From Lexical Semantics to Conceptual Metaphors: Mapping Principle Verification with WordNet and SUMO

Kathleen Ahrens
National Taiwan University
No. 1, Sec. 4, Roosevelt Road
Taipei, 106 Taiwan
ahrens@ntu.edu.tw

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

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

Abstract

The goal of this paper is to further develop methods for verifying Mapping Principles between source and target domain pairings of conceptual metaphors. Previous work (Ahrens, Chung & Huang, 2003) integrated the Conceptual Mapping Model with an ontology-based knowledge representation (i.e. SUMO) in order to demonstrate that conceptual metaphor analysis could be restricted and eventually, automated. However, that study relied on an operational definition that required a large number of instances of source-target domain pairings in order to determine the most frequent mapping. In this study, we further operationalize the frequency-based definition and examine the possibility of using information derived from WordNet and SUMO to verify Mapping Principles for source-target domain pairings that do not occur frequently enough (if at all) in a corpus.

1 Introduction

Lexical resources are central to any natural language processing system and the plethora of multi-lingual resources such as EuroWordNet (EWN) and Academia Sinica's Bilingual Ontological WordNet (BOW) are based firmly in the lexicon. Recent work in figurative language processing, however, has pointed out shortcomings of these resources for processing of conceptual metaphor. Veale (2003), for example, points out that determining the *aptness* is dependent upon parameters that are, as yet, hidden. Lonneker (2003) decries the lack of relationship between literal and figurative synsets in WordNet, while Alonge and Castelli (2003) suggest that the EWN Top Ontology needs to be extended with more concepts in order to deal with figurative language, since at this point in time it only notes general semantic distinctions.

Ahrens (2002), taking an intuition-based approach, proposed that the linguistic expressions that are used metaphorically can be analyzed in terms of the entities, qualities and functions that can map between a source and a target domain. When these conventionalized metaphorical expressions have been analyzed, an underlying reason for these mappings can then be postulated. This study relied on native speaker intuition regarding what is mapped conceptually from the source to the target domain. For example, in the four examples from the metaphor LOVE IS PLANT, given below, the Mapping Principle (MP) of "Love is understood as plant because plants involve physical growth and love involves emotional growth" was extracted based on the fact that all the examples in some way had to do with growth.

1. (a) *liang ren de ai miao zuijin cai gang* <u>*mengya*</u>
two people MOD love seedling lately just recently sprout
'Their love just begins to sprout lately.'

(b) *wo dui tade ai-yi jianjian* <u>*zizhang*</u>
I for his love gradually grow
'My love for him has grown gradually.'

(c) *aiqing xuyao xinqin de quanqai*
love need industriously water
'Love needs to be watered industriously.'

Ahrens, Chung & Huang (2003), followed up on this study and presented a corpus-based approach to determining the systematicity between source and target domain pairings (i.e. Mapping Principles (MPs)). They propose that each source-target domain pairing will have a prototypical instance of mapping as indicated by a lexical item that is frequently mapped, as compared with other mappings. In addition they use the Suggested-Upper-Merged-Ontology (SUMO) to delimit the source domain knowledge. SUMO has the advantages of being able to be used to infer knowledge through automatic reasoning as well as to constrain the falsifiability of the MP.

For example, when they looked at the metaphor ECONOMY IS A BUILDING, they find 102 instances (in their randomly extracted corpora of 2000 instances from Sinica Corpus) of this source-target domain pairing. Of these instances, 39 are mapped with the lexical item *jianshe* 'construction', and the suggested Mapping Principle is therefore: *Economy is a building because buildings involve a (physical) structure and economy involves an (abstract) structure* (p. 38)." However, clear numerical criteria for determining the minimal number of tokens necessary, as well as the percentage of tokens necessary to reach consensus for a mapping principle was not established. In addition, if enough conventional metaphor examples cannot be found in the corpora, there is currently no way to verify the intuition-based mapping principle. Thus, the determination of mapping principles by means of frequency needs to be researched further. In this paper, then, we will create guidelines for determining when MPs can be established on the basis of frequency, as well as extend Ahrens et al.'s (2003) scope and examine the possibility of using information derived from WordNet and SUMO to verify MPs for source-target domain pairings that do not occur frequently enough in a corpus.

2 Using Frequency to Verify Mapping Principles

In this section, we attempt to determine frequency criteria for verifying MPs (Table 1). Two thousand examples were extracted for ECONOMY metaphors from the Academia Sinica Balanced Corpus (http://www.sinica.edu.tw/SinicaCorpus/), and then manually sorted into metaphorical and literal examples. In addition, we also look at data on STOCK MARKET extracted from on-line newspapers in Taiwan (Chung, Ahrens & Sung 2003).

