Conceptual Metaphors: Ontology-based representation and corpora driven Mapping Principles

Kathleen Ahrens
National Taiwan University
kathleenahrens@yahoo.com

Siaw Fong Chung
National Taiwan University
claricefong6376@hotmail.com

Chu-Ren Huang
Academia Sinica
churen@sinica.edu.tw

Abstract

The goal of this paper is to integrate the Conceptual Mapping Model with an ontology-based knowledge representation (i.e. Suggested Upper Merged Ontology (SUMO)) in order to demonstrate that conceptual metaphor analysis can be restricted and eventually, automated. In particular, we will propose a corpora-based operational definition for Mapping Principles, which are explanations of why a conventional conceptual metaphor has a particular source-target domain pairing. This paper will examine 2000 random examples of 'economy' (*jingji*) in Mandarin Chinese and postulate Mapping Principles based frequency and delimited with SUMO.

1 Introduction

A theory of metaphor has been the focus of study on lexical and figurative meaning for the past two decades. Are conventional conceptual metaphors a cognitive rather than a linguistic phenomenon? Work within Cognitive Linguistics would seem to say that this is the case. For example, Lakoff (1993) writes with respect to the source-target domain mapping of the conventional conceptual metaphor LOVE IS A JOURNEY:

> Is there a general principle governing how these *linguistic expressions* about journeys are used to characterize love…. [Yes], but it is a *general principle* that is neither part of the grammar of English, nor the English lexicon. Rather it is *part of the conceptual system underlying English*…. (page 306, italics added)

Thus, the onus of dealing with metaphorical meaning in the lexicon is not necessary. Metaphor must be treated at a different (i.e. higher) cognitive level.

But is it really the case that there are no general principles that can be extracted and proposed at the lexical level? The Conceptual Mapping (CM) Model (Ahrens 2002) was proposed to constrain the Contemporary Theory of Metaphor (Lakoff 1993). This model analyzes the linguistic correspondences between a source and target (knowledge) domain in order to determine the underlying reason for the source-target pairings. The underlying reason is formulated in terms of a Mapping Principle. The theory also postulates a Mapping Principle Constraint, which says that a target domain will select only source domains that involve unique mapping principles. For example, the target domain of IDEA uses the source domains of BUILDING and FOOD, but it does so for different reasons (as we will discuss in the next section). With the addition of this constraint, the CM model is able to explicate the polysemy inherent in a given target domain. In addition, the CM Model presupposes that Mapping Principles are conventionalized linguistically but not conceptualized *a priori*. This model is supported in psycholinguistic experiments because it correctly predicted the processing differences involved between conventional and novel metaphors (Ahrens 2002). In this paper, we propose a new approach to conceptual metaphors that incorporates two computationally trackable elements. First, the data analysis is corpus-based, following the example of MetaBank

(Martin 1992). Second, the representation is ontology-based. Both elements strengthen the empirical basis of the account.

In this paper, we propose that the most frequent mapping instance within a source domain indicates the basis of the reason for the source-target domain pairing, i.e. the mapping principle. We test this empirical prototype (EP) hypothesis by running extracting a dataset of 2000 examples of *jingji* 'economy' from the Academia Sinica Balanced Corpus. We hypothesize that each source-target domain pairing will have a prototypical instance of mapping as evidenced by an individual lexical item that is highly frequent as compared with other mappings. In addition, we propose using an ontological-based knowledge representation, such as SUMO, to define and delimit the source domain knowledge in the CM Model. This has the advantage of using SUMO to infer knowledge through automatic reasoning, and as well as constraining the scope and falsifiablity of the conceptual metaphor.

2 The Conceptual Mapping Model and Ontology

Ahrens (2002) proposed that the question asked by Lakoff above ('Is there a general principle governing how these *linguistic expressions* about journeys are used to characterize love?') should be answered by examining the lexical correspondences that exist between a source and target domain. She proposes 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, they are compared with the real world knowledge that the source domain entails, and an underlying reason for these mappings is then postulated.

