Using WordNet and SUMO to Determine Source Domains of Conceptual Metaphors

Siaw-Fong Chung
Graduate Institute of Linguistics
National Taiwan University
No. 1, Sec. 4, Roosevelt Road
Taipei, 106 Taiwan
claricefong6376@hotmail.com

Kathleen Ahrens
Graduate Institute of Linguistics
National Taiwan University
No. 1, Sec. 4, Roosevelt Road
Taipei, 106 Taiwan
kathleenahrens@yahoo.com

Chu-Ren Huang
Graduate Institute of Linguistics
Academia Sinica
No.128, Sec. 2, Academia Road, Nankang
Taipei, 115 Taiwan
churen@gate.sinica.edu.tw

Abstract

In previous work by Ahrens et al. (2003) and Chung et al. (2003a, 2003b), the extraction of conceptual metaphors was carried out based on prototypicality. Based on this view, the most prototypical mappings of source to target domains in the corpora help determining the Mapping Principle of the conceptual metaphors (Ahrens 2002). However, this approach is limited in the aspect that the identification of source domains must be manually determined. Therefore, the purpose of this paper is to suggest a way to reduce the manual work required for the determination of the source domain as well as to operationalize the steps to define different source domains. This paper incorporates the WordNet lexical representation and SUMO ontology in the identification of the related nodes to a group of linguistics expressions. The lexicons in the WordNet and SUMO are linked through works such as Niles and Pease (2003).

1 Introduction

The Conceptual Mapping (CM) Model (Ahrens 2002) proposed that a specific Mapping Principle can be generated through observing the source and target domain mappings in conceptual metaphors. For instance, the list of words in Table 1 below were suggested by Ahrens, Chung & Huang (2003) to be related to the source domain of PERSON. They extracted a total of 2000 instances from the Academic Sinica Balanced Corpus of Mandarin Chinese. From these instances, they extracted metaphorical expressions through observing the source-target domain mappings. For instance, *chengzhang* 'growth' below shows a mapping between the concrete source domain of PERSON and the abstract domain of ECONOMY. The analyses were carried out manually.

Table 1: ECONOMY IS A PERSON

成長 (growth)	衰退 (dysfunction)
成長期 (growth period)	病狀 (symptoms)
命脈 (lifeblood)	衰頹(weakness and degeneration)
成長 (grow)	衰退 (to become dysfunctional)
復甦 (regain consciousness)	惡化 (deteriorate)
恢復 (recover)	

Ahrens, Chung and Huang (2003) then counted the number of instances of which these expressions occurred. Based on the most frequent (i.e., the most prototypical) mappings, the Mapping Principles (MP) were suggested. For instance, the MP for ECONOMY IS A PERSON in Table 1 is '*economy is person because people have a life cycle and economy has growth cycle*.'

In this paper, we further extend the work of Ahrens et al. (2003) but minimize the manual work required when determining the source domains. We suggest that the identification of source domains can be solidified through examining a) the WordNet senses and explanations and b) the SUMO nodes

and their definitions. By doing so, we propose a way to operationalize the steps in defining source domains.

This extension of the CM Model is aided by studies linking the lexicon in WordNet and SUMO such as in Niles and Pease (2003). After the links were established, the ontological nodes in SUMO were provided hyperlinks to their semantic definitions by WordNet. Also facilitating this analysis is the work of Huang (2004) in which all the nodes in SUMO are available in both Chinese and English through the Academia Sinica Bilingual Ontological WordNet (BOW; available at http://bow.sinica.edu.tw).

2 Methodology

Our methodology involves two major steps, i.e., extracting metaphorical instances from the corpus and examining the WordNet explanation and SUMO nodes.

For the corpus search, a total of 1062 instances of *shiye*事業 'career' were extracted from the Academia Balance Corpus of Mandarin Chinese. All metaphorical instances were manually marked. A metaphorical mapping occurred when there was a mapping from the source (such as PERSON) to the target domain (CAREER). For instance, in the phrase *shiye youhuan yishi* 事業憂患意識 'the worried consciousness of career,' the abstract target domain CAREER is mapped onto the consciousness of PERSON. What makes this paper different is that we did not categorize all the expressions according to source domains at this stage. We used WordNet and SUMO instead to determine the categorization of the metaphorical expressions found.

