Comparing Semantic Role Labeling with Typed Dependency Parsing in Computational Metaphor Identification

Eric P. S. Baumer
Department of Informatics
Univ of California, Irvine
5029 Donald Bren Hall
Irvine, CA 92627-3440 USA
ebaumer@ics.uci.edu

James P. White
School of Information and
Computer Sciences
Univ of California, Irvine
Irvine, CA 92627
jpwhite@uci.edu

Bill Tomlinson
Department of Informatics
Univ of California, Irvine
5068 Donald Bren Hall
Irvine, CA 92627-3440 USA
wmt@uci.edu

Abstract

Most computational approaches to metaphor have focused on discerning between metaphorical and literal text. Recent work on computational metaphor identification (CMI) instead seeks to identify overarching conceptual metaphors by mapping selectional preferences between source and target corpora. This paper explores using semantic role labeling (SRL) in CMI. Its goals are two-fold: first, to demonstrate that semantic roles can effectively be used to identify conceptual metaphors, and second, to compare SRL to the current use of typed dependency parsing in CMI. The results show that SRL can be used to identify potential metaphors and that it overcomes some of the limitations of using typed dependencies, but also that SRL introduces its own set of complications. The paper concludes by suggesting future directions, both for evaluating the use of SRL in CMI, and for fostering critical and creative thinking about metaphors.

1 Introduction

Metaphor, the partial framing of one concept in terms of another, pervades human language and thought (Lakoff and Johnson, 1980; Lakoff, 1993). A variety of computational approaches to metaphorical language have been developed, e.g., (Martin, 1990; Fass, 1991; Gedigian et al., 2006; Krishnakumaran and Zhu, 2007). However, most such methods see metaphor as an obstacle to be overcome in the task of discerning the actual, literal meaning of a phrase or sentence.

In contrast, the work presented here approaches conceptual metaphor not as an obstacle but as a resource. Metaphor is an integral part in human understanding of myriad abstract or complex concepts (Lakoff and Johnson, 1980), and metaphorical thinking can be a powerful component in critical and creative thinking, cf. (Gordon, 1974; Oxman-Michelli, 1991). However, "because they can be used so automatically and effortlessly, we find it hard to question [metaphors], if we can even notice them" (Lakoff and Turner, 1989, p. 65). Computational metaphor identification (CMI) (Baumer, 2009; Baumer et al., under review) addresses this difficulty by identifying potential conceptual metaphors in written text. Rather than attempting to discern whether individual phrases are metaphorical or literal, this technique instead identifies larger, overarching linguistic patterns. The goal of CMI is not to state definitively *the* metaphor present in a text, but rather to draw potential metaphors to readers' attention, thereby encouraging both critical examination of current metaphors and creative generation of alternative metaphors.

CMI identifies potential metaphors by mapping selectional preferences (Resnik, 1993) from a source corpus to a target corpus. Previous work on CMI utilized typed dependency parses (de Marneffe et al., 2006) to calculate these selectional preferences. This paper explores the use of semantic role labeling (SRL) (Gildea and Jurafsky, 2002; Johansson and Nugues, 2008) to calculate selectional preferences. Typed dependencies focus on syntactic structure and grammatical relations, while semantic roles emphasize conceptual and semantic structure, so SRL may

be more effective for identifying potential conceptual metaphors. This paper describes how SRL was incorporated into CMI and compares both the relational data and the metaphors identified with typed dependency parsing and semantic role labeling. The results show that semantic roles enabled effective identification of potential metaphors. However, neither typed dependencies nor semantic roles were necessarily superior. Rather, each provides certain advantages, both in terms of identifying potential metaphors, and in terms of promoting critical thinking and creativity.