For example, she points out that in the conceptual metaphor IDEA IS BUILDING in Mandarin, the linguistic expressions relating to the concept of foundation, stability and construction were mapped (i.e. are conventional linguistic examples) while concepts relating to position of the building, internal wiring and plumbing, the exterior of the building, windows and doors were not (and these are the concepts that are in the real world knowledge of the source domain). Thus she postulated that the target domain of IDEA uses the source domain of BUILDING in order to emphasize the concept of structure. Thus, when someone talks about ideas and want to express positive notions concerning organization, they use the source domain of BUILDING. The Mapping Principle formulated in this case was therefore the following:

(1) Mapping principle for IDEA IS BUILDING:
Idea is understood as building because **buildings involve a (physical) structure and ideas involve an (abstract) structure.** (Ahrens 2002)

When IDEA is talked about in terms of FOOD, however, the expressions that are mapped are 'ingredient', 'spoil', 'flavorless', 'full', 'taste', 'chew', 'digest' and 'absorb'. Mandarin Chinese, in contrast with English, does not have conventional expressions relating to 'cooking' or 'stewing' of ideas. Thus, the postulated Mapping Principle is: Idea is understood as food because **food involves being eaten and digested (by the body) and ideas involved being taken in and processed (by the mind)** (Ahrens 2002).

Thus, IDEA uses the source domains of BUILDING and FOOD for different reasons, namely to convey information related to 'structure' or 'processing' (i.e. 'understanding') respectively. Thus, it is similar to the Contemporary Theory of metaphor in that it supposes that there are systematic mappings between a source and target domain, but it goes a step further in postulating an underlying reason for that mapping. The CM Model predicts that conventional metaphors, novel metaphors that follow the mapping principle and novel metaphors that don't follow the mapping principle will be rated differently on interpretability and acceptability scales when other factors, such as frequency are controlled for. This was, in fact, found to be the case (Ahrens 2002). Other theories of metaphor processing such as Gentner's Structure Mapping Model (Gentner and Wolff 2000), or the Attributive Categorization Hypothesis (McGlone 1996) do not distinguish between novel and conventional metaphors, nor do they suppose that there might be different types of novel metaphors.

The CM model of metaphor presupposed structured shared source domain knowledge. For a mapping to be conventionalized and understood by speakers, the content and structure of the source domain knowledge must be *a priori* knowledge and should not have to be acquired. How to define and verify such structured knowledge is a challenge to this theory. We attempt to meet this challenge in two ways: first, by assuming that source domain knowledge representation is instantiated by a shared upper ontology, such as SUMO. If the source domain knowledge representation is indeed ontology-based, we can adopt the null hypothesis that the mapping principle is based on one of the inference rules encoded on that particular conceptual node. In consequence, we can take the second step by examining actual mappings of linguistic expressions in corpora, and extract the most frequent mappings to verify the null hypothesis. This will also allow us to investigate if it is the case that frequency of use in a newspaper corpora necessarily reflects the underlying mapping principle, an issue which is currently open to interpretation.

The integration of an upper ontology to the CM model has the following theoretical implications:

First, the source domain knowledge representation is now pre-defined and constrained. Second, the validity of such hypothesis will in turn support the robustness and universality of the proposed upper ontology.

3 SUMO

SUMO (Suggested Upper Merged Ontology – http://ontology.teknowledge.com) is a shared upper ontology developed by the IEEE sanctioned IEEE Standard Upper Ontology Working Group. It is a theory in first-order logic that consists of approximately one thousand concepts and 4000 axioms. Its purpose is to be a shared and inter-operable upper ontology (Niles and Pease 2001, Pease and Niles 2002, Sevcenko 2003) Since ontologies are formalized descriptions of the structure of knowledge bases, SUMO can also be viewed as a proposed representation of shared human knowledge, and thus a good candidate for mapping information about the source domain to the target domain. What we will look at below is whether the SUMO conceptual terms and inferences are candidates for knowledge representation in the source domain. In order to analyze this, we first need to extract from a corpora the linguistic terms that are used for mappings between a source and a target domain.

The application of SUMO in NLP and in processing of lexical meaning is facilitated by its interface with WordNet. The SUMO interface allows users to search and map each English lexical meaning defined in WordNet to a concept node on the SUMO ontology. Similarly, one can also search for a Chinese lexical meaning and map it to a SUMO concept node through a Chinese-English bilingual translation equivalents database (http://ckip.iis.sinica.edu.tw/CKIP/ontology/).