3 Results

For the A total of 59 types (84 tokens) of metaphorical expressions were found. These instances are listed in Table 2 on the next page.

Upon keying in the target word such as *yishi* 'consciousness' (highlighted in the Table 2) into Sinica BOW, there were a list of senses suggested by the WordNet lexical knowledge database. The most concrete sense was pulled out from the list to further search for their related SUMO nodes (which were provided after the explanation of each WordNet sense).

Table 2 : Metaphorical expressions related to *shiye* 'career'

Conceptual Metaphors	Tokens	Conceptual Metaphors	Tokens	Conceptual Metaphors	Tokens
新創	1	紮實	1	溶進...之中	1
創造	5	起(步)	1	收起來	1
開創	1	(走)向	2	投身...中	1
共創	1	第一(步)	1	基礎	5
再創	2	闖	2	追求	4
挑戰	1	關	5	風險	1
策略	2	關卡	1	供輸	1
趨勢	3	過程	1	賭	1
(幕後)功臣	1	前途	2	投向	1
異(軍)突起	1	前程	1	投下	1
改革	1	競爭	1	衝刺	1
兵符	1	抗爭	1	角色	1
憂患意識	1	打拚	1	大舞台	2
搖身一變	1	掙	1	階梯	1
創傷	1	拼	1	登上...位子	1
躍進	1	軌道	1	投入	1

放手	1	轟轟烈烈	2	玩掉	1
生命力	1	走(上)	1	包袱	1
壯大	1	火車頭	1	開發	1
成長	1	退出	2		

Table 3 shows the example of *yishi* 'consciousness' as well as other examples taken from Table 2 in which their related WordNet senses and SUMO nodes are listed.

Table 3: Defining Source Domains through WordNet and SUMO

Expressions	**WordNet Senses**	**Explanations**	**SUMO Nodes**
憂患(意識)	1: consciousness	an alert cognitive state in which you are aware of yourself and your situation	Awake (清醒)
挑戰	4: challenge	a call to engage in a contest or fight	Requesting (請求)
策略	6: ambush	the act of concealing yourself and lying in wait to attack by surprise	ViolentContest (暴力性的競爭)
基礎	1: foundation	lowest supporting part of a structure	Region(區域)
軌道	2: track	a pair of parallel rails providing a runway for wheels	TransportationDevice (運輸工具)

For the expressions in Table 3, we first searched for their WordNet senses and examined their WordNet explanations. The shaded words in Table 3 above provided the cue to which source domains these expressions belong. For instance, the concept of *cognitive state* might refer to a PERSON; *contest* and *attack* to WAR whereas *rails* and *wheels* to VEHICLE.

However, the sorting of source domains using the WordNet alone was not enough. For instance, the WordNet explanation regarding *jichu* 'foundation'基礎in Table 2 yielded the idea *structure*. An analysis using the WordNet alone cannot identify whether it is referring to the *structure* of an artifact or any other object. Hence, the second step involved examining the related SUMO nodes. All WordNet senses are linked to the related SUMO nodes, which are listed on the rightmost column of Table 2. Through examining the SUMO nodes, we decide whether a group of expressions belongs to the same concept, i.e, the concept of the source domain onto which the target term CAREER was mapped.

In order to display more specifically which ontological nodes correspond to one another, this paper shows the mechanisms through using the instances from CAREER IS A PERSON

4 Defining the Source Domain of PERSON

In this section, we unfold how the source domain of PERSON was obtained. The expressions in (1) below were found related to the source domain of PERSON.