2 Related Work

2.1 Computational Approaches to Metaphor

Many computational approaches have been taken toward identifying metaphor in written text. MIDAS (Martin, 1990) attempts to detect when users of the Unix Consultant command line help system use metaphors, for example, "How do I enter Emacs?" is interpreted as "How do I invoke Emacs?" Another system, met* (Fass, 1991), is designed to distinguish both metaphor and metonymy from literal text, providing special techniques for processing these instances of figurative language. More recently, Gedigian et al. (2006) used hand-annotated corpora to train an automatic metaphor classifier. Krishnakumaran and Zhu (2007) used violations of WordNet-based (Fellbaum, 1998) verb-noun expectations to identify the presence of a metaphor, e.g., "he is a brave lion," would be considered metaphorical, because "he," taken to mean a "person," which is not a WordNet hyponym of "lion."

These and similar approaches ascribe to some degree to the literal meaning hypothesis (Reddy, 1969), which states that every sentence has a literal meaning, as derived from the meanings of its constituent words, while some also have a figurative meaning that goes beyond the meanings of the words themselves. In this view, a figurative interpretation is only sought only after a literal interpretation has been formed and found inconsistent, nonsensical, or otherwise faulty. However, experimental evidence has made this account suspect (Gibbs, 1984; Gentner et al., 2001). Even distinguishing whether a given expression is literal or figurative can be difficult at best. For example, "the rock is becoming brittle with age" (Reddy, 1969, p. 242), has "a literal interpretation when uttered about a stone and a metaphorical one when said about a decrepit professor emeritus" (Fass, 1991, p. 54).

One previous metaphor system avoids making such literal/metaphorical distinctions. CorMet (Mason, 2004) is designed to extract known conventional metaphors from domain-specific textual corpora, which are derived from Google queries. CorMet calculates selectional preferences and associations (Resnik, 1993) for each corpus's characteristic verbs, i.e., those verbs at least twice as frequent in the corpus as in general English. Based on these selectional associations, CorMet clusters the nouns for which the characteristic verbs select. To identify metaphors, mappings are sought from clusters in the source corpus to clusters in the target corpus, based on the degree to which the same verbs select for members of both clusters. For example, CorMet was used to extract the metaphor MONEY IS A LIQUID[1] by mapping from a cluster for the concept *liquid* in a corpus for the domain LABORATORY to a cluster for the concept *money* in a corpus for the domain FINANCE, based on the selectional associations of verbs such as "pour," "flow," "freeze," and "evaporate." The CMI system described in this paper is informed largely by CorMet (Mason, 2004).

2.2 Semantic Role Labeling

While interpretations vary somewhat, semantic role labeling (SRL) generally aims to represent something about the meaning of a phrase at a deeper level than surface syntactic structure. One of the most common approaches to performing SRL automatically is to use a statistical classifier trained on labeled corpora (Gildea and Jurafsky, 2002), with FrameNet (Baker et al., 1998) and PropBank (Palmer et al., 2005) being the primary sources. An important result of the Gildea and Jurafsky work was identifying the significant utility of using pre-segmented constituents as input to their labeler, and accordingly most SRL systems perform a syntactic analysis as an initial step.

The principal alternative to using a statistical classifier is to use a rule-based labeler for operating on

---

[1]SMALL CAPS are metaphors, *italics* are concepts, CAPS are domains, and "quotes" are example phrases.

the syntactic parse tree. For example, Shi and Mihalcea (2004) extract explicit SRL rules by analyzing FrameNet cases. Another system, RelEx (Fundel et al., 2006) also uses rules and is structured like the implementation used here (see below for details), but despite having the same name, is a different system. Statistical and rule-based methods may also be used within the same system, such as in LTH (Johansson and Nugues, 2008).

One reason for preferring a rule-based SRL system is that rule-based approaches may be less susceptible to the loss of accuracy that statistically trained classifiers suffer when applied to domains that are different than the corpora they are trained on (Johansson and Nugues, 2008). That problem is compounded by the limited domain coverage provided by the labeled corpora currently available for SRL classifier training (Gildea and Jurafsky, 2002).

3 Computational Metaphor Identification

While space precludes a fully detailed description of the algorithms involved, this section provides a high-level summary of the techniques employed in CMI (Baumer, 2009; Baumer et al., under review).

