

A Corpus of Rich Metaphor Annotation

Jonathan Gordon¹, Jerry R. Hobbs¹, Jonathan May¹, Michael Mohler²,
Fabrizio Morbini³, Bryan Rink², Marc Tomlinson², and Suzanne Wertheim⁴

¹ USC Information Sciences Institute, {jgordon, hobbs, jonmay}@isi.edu

² Language Computer Corp., {michael, bryan, marc}@languagecomputer.com

³ USC Institute for Creative Technologies, morbini@ict.usc.edu

⁴ Worthwhile Research & Consulting, worthwhileresearch@gmail.com

Abstract

Metaphor is a central phenomenon of language, and thus a central problem for natural language understanding. Previous work on the analysis of metaphors has identified which *target* concepts are being thought of and described in terms of which *source* concepts, but this is not adequate to explain what motivates the use of particular metaphors. This work proposes the use of *conceptual schemas* to represent the underspecified scenarios that motivate a metaphoric mapping. To support the creation of systems that can understand metaphors in this way, we have created and are publicly releasing a corpus of manually validated metaphor annotations.

1 Introduction

Lakoff and Johnson (1980) wrote that “the essence of metaphor is understanding and experiencing one kind of thing in terms of another.” Our mental, *conceptual metaphors* (CMs) relate the immediate subject matter of a *target* conceptual domain such as *Argument* to a (usually more concrete) *source* domain such as *War*. These conceptual links are exhibited in speech or text as *linguistic metaphors* (LMs), such as “Your claims are indefensible”. Metaphors include both fixed-form expressions, such as “won the argument”, and novel expressions, such as “threw a fragmentation grenade at her premise”. Natural language systems that are not specially equipped to interpret metaphors will behave poorly on metaphoric text. For instance, a document classifier should distinguish between text about a war and text using war as a metaphor to talk about poverty.

The metaphors people use can also provide insights into their attitudes and preconceptions. Examining metaphors allows us to empirically contrast individual speakers, groups, or entire cultures. Even when an expression is common and might not be consciously metaphoric for the speaker, it can fit a metaphoric structuring of the concept being discussed.

Metaphors are not produced at random. Rather, different metaphoric structurings highlight—and hide—different aspects of the target concept. Furthermore, based on theoretical frameworks developed as part of our research, there seems to be a small set of scenarios, central to human existence, that are called upon metaphorically to make sense of more complex or abstract concepts. These include, e.g., threats to a person’s health and safety and mitigators of such such threats.

It is in light of this analysis that we approach the annotation of metaphors for interpretation, explanation, and comparison. While previous work, e.g., Shaikh et al. (2014), has produced collections of linguistic metaphors annotated with source and target concepts, such annotations do not capture the scenarios that inform a metaphoric mapping. For instance, the sentences

Democracy has crushed our dreams.

Extremists have crushed our democracy.

are both about the source–target pair $\langle \textit{Physical Harm}, \textit{Democracy} \rangle$, but with contrasting views: In the first, democracy is seen as a threat, while in the second it is being threatened. In this paper, we present an annotation scheme that draws such distinctions, and

we use it to create a corpus of metaphor annotations, which we are releasing to support the creation of tools for the deeper automated analysis of metaphors.

2 Related Work

All corpora contain metaphors, however even those general corpora that give semantic interpretations, e.g., the AMR corpus (Knight et al., 2014), do not address—or do not consistently address—metaphors. Computational research on metaphor has focused on the problems of (1) identifying linguistic metaphors in text (e.g., Fass, 1991; Birke and Sarkar, 2006; Steen et al., 2010; Shutova et al., 2010; Li and Sporleder, 2010) and then (2) identifying the source and target concepts invoked by each linguistic metaphor (e.g., Narayanan, 1997; Barnden and Lee, 2002; Agerri et al., 2007; Shutova, 2010; Ovchinnikova et al., 2014; Gordon et al., 2015).

