese and English terms. Through MetCon, conceptual relatedness for monolingual and bilingual data can be displayed. A second stage after the creation of this database is to run an evaluation of the data in the system. For this purpose, we have created a management function so that two or more blind annotators will be employed to check each Chinese-English or Chinese-English relation. In Figure 10, for example, some of the relations between a Chinese metaphor type and its English concepts are displayed. Annotators will be trained to examine each link.

**FIGURE 10. MANAGEMENT INTERFACE OF METCON**



Another way how evaluation can be carried out is through running psycholinguistic experiment. Chung [10], for instance, has asked subjects to rate whether two terms (Metaphor Type and Source Domain) are related with a scale of 1 to 7 with 1 being the least related and 7 being the most relatedness. Similar design could be adopted to examine a portion of the data from MetCon.

This current work will have implication in metaphor research especially in determining which conceptual domain an expression may belong to. Our data has advantage in that they are based on linguistic analyses of metaphors. Therefore, a study of metaphorical conceptual network is also possible in the future.

## ACKNOWLEDGMENTS

Siaw-Fong Chung would like to thank the National Science Council Project (NSC 97-2410-H-004 -001-) for supporting the research herein.

## REFERENCES

- [1] G. Lakoff and M. Johnson, *Metaphors We Live By*. Chicago and London: The University of Chicago Press, 1980.
- [2] G. Lakoff, “The Contemporary Theory of Metaphor,” in *Metaphor and Thought*, 2nd ed., A. Ortony, Ed. Cambridge: Cambridge University Press, 1993, pp. 202-251.
- [3] C. J. Fillmore, “Frame semantics,” in *Linguistics in the Morning Calm*. Seoul: Hanshin Publishing Co., 1990, pp. 111-137.
- [4] J. H. Martin, *A Computational Model of Metaphor Interpretation*. Cambridge, MA: Academic Press, 1990.
- [5] Z. Wang, *Chinese Noun Phrase Metaphor Recognition*, Ph.D. thesis. School of Electronics Engineering and Computer Science, Peking University, China, 2006.
- [6] D. Fass and Y. Wilks, “Preference Semantics, Ill-formedness, and Metaphor,” in *Computational Linguistics*, vol. 9, issue 3-4, 1983, pp. 178-187.
- [7] Z. J. Mason, “CorMet: A Computational, Corpus-Based Conventional Metaphor Extraction System,” in *Computational Linguistics*, vol. 30, 2004, pp. 23-44.
- [8] C. Fellbaum, Ed., *WordNet: An Electronic Lexical Database*. MIT Press, 1998.
- [9] I. Niles and A. Pease, “Towards a Standard Upper Ontology,” *Proc. Formal Ontology in Information Systems. (FOIS-2001)*, 2001, pp. 2-9.
- [10] S. F. Chung, “A Corpus-driven Approach to Source Domain Determination,” in *Language and Linguistics Monograph Series*. Nankang: Academia Sinica, in press.
- [11] D. Graff and C. Ke, *Chinese Gigaword*, Linguistic Data Consortium. Philadelphia, 2003.
- [12] A. Kilgariff, C.R. Huang, S. Smith, and D. Tugwell, “Chinese word sketches,” *Proc. Asialex*, Singapore, 2005.