Metaphors are conceptual mappings wherein a source concept partially structures the understanding of a target concept. In ELECTION IS WAR, the target concept *election* is partially framed in terms of the source concept *war*. CMI begins by gathering corpora for the source and target domains. In this paper, the target corpus consists of posts from political blogs, described in more detail in the methods section below. Source corpora are composed of Wikipedia articles, as they provide a readily available, categorically organized, large source of content on a wide variety of topics. A source corpus for a given domain consists of all the Wikipedia articles in the category for that domain, as well as all articles in its subcategories. All documents in the source and target corpora are parsed to extract sentence structure and typed dependencies (Klein and Manning, 2003; de Marneffe et al., 2006).

The crux of CMI is selectional preference learning (Resnik, 1993), which quantifies the tendency of particular words to appear with certain other classes of words in specific grammatical relationships. For example, words for the concept of food are often

the direct object of the verb "eat." Using the parsed documents, CMI calculates selectional preferences of the characteristic nouns in a corpus, where characteristic means that the noun is highly frequent in the corpus relative to its frequency in general English, as derived from (Kilgarriff, 1996). Selectional preference is quantified as the relative entropy of the posterior distribution conditioned on a specific noun and grammatical relation with respect to the prior distribution of verbs in general English:

$$S(c) = \sum_v P(v|c) \log \frac{P(v|c)}{P(v)} \qquad (1)$$

where c is a class of nouns (i.e., a concept like food) and a grammatical relation (such as direct object), and v ranges over all the verbs for which c appears in the given relation. These selectional preference strengths are then divided among the verbs that appear in each grammatical relation to determine the noun class's selectional association for each verb in each relation (Resnik, 1993).

Selectional associations are calculated for classes of words, but the corpora consist of words that may represent many possible classes of nouns. Thus, individual nouns count as partial observations of each word class that they might represent using WordNet (Fellbaum, 1998). For example, "vote," "primary," and "runoff" can all represent the concept of *election*. Here we use a customized version of WordNet that includes major political figures from the 2008 US Election. These word classes are then clustered using two-nearest-neighbor clustering based on the verbs for which they select. Each cluster represents a coherent concept in the corpus, and each is automatically labeled based on the synsets it contains.

This approach of using clustered hypernyms resonates with Lakoff's argument that metaphorical mappings occur not at the level of situational specifics, but at the superordinate level. For example, in the metaphor LOVE IS A JOURNEY, the relationship is a vehicle. Although specific instantiations of the metaphor may frame that vehicle variously as a train ("off the track"), a car ("long, bumpy road"), or a plane ("just taking off"), "the categories mapped will tend to be at the superordinate level rather than the basic level" (Lakoff, 1993, p. 212). This method of counting each word observed as a partial observation of each of the synsets

it might represent causes observations at the basic level to accumulate in the superordinate levels they collectively represent. This is not to say that hierarchical conceptual relations capture every possible metaphor, but rather that these are the relations on which we focus here.

To identify metaphors, CMI looks for correspondences between conceptual clusters in the source and target corpora. For example, in the Military corpus, the cluster for *war* would frequently select to be the direct object of "win," the object of the preposition "during" with the verb "fight," the object of the preposition "in" with the verb "defeated," and so on. In some blog corpora, the cluster for *election* also selects for those same verbs in the same grammatical relationships. Based on the similarity of these selectional associations, each mapping is given a confidence score to indicate how likely the linguistic patterns are to evidence a conceptual metaphor. One of the strengths of CMI is that it works in the aggregate. While individual instances of phrases such as "fought during the election" and "defeated in the primary" may not at first glance appear metaphorical, it is the systematicity of these patterns that becomes compelling evidence for the existence of a metaphor.