The corpora of (variably annotated) metaphors that have been released to date are not sufficient to tell the story of why a metaphor was invoked or to allow the meaningful comparison of metaphors used by different individuals or even entire cultures. MetaBank (Martin, 1994) provided sets of cross-domain mappings, organized by a small set of important abstract metaphors, adapted from the Berkeley Metaphor List. The latter was expanded into the Master Metaphor List (Lakoff et al., 1991), which gives a hierarchically organized set of conceptual metaphors (i.e., source–target mappings) and supporting examples of linguistic metaphors. The Italian Metaphor Database (Alonge, 2006), the work of Shutova et al. (2013) annotating the British National Corpus (BNC Consortium, 2001), and the large-scale, multilingual work of Shaikh et al. (2014) and Mohler et al. (2014) all focus on the identification of source and target concepts.

Unlike other work manually or automatically annotating metaphors in text, the focus of this paper is not on whether text is metaphorical or not, or which concept domains the metaphors involve, but on the meaning (or “story”) of each metaphor. For instance, from the text

...changes over the last couple centuries
have increased how much democracy in-
fests our [political] system ...

we can identify the verb “infests” as an instance of the source concept *Parasite* and “democracy” as the

target *Democracy*. What is absent from this annotation is why we consider this to be so, and what underlying story gives the metaphorical expression heft, e.g.,

Democracy is seen as a parasite because it poses a threat to the health of the political system that it “infests”.

In the following section we describe the annotation we use to produce a corpus that meets this standard.

3 Metaphor Representation

We desire a rich structured markup to represent the meaning of metaphors in terms of the speaker’s motivation. For this, we introduce a set of ontological categories with associated schema representations.

This builds on previous work mapping linguistic metaphors to their conceptual source and target domains. The *source domain* of the metaphor is the loose set of related concepts used to metaphorically describe the *target domain*. There are no theoretical or practical limitations on the set of target domains that can be described by metaphors. The set of possible source domains is also unbounded, but people commonly draw upon a small set of familiar scenarios in order to make sense of more abstract or complex experiences.

For this work, we recognize 70 source domains. This list is the result of a three-year bottom-up process, where new metaphors were observed, clustered, and assigned a label. Source domains were split or consolidated to better fit the data. The list is necessarily arbitrary, relying on human judgment of when two source domains are distinct.

While 70 source domains abstract over individual linguistic metaphors, 14 *ontological categories* abstract over the source domains. An ontological category is a collection of one or more scenarios that are conceptually related. The choice of ontological categories was based on extensive data analysis, but, as with the source domains, ultimately relies on human judgment. The category of Health and Safety, for instance, includes metaphors from the source domains *Food*, *Medicine*, *Physical Harm*, and *Protection*, among others. The ontological categories are:

1. Health and Safety
2. Journey

3. Conflict
4. Power and Control
5. Engineering and Business
6. Morality and Justice
7. Systematic Explanations
8. Plants
9. Animals
10. Human Life Cycle and Relations
11. Darkness and Light
12. High and Low
13. Nature
14. Embodied Experience

A scenario in an ontological category is a coherent set of roles. Each scenario can be represented by a *conceptual schema*, the set of properties or components that were deemed essential to represent the roles that elements can play in metaphors about the scenario. Each schema was designed based on the analysis of a variety of metaphors for the scenario. Most categories are sufficiently coherent that they can be described by a single overarching *über schema*, while a few, such as Nature, consist of diverse scenarios and thus require multiple schemas. A scenario is added to a category when it is common and is conceptually distinct from the other scenarios in the category, reflected by a low overlap in schema elements.

Each schema is simplified as much as possible while retaining the ability to capture the basic meaning of the scenario. Additional meaning of the metaphor comes from the specific source and target domains involved, and from the particulars of the sentence and the discourse context. The schema analysis of metaphors cannot capture the full meaning of each linguistic metaphor, but it is a step toward a (notional) complete analysis.

We represent a schema as an unordered set of labeled slots, whose values can be

- null (i.e., the slot is not instantiated);
- a text excerpt from the linguistic metaphor (not altered or rearranged); or
- one of a closed-class set of values defined for that slot.

Then, for each linguistic metaphor in a text, a basic explanation of the metaphor is an instantiation of the schema for an ontological source category. A

successful annotation should contain enough information, including the choice of the schema, to allow a natural language generator (human or otherwise) to construct an explanatory sentence.

For a selection of the categories, we now provide the corresponding annotation guidelines.¹ These include a detailed description of the scenarios and the elements that define their schemas. The scenario descriptions also serve to explain the schema slots to end-users, e.g., for comparing metaphors used by different groups.