Table 1: Total Number of Metaphorical Instances Extracted from 2000 Samples

Metaphor	**Total # of Tokens**	**Most Frequent**	**Percentage**	MPs match
ECONOMY IS A PERSON	121	88	72%	Yes
ECONOMY IS BUILDING	102	39	38%	Yes
STOCK MARKET IS OCEAN WATER	58	53	91.00%	Yes
ECONOMY IS COMPETITION	40	14	35%	Yes
STOCK MARKET IS WAR	25	2	8%	Not clear
STOCK MARKET IS A BALL	24	17	71.00%	Yes
ECONOMY IS WAR	23	5	21%	Not clear
STOCK MARKET IS A PERSON	19	2	10%	Not clear
STOCK MARKET IS EARTHQUAKE	17	16	94.12%	Yes
STOCK MARKET IS DISASTER	16	11	68.75%	Yes
ECONOMY AS JOURNEY	15	5	33%	Yes
ECONOMY IS AN AEROPLANE	10	8	80%	Yes

The second column in Table 1 lists the total number of metaphorical examples found for each source domain. Next, examples with the same lexical item were counted. For example, in ECONOMY IS A

PERSON, 'chengzhang' *growth* is found 88 times or in 72% of the data. In ECONOMY IS A BUILDING, 'jianshe' *construction* occurs 39 times (or in 38% of the data). In both these cases (and in all the shaded cases in the table) the most frequent lexical item is related to what was originally postulated as a Mapping Principle on the basis of linguistic intuition.

However, in the three non-shaded cases, the relationship between the frequency of the lexical items and the Mapping Principle is not straightforward. For example, in STOCK MARKET IS WAR, no one lexical item receives priority in terms of frequency. In ECONOMY IS WAR, there are five instances of 'gua-shuai' *to take position of the commander in the army*, but this does not clearly relate to the MP of "Economy is war because war involves a violent contest for territorial gain and the economy involves a vigorous contest for financial gain." Therefore, we propose that the following two conditions be met for a frequency-based account to be valid: First, at least 10 tokens must found and, in addition, at least 30% of these tokens must be of one particular lexical item.

3 Using WordNet and SUMO to Verify Mapping Principles

In the cases where there are fewer than 10 instances found in the corpora, or in cases where the percentage of a single lexical item is less than 3o%, we need to find another method to verify MPs. Our goal is aided by studies linking the lexicon in WordNet and SUMO (http://ontology.teknowledge.com, Niles & Pease, 2003). After the links are established, the ontological nodes in SUMO provide 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 (http://bow.sinica.edu.tw).

We examine two instances here: LOVE IS A PLANT (Table 2) and LOVE IS FOOD (Table 3).

Table 2: LOVE IS PLANT Definitions from WordNet and SUMO

Items	**WordNet Senses**	**Explanation**	**SUMO Category**
萌芽	**2: sprout**	grow sprouts, of a plant	Growth(生長)
苗	**1: seedling**	young plant or tree grown from a seed	FloweringPlant (開花植物)
長	**1: grow**	come to have, of physical features and attributes	Growth(生長)
灌溉	**1: water**	pour water on	Wetting(濕潤)
開花	**1: bloom**	produce or yield flowers	Growth(生長)

Ahrens (2002) proposes the following MP for LOVE IS A PLANT: "Love is understood as plant because plants involve physical growth and love involves emotional growth." Since corpora searches do not come up with any instances of this metaphor, it is difficult to ascertain the validity of this principle. We therefore propose looking at the 1) WordNet sense, 2) the WordNet definition, and the 3) SUMO node for the WordNet sense (Niles & Pease 2003) for the intuition-based examples in order to see if there are any semantic overlaps within, or between, these three types of information. Table 2 shows that the word "Growth" appears three times in the SUMO category, out of the five examples. "Grow" also appears three times in the sense and definition columns from WordNet. Thus, the combination of WordNet information and the SUMO representation agrees with the MP originally given.

In another example that has less than ten corpora examples, LOVE IS FOOD (Table 3), both the WordNet information and the SUMO information again matches up with the Mapping Principle suggested in Ahrens (2002), that "Love is understood as food because food has different tastes as love involves different feelings." Table 3 shows that *taste* is mentioned five times in the WordNet sense and definition, and three out of four times in the SUMO category. Thus, determining the number of overlapping lexical items in WordNet definitions and SUMO categories to verify Mapping Principles seems to hold promise for instances where there are not enough exemplars to make a judgment based on frequency alone.