An important aspect of CMI is that it identifies only linguistic patterns potentially indicative of conceptual metaphors, not the metaphors themselves. As mentioned above, Lakoff (1993) emphasizes that metaphor is primarily a cognitive phenomenon, and that metaphorical language serves as evidence for the cognitive phenomenon. CMI leverages computational power to search through large bodies of text to identify patterns of potential interest, then presents those patterns to a human user along with the potential metaphors they might imply to foster critical thinking about metaphor. To reiterate, this places the job of finding patterns in the hands of the computer, and the job of interpreting those patterns in the hands of the human user.

4 CMI with Semantic Role Labeling

The work presented in this paper attempts to enhance CMI by using SRL to expand the types of relations between nouns and verbs that can be seen as instantiating a metaphor. The prior CMI implementation treats each grammatical dependency type

as a distinct relation. For example, in the sentence, "The city contained a sacred grove for performing religious rites," "rites" is the direct object of "perform," as denoted by the *dobj* dependency. However, the sentence, "The religious rites were once again performed openly," uses a passive construction, meaning that "rites" is the passive subject, or *nsubjpass*, of "perform." With SRL, the relations between "perform" and "rite" are the same for both sentences; specifically, *Intentionally_act:Act* ("rite" is the intentional act being performed) and *Transitive_action:Patient* ("rite" is the recipient of a transitive action). Because the relations in FrameNet are organized into an inheritance structure, both the more general frame *Transitive_action* and the more specialized frame *Intentionally_act* apply here.

This section describes how SRL was incorporated into CMI, compares the component data derived from SRL with the data derived from a typed dependency parse, and compares resulting identified metaphors.

4.1 Implementation Methods

The CMI system used here takes the prior implementation (described in section 3) and replaces the Stanford typed dependency parser (de Marneffe et al., 2006) with the RelEx SRL system (http://opencog.org/wiki/RelEx). RelEx performs a full syntactic parse, then applies a set of syntactic pattern rules to annotate the parse tree with role labels based (not exactly or completely) on FrameNet. This implementation uses a rule-based labeler because CMI hinges on differences in selectional preferences in corpora from different domains, and statistically trained classifiers are biased by the distributions of the corpora on which they are trained.

For syntactic parsing, RelEx uses the Link Grammar Parser (LGP) which is based on the Link Grammar model (Sleator and Temperley, 1993). LGP produces output very similar to typed dependencies. The version of RelEx we use integrates the Another Nearly-New Information Extraction (ANNIE) system (http://gate.ac.uk/sale/tao/splitch6.html#chap:annie) to tag named entities. Sentences are split using the OpenNLP sentence splitter (http://opennlp.sourceforge.net/).

Because CMI's corpora are acquired from public

	Blogs	Religion (Wikipedia)
Docs	546 (604)	3289 (3294)
Sents	5732 (6708)	128,543 (145,193)
Words	148,619	3,300,455

Table 1: Sizes of the target and source corpora; parentheses show totals including documents without valid sentences and sentences with no relations.

Stanford	Blogs	Religion
Reln(v, n)	19,303 (2.88)	425,367 (2.93)
Unique(v, n)	19,303 (2.88)	425,367 (2.93)
RelEx	Blogs	Religion
Reln(v, n)	57,639 (8.59)	1,219,345 (8.40)
Unique(v, n)	20,962 (3.12)	482,997 (3.33)

Table 2: Relations for the target and source corpora; parentheses show average relations per word.

Stanford		RelEx	
Relation	Freq	Relation	Freq
dobj	3815	Transitive_action:Patient	4268
nsubj	3739	Transitive_action:Agent	3597
prep_in	1072	Inheritance:Item_2	1489
prep_to	695	Categorization:Category	1489
prep_on	563	Attributes:Attribute	1488
nsubjpass	528	Existence:Entity	1277
prep_for	491	Categorization:Item	1270
prep_with	435	Inheritance:Item_1	1269
prep_at	285	Attributes:Entity	1268
dep	279	Purpose:Means	569

Table 3: Most common dependencies and frequencies.