4 Corpora Data

In order to test the feasibility of using SUMO to aid the analysis of Mapping Principles within the framework of the CM Model, we searched the Academia Sinica Balanced Corpus, a tagged corpus of over 5 million words of modern Mandarin usage in Taiwan (available on the Internet: http://www.sinica.edu.tw/SinicaCorpus/). The maximum number of responses (i.e. 2000) was obtained for the word '*jingji'* (economy) in Mandarin Chinese. Each of these 2000 was examined and all metaphorical instances were marked. (A metaphorical instance is defined as when an abstract concept such as 'economy' is discussed in terms of a concrete concept, such as 'building'.) All instances of concrete concepts were then grouped into source domains. All source-target domain pairings that had more than 20 instances were then examined. In Tables 1-4 below we show the source domains that were found for *jingji* 'economy' and we give the total number of instances and the number of tokens for each metaphor, as well as a proposed mapping principle based. Also note that the following mappings were manually analyzed and classified.

We first note that the EP (empirical prototype) hypothesis holds up since in all source-target domain pairings except for in ECONOMY IS WAR in Table 4. In the remaining three metaphors, there is one or two lexical items that is/are obviously more frequent than the others.

Table 1: **ECONOMY IS A PERSON** (121 instances)
M.P.: Economy is person because people have a life cycle and economy has growth cycle.

	Metaphor	**Freq.**
Entities	***Chen2zhang3* (growth)**	**67**
	Shuai1tui4 (regression/decay)	8
	Chen2zhang3chi2 (growth period)	2
	Bing4zhuang4 (symptoms)	1
	Ming4ma4i (lifeblood)	2
Quality	*Shuai1tui2* (weaken and degenerate)	1
Functions	*Chen2zhang3* (grow)	21
	Shuai1tui4 (to become weaker)	5
	Fu4shu1 (regain consciousness)	9
	E4hua4 (deteriorate)	4
	Hui1fu4 (recover)	1

Thus, for ECONOMY IS A PERSON, the mapping principle is postulated to have to do with the life cycle of a person (and not, for example, the mental health of a person) because of the frequent occurrence of the lexical item 'chengzhang' (growth).

Table 2: ECONOMY IS A BUILDING (102 instances)
M.P.: Economy is building because buildings involve a (physical) structure and economy involves an (abstract) structure.

	Metaphors	**Frequency**
Entities	***jianshe* (construction)**	**39**
	jiegou (structure)	20
	jiqu (foundation)	15
	zhichu (pillar)	1
	genji (foundation)	2
	guimo (model)	5
	chuxing (model)	1
Qualities	*wengu* (firm)	2
	wending (stable)	8
Functions	*chongjian* (re-build)	9

In the case of ECONOMY IS A BUILDING the mapping principle is postulated to having to do with structure, and not for example, leaky plumbing. This is an interesting case because, as mentioned above, Ahrens (2002) examined IDEA IS A BUILDING and postulated that the mapping principle also had to do with structure (i.e the structure of a building and the structure of ideas). As Ahrens (2002) points out, it is not always the case that different target domains use the same aspect of a source domain. For example, the source domain of FOOD is used differently for IDEAS (to express the notion of digestion and processing) as compared with LOVE which uses FOOD to compare different tastes to different feelings.

Table 3: ECONOMY IS A COMPETITION (40 instances)
M.P.: Economy is competition because a competition involves physical and mental strength to defeat an opponent and an economy requires financial strength in order to prosper against other economies.

	Metaphors	**Frequency**
Entities	***shili* (actual strength)**	**14**
	***jingzheng* (competition)**	**12**
	jingzhengyoushi (advantage in competition)	3
	ruozhe (the weak one)	2
	jingzhengli (power of competition)	3
	ruoshi (a disadvantaged situation)	1
	qiangguo (a powerful nation)	1
	douzheng (a struggle)	2
	tuishi (a declining tendency)	1
Function	*shuaibai* (to lose)	1

Thus, for ECONOMY IS A COMPETITION, the emphasis is on the strength of participant in order to defeat the opponent.

Table 4: ECONOMY IS WAR (23 instances)
M.P.: Economy is war because war involves a violent contest for territorial gain and the economy involves a vigorous contest for financial gain.

	Metaphors	Frequency
Entities	*qinglue* (invasion)	4
	zhan (battle)	2
	laobing (veteran)	1
	gungfangzhan (defend and attack battle)	1
	chelue (tactics)	1
	daquan (immense power)	4
Qualities	*qianchuangbaikong* (one thousand boils and a hundred holes; holes all over)	1
Functions	*quanlichongchi* (to dash with full force)	1
	guashuai (to take command)	5
	(daquan) chaozai shoushang (to grasp the power)	1
	xisheng (sacrifice)	1
	Xishengping (victims)	1

In ECONOMY IS WAR, there is no clear-cut instance of a frequent mapping. We suggest that this is because WAR is a subset of the source domain of COMPETITION (i.e. a violent contest) in the SUMO representation, as discussed below.