(1) 憂患(意識) ' the worried consciousness'
創傷 'wound'
放手 'let go'
生命力 'the force to live'
成長 'grow'
起(步) 'start a step'
(走)向 'walk towards'
第一(步) 'first step'

For these expressions, the search in WordNet and SUMO produced the explanations in Table 4:

Table 4: CAREER IS A PERSON and its related WordNet senses and SUMO Nodes

Expressions	WordNet Senses	Explanations	SUMO Nodes
憂患(意識)	1: consciousness	an alert cognitive state in which you are aware of yourself and your situation	Awake (清醒)
創傷	1: trauma	an emotional wound or shock often having long-lasting effects	EmotionalState (情緒狀態)
放手	1: let_go	be relaxed	EmotionalState (情緒狀態)
生命力	1: life_form	any living entity	Organism (生物體)
成長	3: mature	develop and reach maturity; undergo maturation	Growth (生長)
起(步)	1: pace	a step in walking or running	Walking (行走)
(走)向	1: foot	informal or colloquial synonyms of 'walk'	Walking (行走)
第一(步)	1: pace	a step in walking or running	Walking(行走)

From Table 3, all the metaphorical expressions are related to the following five nodes:

(2) (a) Awake (清醒)
(b) EmotionalState (情緒狀態)
(c) Organism (生物體)
(d) Growth (生長)
(e) Walking (行走)

Although no obvious pattern was found among the nodes in (2), when their definitions were searched for in SUMO, these nodes showed linkage to one another, as in (3). The linking concept, i.e., 'Organism,' is shaded in all the definitions in (3).

(3) (a) Awake: Attribute applies to Organisms that are neither Unconscious nor Asleep.
(b) EmotionalState: The Class of Attributes that denote emotional states of Organisms.
(c) Organism: Generally, a living individual, including all Plants and Animals.
(d) Growth: The Process of biological development in which an Organism or part of an Organism changes its form or its size.
(e) Walking: Any BodyMotion which is accomplished by means of the legs of an Organism on land for the purpose of moving from one point to another.

From investigating the SUMO definitions, all the expressions in Table 4 were found related to 'Organism' in one way or another. Among these five SUMO nodes in (3), 3(a) and (b) are related to the upper node of 'State of mind' or a psychological process. 3(c) and (d) are linked to the 'Organism' and the 'Organism Process' or biological process. 'Walking' in which (3e) is related 'Body motion' in SUMO.

What can be concluded from this discussion is that there were overlaps between all the expressions in Table 4 which were identified through the CM Model. The linking concept between them is 'Organism.' However, the concept of 'Organism' is too broad because it comprises all living things, including all plants and animals. For the conceptual metaphors, we are looking for source domains that are in contact with human conceptualization, i.e., more concrete concepts which human can easily recall when describing the abstract idea. Furthermore, we are not only looking for any organisms which have the abilities to grow and to walk, but one with a state of mind and emotional state. Based on these attributes, we rule out the possibility of PLANT and ANIMAL, but suggested that these expressions are related to HUMAN or PERSON.

5 PERSON in other Metaphor

In order to verify our findings regarding CAREER IS A PERSON, this paper also carried out the same procedures with the target domain of CULTURE. A total of 2000 instances was extracted from the Academia Balance Corpus of Mandarin Chinese. From these 2000 instances, 335 metaphorical expressions were extracted manually. From these 335 instances, 88 were found related to the source domain of PERSON. These instances constituted the following related WordNet senses and SUMO nodes (See Table 5).

Table 5: CULTURE IS A PERSON and its related WordNet senses and SUMO Nodes

Expressions	WordNet Senses	Explanations	SUMO Nodes
生命	3: life	the organic phenomenon that distinguishes living organisms from nonliving ones	Living (活的)
思想	2: thought	the organized beliefs of a period or group or individual	Proposition (命題)
成長	6: growth	the process of an individual organism growing organically; a purely biological unfolding of events involved in an organism changing gradually from a simple to a more complex level	Growth (生長)
反省	3: introspectiveness	thoughtfulness about your own situation and feelings	TraitAttribute (人格特質)
生命力	1: animation	the property of being able to survive and grow	Living (活的)
氣質	1: temperament	excessive emotionalism or irritability and excitability (especially when displayed openly)	PsychologicalAttribute (心理屬性)
命脈	1: lifeblood	the blood considered as the seat of vitality	Blood (血液)
心靈	2: soul	the immaterial part of a person; the actuating cause of an individual life	Human (人類)
(走) 向（v)	1: foot	informal or colloquial synonyms of 'walk' "	Walking (行走)
決定	3: decide	reach, make, or come to a decision about something	IntentionalPsychologicalProcess (意向性心理歷程)
行爲	4: behavior	manner of acting or conducting oneself	BodyMotion (身體的移動)
滅絕	1: die_out	become extinct	Death (死亡)