Internet sources, the text must be cleaned to make it suitable for parsing. Text from Wikipedia articles undergoes many small filtering steps in order to remove wiki markup, omit article sections that do not consist primarily of prose (e.g., "See Also" and "References"), and decompose Unicode letters and punctuation into compatibility form. Wikipedia articles also tend to use bulleted lists in the middle of sentences rather than comma-separated clauses. We attempt to convert those constructions back into sentences, which only sometimes results in a reasonable sentence. However, it helps to ensure that the following sentence is properly recognized by the sentence splitter. For blog posts, HTML tags were removed, which at times required multiple decoding passes due to improperly configured blog feeds, and characters decomposed into compatible form.

4.2 Data

Table 1 shows statistics on the sizes of the source and target corpora. Numbers in parentheses are totals, including blank documents and sentences with no valid relations. There are some sentences for which RelEx does not produce any parse, e.g., long sentences that LGP deems ungrammatical. The Stanford parser produced some result for every sentence, because it will produce a result tree for any kind of text, even if it does not recognize any grammatically valid tokens.

Table 2 lists the number of verb-noun relations for each corpus, with parentheses showing average relations per word. Since RelEx often labels the same verb-noun relation with multiple hierarchically-related frames (as described above), Table 2 also lists the number of unique verb-noun pairs labeled. For the blogs corpus, the Stanford parser generated 111 distinct dependency types, while RelEx labeled 1446 distinct roles. The ten

most common of each are listed with their frequencies in Table 3.

These data show that RelEx provides more information, both in terms of successfully parsing more sentences, and in terms of relations-per-word. The next section explores the impact of these differences on identified metaphors.

4.3 Results

This section describes metaphors identified when mapping from the RELIGION source corpus to the political blogs target corpus. CMI results are usually culled to include only the upper one percentile in terms of confidence, but space constraints prohibit a full analysis of even this upper one percentile. Instead, this section compares mappings with the highest confidence score from the typed dependency data and from the semantic role data. RELIGION was chosen as the source domain because the highest confidence metaphors from both typed dependencies and semantic roles had similar target and source concepts, facilitating a better comparison. This analysis

Target (label and cluster)	Source (label and cluster)	Conf
medicine - {medicine, medical specialty}, {medicine, medication, medicament, medicinal drug}, {music, medicine}, {medicine, practice of medicine}, {learned profession}, {drug}, {social control}, {profession}, {punishment, penalty, penalization, penalisation}, {life science, bioscience}	**sacrament** - {sacrament}, {baptism}, {religious ceremony, religious ritual}	1.968
	ritual - {ceremony}, {practice, pattern}, {custom, usage, usance}, {ritual, rite}, {survival}	1.465

Table 4: Metaphors for *medicine* from RELIGION using typed dependencies.

is not intended to demonstrate that either technique is superior (for more on possible evaluation methods, see Discussion section below). Rather, it provides a detailed depiction of both to ascertain potential benefits and drawbacks of each.

Table 4 presents the strongest two mappings from RELIGION: MEDICINE IS A SACRAMENT and MEDICINE IS A RITUAL; these were the only mappings for *medicine* in the upper one percentile. Each mapping lists both the automatically identified labels and the full cluster contents for source and target, along with the confidence score. The table can be read left-to-right, e.g., "medicine is like a sacrament." Confidence scores typically fall in the range (0, 5) with a few high-confidence mappings and many low-confidence mappings; see (Baumer, 2009; Baumer et al., under review) for details of confidence score calculation. Table 5 shows details for each mapping, including the verb-relation pairs that mediate the mapping, along with an example fragment from the target and source corpora for each verb-relation. These examples show why and how medicine might be like, variously, a sacrament or a ritual; both are "practiced," "administered," "performed," etc. Note that the passive subject and direct object relations are treated as distinct, e.g., "Eucharist is variously administered" involves a different grammatical relation than "administer the sacrament," even though the word for *sacrament* plays a similar semantic role in both fragments.

Tables 6 and 7 show mappings resulting from semantic roles labeled by RelEx, with formats similar to those of tables 4 and 5, except that the verb-relations in table 7 are semantic roles rather than grammatical relations. The mapping in table 6 was the strongest mapping from RELIGION and the only mapping for *medication*.