Each schema slot is accompanied by an illustrative list of *lexical triggers*, which are text constructions that may identify the slot's value. For closed-class slots, the legal values are given in uppercase. Each schema is followed by example metaphor annotations, primarily drawn from the US gun control debate. These analyses include the identification of source and target conceptual domains.

Category: Health and Safety

People want to stay safe and healthy. Some things in the world are *threats* to the health and safety of the *threatened*. Other things are *protection* against these threats or are *beneficial*.

- **Threat**, e.g., monsters, tsunamis, diseases, parasites, “overdose of *x*”, “evil *x*”.
- **Threatened**, e.g., “sick *x*”, “*x* overdoses”, “*x* is threatened”, “*x* is infested”, “*x* is contaminated”.
- **Protection** or mitigation of threats, e.g., medicine, protection, shelter, “*x* alleviates”.
- **Beneficial** or necessary things, e.g., “*x* is the beating heart of *y*”, doctors, “appetite for *x*”.

Examples:

“This is how irrationally fearful are [sic] of guns some people are. Seems any *exposure to firearms* is a horrific tragedy. Even teaching gun safety is a travesty.”

- **Source:** *Disease*
- **Target:** *Guns*
- **Threat:** “firearms”
- **Threatened:** “some people”

¹ The full set of annotation guidelines is included with the released corpus.

“Back in the 1760’s there was a far greater amount of threat that a *gun* could potentially *alleviate* than there is for a person at a local starbucks [sic].”

- **Source:** *Medicine/Physical Burden*
- **Target:** *Guns*
- **Protection:** “gun”

Category: Journey

An *agent* on a journey wants to reach their *goal*. Some things—*vehicles*—facilitate movement towards a destination. Other things—*barriers*—hinder movement towards a destination. Any movement forward or back causes *change* to increase or decrease (*change type*).

- **Agent:** the person on a journey, e.g., “*x* flies”, “*x* travels”, “*x* marches”, “journey of *x*”, “progression of *x*”.
- **Goal:** the destination of the journey, e.g., destination, escape, summit, “advance toward *x*”, “road to *x*”, “steps toward *x*”, “door to *x*”.
- **Vehicle:** the facilitator of the journey, e.g., “straight pathway of *x*”, “engine of *x*”, “*x* provides access to”.
- **Barrier:** a thing that impedes the journey, e.g., “maze of *x*”, road block, “obstructive *x*”, obstacle, “*x* restrains”, “*x* ensnares”, labyrinthine.
- **Change,** e.g., “*x* advances”, “*x* progresses”, “retreat of *x*”, “*x* go backwards”, “*x* reversed course”.
- **Change Type:**
 - INCREASE, e.g., advances, increases, progresses.
 - DECREASE, e.g., retreats, decreases, diminishes.

Examples:

“... we need to accept the historical account of the second amendment for what it actually is, and move forward *in advancing gun rights* with a full understanding of the historical truth. Only then will gun rights be on a solid foundation.”

- **Source:** *Forward Movement*
- **Target:** *Gun Rights*
- **Agent:** “we”

- **Goal:** “gun rights be on a solid foundation”
- **Vehicle:** “accept the historical account of the second amendment for what is actually is”
- **Change:** “gun rights”
- **Change Type:** INCREASE

“The *retreat* of *gun control* is over.”

- **Source:** *Backward Movement*
- **Target:** *Control of Guns*
- **Change:** “gun control”
- **Change Type:** DECREASE

Category: Conflict

There are two opposing *sides* in a conflict, one of which may be the *enemy* of the speaker. The conflict has structure and plays out according to the rules of engagement. A component of the conflict may be an *aid*, which helps progress toward the goal of winning. At the end, one side is a *winner* and the other side is a *loser*. If your side wins, it is success; if it loses, it is failure. The conflict may have different stakes, e.g., losing a war is more serious than losing a football game.