Table 3: LOVE IS FOOD: Definitions from WordNet and SUMO

Items	**WordNet Senses**	**Explanation**	**SUMO Nodes**
滋味	**1: taste**	distinguishing a taste by means of the taste buds	Tasting(嚐)
苦	**1: bitter**	one of the four basic taste sensations; sharp and disagreeable; like the taste of quinine	TasteAttribute (味覺屬性)
味道	**1: taste**	distinguishing a taste by means of the taste buds	Tasting(嚐)
吃	**1: eating**	the act of consuming food	Eating(吃)

4 Testing the Verification Principle

In Section 2, we pointed out that there were three instances in the STOCK MARKET and ECONOMY metaphors that did not meet the 30% criteria. We look at these three metaphors below in order to determine if WN and SUMO can be used to verify the MPs.

Table 4: ECONOMY IS WAR: Definitions from WordNet and SUMO

Items	**WordNet Senses**	**Explanation**	**SUMO Nodes**
侵略	**4: invasion**	the act of invading; the act of an army that invades for conquest or plunder	ViolentContest (暴力性的競爭)
戰	**1: war**	the waging of armed conflict against an enemy	War (戰爭)
老兵	**1: veteran**	a serviceman who has seen considerable active service	SocialRole (社會角色)
攻防戰	**x**		
策略	**6: ambush**	the act of concealing yourself and lying in wait to attack by surprise	ViolentContest (暴力性的競爭)
大權	**x**		
千瘡百孔			
全力(衝刺)	**x**		
掛帥	**x**		
大權…操在…的手上	**x**		
犧牲	**1: sacrifice**	kill or destroy	Killing(殺害)
犧牲品	**1: sacrifice**	personnel that are sacrificed (e.g., surrendered or lost in order to gain an objective)	Human(人類)

Table 4 shows that there are three instances of the concept of 'Invasion' found in the WN definitions, but they are all in the same definition. An alternate hypothesis is that 'ViolentContest' is the critical issue since it occurs in the SUMO nodes of two different words. (The blank cells indicate that this word was not found in BOW and so does not have a listed sense or associated SUMO node.) In addition, Ahrens et al. (2003) noted that ECONOMY IS WAR is a subset of the ECONOMY IS CONTEST metaphor, with the MP of 'Economy is war because war involves a violent contest for territorial gain and the economy involves a vigorous contest for financial gain.' Moreover, the SUMO node of WAR is linked to ViolentContest. This example demonstrates that not only do we need to have an expansion of the SinicaBow to link to more items in WordNet, we also need to expand our notion of semantic space to include related Sumo nodes. In sum, our current analysis suggests that the previous MP was correct.

Table 5 gives instances of STOCK MARKET IS WAR. Although different lexical items are mapped as compared with ECONOMY IS WAR, the mapping of the concept of 'contest' is the same.

Table 5: STOCK MARKET IS WAR: Definitions from WordNet and SUMO

Items	**WordNet Senses**	**Explanation**	**SUMO Nodes**
戰	**1: war**	the waging of armed conflict against an enemy	War (戰爭)
策略	**6: ambush**	the act of concealing yourself and lying in wait to attack by surprise	ViolentContest (暴力性的競爭)
焦土	**x**		
地雷	**1: land_mine**	an explosive mine hidden underground; explodes when stepped on or driven over	Weapon (武器)
攻勢	**x**		
關卡	**1: checkpoint**	a place (as at a frontier) where travellers are stopped for inspection and clearance	LandArea (陸地)
大關	**x**		
防衛	**2: defend**	be on the defensive; act against an attack	Contest (競爭)
功臣	**x**		
全軍覆沒	**x**		
殺空	**x**		
撤出	**x**		
殺出	**x**		
上(攻)	**1: attack**	take the initiative and go on the offensive: "The Serbs attacked the village at night"	Contest (競爭)
挑戰	**4: challenge**	a call to engage in a contest or fight	Requesting (請求)
撤離	**1: evacuation**	the act of evacuating; leaving a place in an orderly fashion; esp. for protection	Motion (移動)

Thus, the proposed MP is the same; ‘Stock market is war because war involves a violent contest for territorial gain and the stock market involves a vigorous contest for financial gain.’ Note that this does not violate the Mapping Principle Constraint proposed in Ahrens (2002) since two different *target* domains are recruiting the same source domain for the same reason. The Mapping Principle Constraint says that ‘A target domain will select only source domains that involve unique mapping principles.’ This constraint applies to the source domains that are selected by a single target domain.