Table 7 shows how RelEx can treat different grammatical relations as the same semantic role. For example, "medicine is practiced" and "practice the rites" use passive subjective and direct object, respectively, but are both treated as the patient of a transitive action. Such examples confirm that SRL is, at least to some extent, performing the job for which it was intended.

However, these results also expose some problems with SRL, or at least with RelEx's implementation thereof. For example, the phrase "dispose of prescription drugs" is labeled with four separate semantic roles, which is an instance of a single verb-noun relation being labeled with both a superordinate relation, *Physical_entity:Entity*, and a subordinate relation, *Physical_entity:Constituents* (the constituents of a physical entity are themselves an entity). While various approaches might avoid multiple labels, e.g., using only the most general or most specific frame, those are beyond the scope here.

5 Discussion

As mentioned above, these results do not provide conclusive evidence that either typed dependencies or semantic roles are more effective for identifying potential metaphors. However, they do provide an understanding of both techniques' strengths and weaknesses for this purpose, and they also suggest ways in which each may be more or less effective at fostering critical and creative thinking.

For metaphor identification, the previous section described how typed dependency parsing treats passive subjects and direct object as distinct relations, whereas SRL will at times conflate them into identical patient roles. This means that the typed dependency-based metaphors appear to be mediated by a greater number of relations. However, it also

Target	Source	Verb-Reln	Target Ex Frag	Source Ex Frag
medicine	sacrament	practice - nsubjpass	"**medicine** is **practiced**"	"**rites** were **practiced**"
		administer - nsubjpass	"**antibiotics** are regularly **administered**"	"**Eucharist** is variously **administered**"
		administer - dobj	"**administered medicines**"	"**administer** the **sacrament**"
		perform - dobj	"**perform** defensive **medicine**"	"**performed** the last **rites**"
		receive - dobj	"**received** conventional **medicines**"	"**received** the **rites**"
	ritual	perform - dobj	"**perform** defensive **medicine**"	"**performed** the last **rites**"
		practice - nsubjpass	"**medicine** is **practiced**"	"**ceremonies** are also **practiced**"
		administer - dobj	"**administered medicines**"	"**administering** the **rites**"
		administer - nsubjpass	"**antibiotics** are regularly **administered**"	"**sacrament** is ordinarily **administered**"

Table 5: Details of RELIGION metaphors from typed dependencies, including mediators and example phrases.

Target (label and cluster)	Source (label and cluster)	Conf
medication - {medicine, medication, medicament, medicinal drug}, {drug}, {agent}	**ceremony** - {ceremony}, {sacrament}, {rite, religious rite}, {religious ceremony, religious ritual}	2.570

Table 6: Metaphor for *medication* from RELIGION using semantic roles.

Target	Source	Verb-Reln	Target Ex Frag	Source Ex Frag
medication	ceremony	practice - Transitive_action:Patient	"**medicine** is **practiced**"	"**practice** the **rites**"
		perform - Transitive_action:Patient	"**perform** defensive **medicine**"	"**perform** most religious **rites**"
		include - Transitive_action:Agent	"**medicine including**"	"**liturgies included**"
		dispose - Physical_entity:Constituents	"**dispose** of prescription **drugs**"	"**disposed** of without **ceremony**"
		dispose - Inheritance:Instance	"**dispose** of prescription **drugs**"	"**disposed** of without **ceremony**"
		dispose - Inheritance:Group	"**dispose** of prescription **drugs**"	"**disposed** of without **ceremony**"
		dispose - Physical_entity:Entity	"**dispose** of prescription **drugs**"	"**disposed** of without **ceremony**"

Table 7: Details of RELIGION metaphors from semantic roles, including mediators and example phrases.

means that less data are available to the selection preference calculation, in that there are fewer observations for each relation. On the other hand, SRL is a much finer-grained classification than typed dependencies. The implementation used here included 111 grammatical relations, whereas RelEx labeled 1446 distinct roles. Thus, overall, RelEx may be providing fewer observations for each relation, but those relations may have more semantic import.