- **Conflict,** e.g., “game of *x*”, “battle of *x*”, “*x* competition”, “*x* debate”, “fight in *x*”, “the *x* war”, “inning of *x*”, “struggle of *x*”.
- **Side,** e.g., “*x* team”, “*x* forces”, “compete with *x*”, “challenge *x*”, “winning against *x*”, “rival to *x*”, “*x* combats *y*”, “*x* scored”, “*x* is battling”.
- **Enemy** or competitor of the speaker, e.g., “*x* are terrorists”, “*x* are evildoers”, “opposing *x*”, “fighting *x*”, “*x* is our enemy”.
- **Winner,** e.g., “*x* wins”, “*x* victory”, “victorious *x*”, “*x* conquered”, “victory of *x* over...”
- **Loser,** e.g., “*x* loses”, “defeated *x*”, “surrender of *x*”, “defeat of *x*”, “*x* capitulates”.
- **Aid,** a component of the conflict that helps toward winning, e.g., a home run, “sword of *x*”, “brandish *x*”, “wield *x*”, “*x* is a useful weapon”.

Examples:

“Whether George W. Bush or Al Gore ends up **winning** the *presidency*, the Constitution charts a course for him ...”

- **Source:** *Competition*
- **Target:** *Government*

- **Conflict:** presidency
- **Side:** “George W. Bush”
- **Side:** “Al Gore”

“We agree that *gun control* will *win* because Americans aren’t willing to kill over it.”

- **Source:** *Competition / Game*
- **Target:** *Control of Guns*
- **Side:** “Americans”
- **Winner:** “gun control”

Category: Power and Control

A being may have power and control over another being. Levels of control and the autonomy of the subservient person vary. There is resentment and anger when autonomy levels are perceived as too low.

There are two distinct—but related—scenarios in Power and Control, which are annotated differently:

Scenario: Power and Control: God

A *god* is a sacred being that a *worshipper* considers to rightly have power over humans due to its innate superiority, including its holiness, wisdom, or might. The *legitimacy* of a non-divine being that has been elevated and worshipped as a god is false; otherwise, it is true.

- **God**, e.g., “*x* cult”, “*idolize x*”, “*sacred x*”, “*worship of x*”, “*temple of x*”, “*divine x*”, “*x* idol”, “*holy x*”.
- **Worshipper**, e.g., “*x* idolizes”, “*x* praises”, “*x* worships”.
- **Legitimacy:**
 - TRUE, e.g., divine, sacred.
 - FALSE, e.g., cult, false god, idol.

Examples:

“On the other hand when *guns* become *idols* we can document how their presence transforms the personalities of individuals and entire communities.”

- **Source:** *A God*
- **Target:** *Guns*
- **God:** “guns”
- **Legit:** FALSE

“Thus, independence of the *Judiciary* is *enshrined* in the Constitution for the first time, which is rightly considered a historic landmark.”

- **Source:** *A God*
- **Target:** *Government*
- **God:** “independence of the Judiciary”
- **Legit:** TRUE

Scenario: Power and Control: Human

Sometimes there is a clearly marked hierarchy among people, where a *servant* serves the will of a *leader*. The *degree* of oppression or submission may be low, in which case the servant is more thoroughly controlled, like a slave. Higher degrees of oppression are generally seen more negatively.

- **Leader:** who or what has power, e.g., “*x* ordered”, “*assisted x*”, “*served x*”, “*x* enslaves”, “*x* oppression”, “*x* reigns”, “*x* is king”.
- **Servant:** who or what is assisting or being controlled, e.g., “*x* assisted”, “*x* served”, “*enslaves x*”, “*x* obeys”, “*servile x*”, “*x* works for”.
- **Degree:**
 - HIGH: like a slave, e.g., slave, slave driver, dominance.
 - LOW: like a servant, e.g., served, assisted, helped.

Examples:

“Instead we watch *gun control* command the media filling every mind in the world with its hatred and fear of guns.”

- **Source:** *Leader*
- **Target:** *Control of Guns*
- **Leader:** “gun control”
- **Servant:** “the media”
- **Degree:** HIGH

“Guns prevent crime, *guns* assist crime, guns cause accidental deaths, guns turn minor disagreement into a [sic] deadly encounters.”