In short, the corpora data support the CM model's hypothesis that there is a subset of linguistic expressions within a particular source domain that map to a target domain. It is not the case that 'anything goes'. In fact, the corpora data presented above, suggest an even more restricted view – that there are usually one or two linguistic expressions that frequently map between the source and target domains and 'drive' the motivating relationship between them. In the next section, we look at whether or not the source domain knowledge can be defined *a priori* through an upper ontology such as SUMO.

5 Defining Source Domain Knowledge with Shared Upper Ontology

The research on Shared Upper Ontology offers a potential answer to the challenge of how to define and verify the structured knowledge in a source domain. A shared upper ontology is designed to represent the shared knowledge structure of intelligent agents and allows knowledge exchange among them. In computational application, it is an infrastructure for knowledge engineering. In cognitive terms, we can view it as a candidate for he description of shared human knowledge. In this paper, we adopt SUMO.

In SUMO, conceptual terms are defined and situated in a tree-taxonomy. In addition, a set of first order inference rules can be attached to each conceptual node to represent the knowledge content encoded on that term. The conceptual terms of SUMO are roughly equivalent to the source domains in MP theory. Hence the well-defined SUMO conceptual terms are candidates for knowledge representation of the source domain in the MP theory of metaphor. In other words, SUMO provides a possible answer the question of how source domain knowledge is represented and how does this knowledge allows the mapping in conceptual metaphors. We examine how this might be possible by looking at two conceptual terms that are represented in SUMO that related to our source domains – CONTEST and ORGANISM.

Economy is Contest

First, we found that what we intuitively termed as 'competition' above has a corresponding ontological node of Contest. The term Contest is documented as 'A SocialInteraction where the agent and patient are CognitiveAgents who are trying to defeat one another.' Its only inference rule is quoted here:

(=> (instance ?CONTEST Contest) (exists (?AGENT1 ?AGENT2 ?PURP1 ?PURP2) (and (agent ?CONTEST ?AGENT1) (agent ?CONTEST ?AGENT2) (hasPurposeForAgent ?CONTEST ?PURP1 ?AGENT1) (hasPurposeForAgent ?CONTEST ?PURP2 ?AGENT2) (not (equal ?AGENT1 ?AGENT2)) (not (equal ?PURP1 ?PURP2)))))

The knowledge inference rule stipulates that each instance of Contest is carried out by two agents, each has his own non-equal purpose. This is exactly the source knowledge needed for the metaphor mapping. When the conceptual metaphor is linguistically realized, lexical expressions are then chosen to represent the conceptual terms of both purposeful agents, and conflicting purposes for the agents. Notice that in contest, as in economy, it is not necessary to have only one winner. There may be multiple winners and perhaps no winners. In other words, the agents' purpose may not be conflicting. But the purposes-for-agent are definitely different for each agent.

In addition to the 40 instances of economy metaphors involving contest. There are also 23 instances of metaphors involving War. In these cases, it is interesting to observe that the central concept is still the conflicting purposes (one's gain is another's loss) of the warring party. This is confirmed by the shared ontology. In SUMO, a War is a kind of ViolentContest, which in term is a kind of Contest.

Contest—ViolentContest—War

The term War is defined as 'A military confrontation between two or more Nations or Organizations whose members are Nations.' And the term ViolentContest is defined as 'Contest where one participant attempts to physically injure another participant.' As can be seen from the definition and the metaphoric uses involving War, the ontological source domain knowledge is not involved.

In fact, when examined more closely, it is clear that when the domain knowledge of War is used, it either further specifies the conflicting purposes by elaborating on the quality and manner of the conflict, or elaborating on the agent participants as combatants. In other words, Economy is War is not a different mapping. It is subsumed under the mapping of Economy is Contest, and added elaborations on the participants.

By carefully examining the mapping from source domain knowledge based on SUMO, we discovered that not only mapping is indeed based on a priori source domain knowledge. We also discovered that a metaphor can often involve additional and more specified terms within a domain. In these cases, no additional mapping is required. The same structured domain knowledge is used, and the subsumed terms offers only elaborations based on the same knowledge structure.

It is appropriate to note here that based on WordNet to SUMO mapping, economy is a SocialInteraction, and Contest is a subclass of SocialInteraction. In other words, economy is a related concept to Contest, although it does not belong to that conceptual domain. That a metaphor chooses a related domain is to be expected.