For fostering critical thinking and creativity, a key concern is making identified metaphors readily comprehensible. Ortony (Ortony, 1980) and others have suggested that selectional restriction violations are an important component of metaphor comprehension. Therefore, tools that employ CMI often present parallel source and target fragments side-by-side to make clear the selectional restriction violation, e.g., metaViz, a system for presenting computationally identified metaphors in political blogs (Baumer et al., 2010). One might assume that typed dependencies are more readily comprehensible, since they are expressed as relatively simple grammatical relations. However, when presenting example fragments to users, there is no need to explicate the nature of the relationship being demonstrated, but rather the parallel examples can simply be placed side-by-side. It is an empirical question whether users would see phrases such as "medicine is practiced" and "practice the rites" as parallel examples of the same psycholinguistic relationship. Thus, the question of whether typed dependencies or semantic roles better facilitate metaphor comprehension may not be as important as the question of whether example phrases are perceived as parallel.

6   Future Work

This paper is only an initial exploration, demonstrating that semantic role labeling is viable for use in CMI. For the sake of comparison, the analysis here focuses on examples where metaphors identified using the two techniques were relatively similar. However, such similarity does not always occur. For example, using MILITARY as the source domain, typed dependencies led to results such as A NOMINEE IS A FORCE and A NOMINEE IS AN ARMY, whereas semantic roles gave mappings including AN INDIVIDUAL IS A WEAPON (here, the label "indi-

vidual" is a superordinate category including mostly politicians), and THE US IS A SOLDIER. Future work should analyze these differences in more detail to provide a broad and deep comparison across multiple source domains and target corpora.

But how should such an analysis be conducted? That is, how does one determine which identified metaphors are "better," and by what standard? In suggesting a number of potential evaluation methods for CMI, Baumer et al. (under review) argue that the most sensible approach is asking human subjects to assess metaphors, potentially along a variety of criteria. For example: Does the metaphor make sense? Is it unexpected? Is it confusing? Such assessments could help evaluate semantic roles vs. typed dependencies in two ways. First, does either parsing technique lead to metaphors that are consistently assessed by subjects as better? Second, does either parsing technique lead to better alignment (i.e., stronger correlations) between human assessments and CMI confidence scores? Such subjective assessments could provide evidence for an argument that either typed dependencies or semantic roles are more effective at identifying conceptual metaphors.

7   Conclusion

This paper explores using semantic role labeling (SRL) as a technique for improving computational metaphor identification (CMI). The results show that SRL can be successfully incorporated into CMI. Furthermore, they suggest that SRL may be more effective at identifying relationships with semantic import than typed dependency parsing, but that SRL may also make distinctions that are too fine-grained to serve as effective input for the selectional preference learning involved in CMI. The results also demonstrate that, even though the notion of semantic roles may seem more complex than typed dependencies from a user's perspective, it is possible to present either in a way that may be readily comprehensible. Thus, while more work is necessary to compare these two parsing techniques more fully, semantic role labeling may present an effective means of improving CMI, both in terms of the technical process of identifying conceptual metaphors, and in terms of the broader goal of fostering critical thinking and creativity.

Acknowledgments

References

Colin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *Proc 17th Int'l Conf on Comp Ling*, pages 86–90, Montral, Quebec, Canada.

Eric P. S. Baumer, Jordan Sinclair, and Bill Tomlinson. 2010. "America is like Metamucil:" Critical and creative thinking about metaphor in political blogs. In *ACM SIGCHI Conf*, Atlanta, GA. ACM Press.

Eric P. S. Baumer, David Hubin, and Bill Tomlinson. under review. Computational metaphor identification.

Eric Baumer. 2009. *Computational Metaphor Identification to Foster Critical Thinking and Creativity*. Dissertation, University of California, Irvine, Department of Informatics.

Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *Lang Res and Eval (LREC)*, Genoa, Italy.

Dan Fass. 1991. Met*: A method for discriminating metonymy and metaphor by computer. *Comp Ling*, 17(1):49–90.

Christine Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Katrin Fundel, Robert Kuffner, and Ralf Zimmer. 2006. RelEx-Relation extraction using dependency parse trees. *Bioinformatics*, 23(3):365–371.

Matt Gedigian, John Bryant, Srini Narayanan, and Branimir Ciric. 2006. Catching metaphors. In *3rd Workshop on Scalable Natural Language Understanding*, New York City. Assoc Comp Ling.

Dedre Gentner, Brian F. Bowdle, Phillip Wolff, and C. Boronat. 2001. Metaphor is like analogy. In Dedre Gentner, Keith J. Holyoak, and Boicho Kokinov, editors, *The Analogical Mind*, pages 199–253. MIT Press, Cambridge, MA.

Raymond W. Gibbs. 1984. Literal meaning and psychological theory. *Cognitive Science*, 8:275–304.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Comp Ling*, 28(3):245–288.

W.J.J. Gordon. 1974. Some source material in discovery-by-analogy. *Journal of Creative Behavior*, 8:239–257.

Richard Johansson and Pierre Nugues. 2008. Dependency-based semantic role labeling of PropBank. In *Proc Conf on Empirical Meth in Nat Lang Proc*, pages 69–78, Honolulu, HI. Assoc Comp Ling.

Adam Kilgarriff. 1996. BNC word frequency list. http://www.kilgarriff.co.uk/bnc-readme.html.

Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Mtg of the Assoc for Comp Ling*, Sapporo, Japan.

Saisuresh Krishnakumaran and Xiaojin Zhu. 2007. Hunting elusive metaphors using lexical resources. In Xiaofei Lu and Anna Feldman, editors, *Computational Approaches to Figurative Language, Workshop at HLT/NAACL 2007*, Rochester, NY.

George Lakoff and Mark Johnson. 1980. *Metaphors We Live By*. University of Chicago Press, Chicago, IL, 2003 edition.

George Lakoff and Mark Turner. 1989. *More Than Cool Reason: A Field Guide to Poetic Metaphor*. University of Chicago Press, Chicago and London.

George Lakoff. 1993. The contemporary theory of metaphor. In A. Ortony, editor, *Metaphor and thought, 2nd. ed.*, pages 202–251. Cambridge Univ Press, New York.

James H. Martin. 1990. *A Computational Model of Metaphor Interpretation*. Acad Press, San Diego, CA.

Zachary J. Mason. 2004. CorMet: a computational, corpus-based conventional metaphor extraction system. *Comp Ling*, 30(1):23–44, March.

Andrew Ortony. 1980. Some psycholinguistic aspects of metaphor. In R.P. Honeck and H.R. Robert, editors, *Cog and Fig Lang*, pages 69–83. Erlbaum Associates, Hillsdale, NJ.

Wendy Oxman-Michelli. 1991. Critical thinking as creativity. Technical Report SO 023 597, Montclair State, Institute for Critical Thinking, Montclair, NJ.

Martha Palmer, Paul Kingsbury, and Daniel Gildea. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Comp Ling*, 31(1):71–106, March.

Michael J. Reddy. 1969. A semantic approach to metaphor. In *Chicago Linguistic Society Collected Papers*, pages 240–251. Chicago Univ Press, Chicago.

Philip Resnik. 1993. *Selection and Information: A Class-Based Approach to Lexical Relationships*. Dissertation, University of Pennsylvania, Department of Computer and Information Science.

Lei Shi and Rada Mihalcea. 2004. Open text semantic parsing using FrameNet and WordNet. In *Demonstration Papers at HLT-NAACL 2004*, pages 19–22, Boston. Assoc for Computational Linguistics.

Daniel Sleator and Davy Temperley. 1993. Parsing English with a Link Grammar. In *Proc Third International Workshop on Parsing Technologies*, pages 277–292.