The last metaphor in our study that does not meet the frequency requirement of 30% is STOCK MARKET IS A PERSON. Although no one lexeme predominates, variants of the word ‘life’ occur several times. It is interesting to note that although in the previous examples a case could be made for looking solely at the SUMO nodes (cf. Tables 2-5 above), in this instance, only the words used in the WN definition are found to coalesce around a concept. In this instance, we suggest that the MP is: *Stock market is a person because people have varying degrees of physical activity (liveliness) and the stock market has varying degrees of financial activity*. It may also be beneficial to look at whether analyzing the relationship between the words given in the definitions (i.e. if some words are near-synonyms) would aid in determining the underlying reason for the source-target domain pairing. However, another possibility is that this source domain is too broad or not clearly delimited. In fact, one step that needs to be taken in this regard is to incorporate the steps that Chung, Ahrens and Huang (2004) take to delimit source domains before analyzing MPs.

Table 6: STOCK MARKET IS A PERSON: Definitions from WordNet and SUMO

Items	**WordNet Senses**	**Explanation**	**SUMO Nodes**
活力	**4: pep**	liveliness and energy	BiologicalAttribute (生物屬性)
創傷	**2: wound**	any break in the skin or an organ caused by violence or surgical incision	Injuring(傷害)
瀕 (死) 的	**2: die**	pass from physical life and lose all bodily attributes and functions necessary to sustain life	Death(死亡)
敏感的	**x**		
疲軟	**x**		
激勵	**12: spur**	give heart or courage to	IntentionalPsychologicalProcess (意向性心理歷程)
輸血	**x**		
起死回生	**1: resurrect**	cause to become alive again	OrganismProcess (生命歷程)
變 (臉)	**1: face**	the front of the head from the forehead to the chin and ear to ear	BodyPart (軀體部件)
應聲倒地, 應聲而倒	**x**		
復甦	**1: come_to**	return to consciousness: "The patient came to quickly"	BiologicalProcess (生物歷程)
失血	**x**		

5 Conclusion

We propose the following steps to determine Mapping Principles: First, if the number of tokens of a conceptual metaphor is greater than 10, and the highest number of tokens mapped is greater than 30% of all tokens, then this lexical item is the basis for the MP. Second, if either the first or second criteria are not met, determine the sense of each item in WordNet (extracted through SinicaBow) and the associated SUMO category for that particular sense. Next calculate the number of times that a particular lexical item is used in the WN sense, explanation, or linked SUMO category and determine if that lexical item is related to what intuitively seems to be mapped between the source and target domains in the example sentences.

Further research will focus on examining whether it is necessary to extend the parameters of the second step to include synonyms, thus expanding the semantic space by which a MP can be determined. In addition, the biggest hurdle this type of account faces is that the WN sense must currently be stipulated manually, although we hope that the implementation of semantic tagging can overcome this problem in the future. We hope that the line of research discussed herein will stimulate more research in how a computational approach can help set parameters for determining metaphorical sense and point the way to the application of creating a systematic relationship between literal and figurative synsets in WordNet.

Acknowledgements

We would like to thank the CLSW-5 reviewers for their comments on this paper and NSC grant #NSC92-2411-H-002-076-ME for supporting the research discussed herein. Any remaining errors are the sole responsibility of the authors.

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 *Proceedings of the ACL Workshop on the Lexicon & Figurative Language*. pp. 35-41.

Alonge, Antonietta & Margherita Castelli. 2003. "Encoding Information on Metaphoric Expressions in WordNet-like Resources." In *Proceedings of the ACL Workshop on the Lexicon and Figurative Language*. pp. 10-17.

Chung, Siaw-Fong, Kathleen Ahrens and Chu-Ren Huang. 2004. "Using WordNet and SUMO to Determine Source Domains of Conceptual Metaphors." In *Proceedings of the 5th Chinese Lexical Semantics Workshop (CLSW-5)*, Singapore. June 14-15.

Chung, Siaw Fong, Kathleen Ahrens and Sung Yahui. 2003. "STOCK MARKETS AS OCEAN WATER: A Corpus-based, Comparative Study in Mandarin Chinese, English and Spanish." In the *Proceedings of the 17th Pacific Asia Conference on Language, Information and Computation (PACLIC)*, Singapore. pp. 124-133.

Lonneker, Birte. 2003. "Is There a Way to Represent Metaphors in WordNets? Insights from the Hamburg Metaphor Database." In the *Proceedings of the ACL Workshop on the Lexicon and Figurative Language*. pp. 18-26.

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

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.

Veale, Tony. 2003. "Systematicity and the Lexicon in Creative Metaphor." In *Proceedings of the ACL Workshop on the Lexicon & Figurative Language*. pp. 27-34.