From Table 5, the SUMO nodes in the rightmost columns were found related to HUMAN (or PERSON). These nodes have the following definitions in SUMO, all of which are related to Organism in one way or another:

(4) (a) Living (活的): This Attribute applies to Organisms that are alive.
(b) Proposition (命題) :Propositions are Abstract entities that express a complete thought or a set of such thoughts.
(c) Growth (生長): The Process of biological development in which an Organism or part of an Organism changes its form or its size.
(d) TraitAttribute (人格特質): Attributes that indicate the behavior/personality traits of an Organism.
(e) PsychologicalAttribute (心理屬性): Attributes that characterize the mental or behavioral life of an Organism.
(f) Blood (血液): A fluid present in Animals that transports Nutrients to and waste products away from various BodyParts.

(g) Human(人類): Modern man, the only remaining species of the Homo genus.
(h) Walking(行走): Any BodyMotion which is accomplished by means of the legs of an Organism on land for the purpose of moving from one point to another.
(i) IntentionalPsychologicalProcess(意向性心理歷程): An IntentionalProcess that can be realized entirely within the mind or brain of an Organism.
(j) BodyMotion(身體的移動): Any Motion where the agent is an Organism and the patient is a BodyPart.
(k) Death(死亡): The Process of dying.

The target domain of CULTURE produced more instances than the CAREER metaphor. Therefore, a comparison can be made. From the information in Table 5, the first decision that the source domain could be PERSON was observed through the SUMO nodes, all of which can be related to HUMAN. The further investigation of their SUMO definitions in (4) found that these nodes are related to 'Organism' by possessing 'thoughts,' 'biological development,' 'animals,' 'body parts,' 'modern man,' 'Homo genus,' 'body motion,' 'mind or brain' and 'the process of dying.'

In other words, the analysis showed that there is a connectivity from the beginning linguistic expressions (metaphorical instances) to the WordNet explanations and SUMO nodes (Table 5) and lastly to the explanation of the SUMO nodes (example (4) above). The connectivity can be drawn by first defining the linguistic expressions within a source domain, as shown in Figure 1:

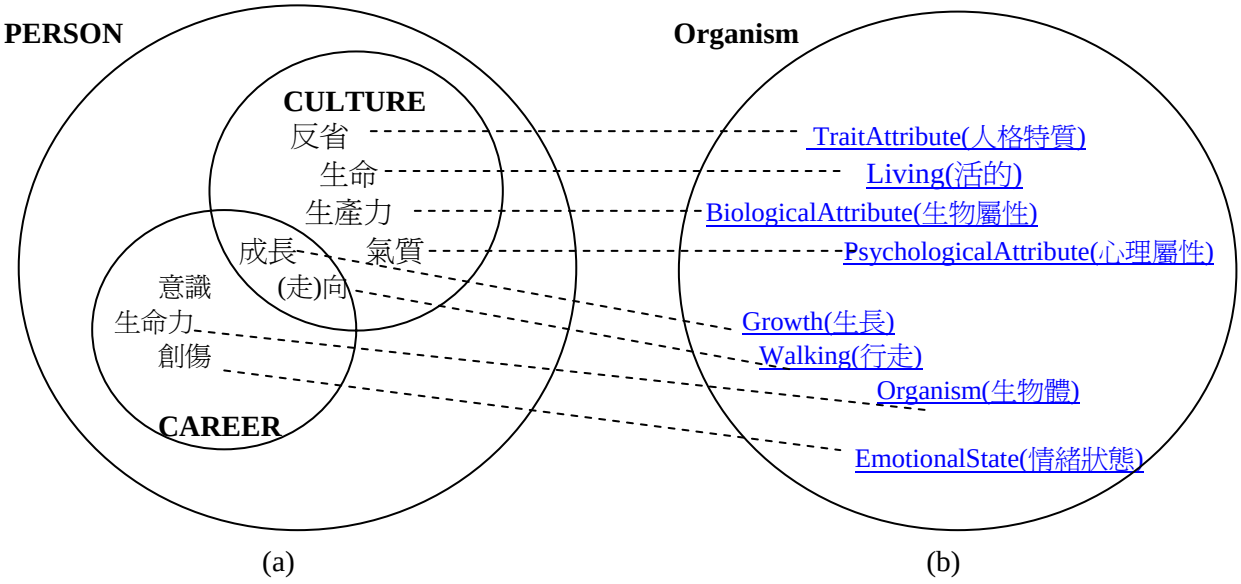

Figure 1: Representations of the Metaphorical Instances According to Source and Target Domians and their Relationship with WordNet and SUMO

The overlapping area between CULTURE and CAREER in Figure 1(a) constitutes lexical items that are more lexicalized as they can apply to more target domains than the other items. In other words, lexical items within several overlapping source domains (such as *chengzhang* 成長 'growth' and *zouxiang* (走)向 'walk towards') tend to be lexicalized faster than the other lexical items. All the lexical items in 1(a) can be mapped onto the ontological concepts of SUMO in 1(b).

6 Conclusion

The results from the WordNet lexical representation and SUMO searches prove two major points: First, the manual analysis made using the CM Model in previous studies can be made more automatized. Second, there is conceptual connectivity between the expressions found within a source domain. This connectivity can be established through observing the overlapped concepts in all the expressions within the same source domain.

The approach proposed by this paper takes a linguistic perspective in which instances are first analyzed based on a linguistic model of metaphors. In order to computerized the process of extracting source domain from the instances, a large amount of data has to be collected. The project has covered numerous source domains with corpora instances of metaphorical examples. When these instances are fully coded with the linguistic information, a computerized program can be designed for the extraction purposes.

From the perspective of the linguists, this analysis provides the evidence to pinpoint at which level of knowledge conceptual metaphors occur. From the perspective of anthropology and language processing, this study provides the linguistic evidence to understand the mental representation of human when expressing metaphors.

Acknowledgements

We would like to thank the CLSW-5 reviewers for their comments on this paper.

References

Ahrens, Kathleen. 2002. "When Love is Not Digested: Underlying Reasons for Source to Target Domain Pairings in the Contemporary Theory of Metaphor." In Yuchau E. Hsiao (ed.). *Proceedings of the First Cognitive Linguistics Conference*. Cheng-Chi University, 273-302.

Ahrens, Kathleen, Chung Siaw-Fong, Huang Chu-Ren. 2003. "Conceptual Metaphors: Ontology-based Representation and Corpora Driven Mapping Principles." In the *Proceedings of the ACL Workshop on the Lexicon and Figurative Language*. pp. 35-41.

Chung, Siaw Fong, Kathleen Ahrens and Huang Chu-Ren. 2003a. "ECONOMY IS A PERSON: A Chinese-English Corpora and Ontological-based Comparison Using the Conceptual Mapping Model." *Proceedings of the 15th ROCLING Conference*, National Tsing-Hwa University, Taiwan.

Chung, Siaw Fong, Kathleen Ahrens and Huang Chu-Ren. 2003b. "ECONOMY IS A TRANSPORTATION_DEVICE: Contrastive Representation of Source Domain Knowledge in English and Chinese." *Proceedings of the special session of UONLP, 2003 International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE)*, Beijing.

Huang, Chu-Ren, Ru-Ying Chang, Shiang-Bin Lee. 2004 "Sinica BOW (Bilingual Ontological Wordnet): Integration of Bilingual WordNet and SUMO." To appear in the *Proceedings of the 4th International Conference on Language Resources and Evaluation* (LREC2004). Lisbon. Portugal. 26-28 May, 2004.

Lakoff, George and Mark Johnson. 1980. *Metaphors We Live By*. Chicago: The University of Chicago Press.

Niles, I. and Pease, A. 2003 "Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology." In *Proceedings of the 2003 International Conference on Information and Knowledge Engineering*,, Las Vegas, Nevada.