Economy is Organism

Among metaphors involving economies, one source domain stands out as being far removed conceptually. These are the metaphors involving Organism. We arrived at this conclusion by re-examining the examples that we generalized as Economy is a Person in the previous section. After closer examination with the help of SUMO knowledge representation, we found that the linguistic realizations of this mapping do not involve any knowledge that is specific to Human. In fact, it only involves the notion of a life cycle, which is the defining knowledge involving an Organism.

Organism is defined in SUMO as 'a living individual, including all Plants and Animals.' And the crucial knowledge encoded in of the attached inference rules follows:

=> (and (instance ?ORGANISM Organism) (agent ?PROCESS ?ORGANISM)) (holdsDuring (WhenFn ?PROCESS) (attribute ?ORGANISM Living)))

The above inference rule encodes the knowledge that 'An organism is the agent of a living process that holds over a duration.' In other words, having a life cycle is the defining knowledge of an Organism. This turns out to be the source domain knowledge that is involved in the mapping.

It is interesting to observe, though this is not encoded by SUMO, that from a Darwinian perspective, the Purpose of an Organism as an Agent is to prolong his own life cycle. We found that in actual linguistic data, when the above two metaphors are used simultaneously, it is only when improving the life cycle (Economy is Organism) is incorporated as the PurposeForAgent (Economy is Contest). In other words, the source domain knowledge is robust in conceptual metaphor and can be automatically mapped to and merged.

6 Conclusion

In this paper, we propose an ontology-based and corpus-driven approach towards predicting lexical meaning of conceptual metaphors. Our theory is thus formally constrained. We also verified our findings with examination of corpora data. In the final version of this paper, we will demonstrate how the process of establishing mapping principles and deriving metaphorical meaning can be semi-automaticized based on both the SUMO ontological databases and corpora data. Such a process has important implications both in cognitive explanation of conceptual metaphors and in the application of SUMO to predict figurative meaning in metaphorical uses.

Acknowledgments

This study is partially supported both by a NSC project 'Sense and Sense-Ability', as well as a NDAP project 'Linguistic Anchoring.' We would like to thank Adam Pease of Teknowledge, the ACL workshop reviewers, as well as colleagues of the two above projects, for their comments. Any remaining errors are our own.

References

Ahrens, K. 2002. When Love is not Digested: Underlying Reasons for Source to Target Domain Pairing in the Contemporary Theory of Metaphor. In YuChau E. Hsiao (ed.) *Proceeding of the First Cognitive Linguistics Conference*, pp 273-302. Taipei: Cheng-Chi University.

Farrar, S., Lewis, W., and Langendoen, T. 2002. A Common Ontology for Linguistic Concepts. In *Proceedings of the Knowledge Technologies Conference*, Seattle, Washington, March 10-13, 2002. (available at http://ontology.teknowledge.com/#pubs)

Gentner, D. and G. Wolff. 2000. "Evidence for Role-Neutral Initial Processing of Metaphors." *Journal of Experimental Psychology*, 26, 529-541.

Lakoff, G. 1993. "The Contemporary Theory of Metaphor." In Andrew Ortony (ed.) *Metaphor and Thought* (2nd ed.). Cambridge: Cambridge University Press. P 202-251.

Martin J., 1992. Metabank: a Knowledge Base of Metaphoric Language Conventions. *Computational Intelligence*, 10, pg. 134-149.

McGlone, M. S. 1996. "Conceptual Metaphors and Figurative Language Interpretation: Food for Thought?" *Journal of Memory and Language*, 35, 544-565.

Niles, I. 2003. Mapping WordNet to the SUMO Ontology. Teknowledge Technical Report.

Niles, I., & Pease, A. 2001. Toward a Standard Upper Ontology. Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001).

Pease, A. & Niles, I. 2002. IEEE Standard Upper Ontology: A Progress Report. *Knowledge Engineering Review, Special Issue on Ontology and Agents*, Volume 17.

Sevcenko, M. 2003. Online Presentation of an Upper Ontology. In *Proceedings of Znalosti 2003*, Ostrava, Czech Republic, February 19-21, 2003.

On-line Resources

Academia Sinica Balanced Corpus
http://www.sinica.edu.tw/SinicaCorpus/

English-Chinese Ontology/WordNet Interface (http://ckip.iis.sinica.edu.tw/CKIP/ontology/).

SUMO (Suggested Upper Merged Ontology) http://ontology.teknowledge.com)