- **Source:** *Servant*
- **Target:** *Guns*
- **Leader:** “crime”
- **Servant:** “guns”
- **Degree:** LOW

4 Metaphor Annotation

While annotated datasets are needed for computational work on metaphor, creating them is a difficult problem. Generating any rich semantic annotation from scratch is a daunting task, which calls for an annotation standard with a potential for high inter-annotator agreement (IAA). Even trained annotators using specialized tools will frequently disagree on the meaning of sentences—see, e.g., Banarescu et al. (2013). Previous work has found it challenging even to manually annotate metaphors with source and target domain labels (Shutova et al., 2013).

Part of the difficulty is a lack of consensus about annotation schemes. The specification given above is particular to the features that motivate metaphor production and thus allow us to readily explain the meanings of metaphors. It is worth considering the relation of this metaphor-specific annotation to general semantic representation. A significant amount of work has gone into the creation of semantically annotated corpora such as the 13,000-sentence AMR 1.0 (Knight et al., 2014) or the 170,000 sentences annotated in FrameNet (Fillmore et al., 2003). This kind of sentential semantic representation captures the literal meaning of metaphoric sentences. Combined with an LM identification system, such semantic analyses could provide a basis for automatic metaphor interpretation. However, this interpretation would still need to be within a framework for metaphoric meaning, like the one outlined above.

4.1 LM Discovery

This work does not begin with a standard corpus and annotate it. Rather, we rely on the work of Mohler et al. (2015) to manually and automatically gather a diverse set of metaphors from varied text. These metaphors pertain to target concepts in the areas of governance (e.g., *Democracy, Elections*), economic inequality (e.g., *Taxation, Wealth*), and the US debate on gun control (e.g., *Gun Rights, Control of Guns*). Some metaphors were identified manually, using web searches targeted at finding examples or particular source–target pairs. More potential metaphors were found automatically in text by a variety of techniques, described in that work, and were partly verified by human annotators in an ad hoc active learning setting.

The sentences are intentionally varied in the viewpoints of the authors as well as the genres of writing, which include press releases, news articles, weblog posts, online forum discussions, and social media. There are trade-offs in the use of this data set versus the annotation of metaphors in a general-purpose corpus: We will necessarily miss some kinds of metaphors that we would find if we annotated each sentence in a corpus, but this also lets us find more interesting and unique metaphors about a small set of target concepts than we would in that approach. That is, our choice of sentences exhibits the diversity of ways that people can conceptualize the same set of target concepts.

4.2 Manual Annotation and Active Expansion

The corpus of annotated metaphoric sentences are all manually validated. Some of these were annotated entirely manually: An initial set of 218 metaphors were annotated by two of the authors, including three to five examples instantiating each schema slot for the most common schemas. Along with these annotations, we identified potential lexical triggers for each slot, like the examples given in section 3.

A prototype classifier was created and was trained on these annotations. It suggested possible annotations, which were then manually verified or corrected. The use of an automatic classifier is important as it (1) allowed for the more rapid creation of a human-verified dataset and (2) suggests the suitability of these annotations for the creation of future tools for automated metaphor analysis.

The prototype classifier used an instance-based classification approach. Each scenario and schema slot were represented by one to three example linguistic metaphors. The example metaphors were then automatically expanded based on four key features:

1. *The source concept used in the linguistic metaphor*: Each source concept is associated with one or more schemas and their scenarios. These were identified during the development of the ontological categories and the annotation guidelines. An example of an ambiguous source concept is *Body of Water*. A linguistic metaphor about “an ocean of wealth” would be classified in the Nature category, while threatening water, e.g., “a tsunami of guns”, would be classified in Health and Safety.

2. *Grammatical and morphological information in the linguistic metaphor's context.* The sentence containing the linguistic metaphor was annotated with the Malt parser (Nivre et al., 2006), giving the dependency path between the LM's source and target. Custom code collapses the dependency relations across prepositions, integrating the preposition into the relation. The dependency path is used in combination with the lexical item, its voice (if applicable), part of speech, and valence. Expansion is allowed to linguistic metaphors with similar characteristics. For instance, "we were attacked by gun control" (attacked-VERB-passive PREP-BY target, transitive) indicates that the target concept is agentive and should fill a slot that supports agency of the target, such as Threat in Health and Safety or Enemy in the Conflict schema.

3. *The affective quality of the linguistic metaphor.* This feature further disambiguates between potential schema slots, e.g., Threat vs Protection in the Health and Safety schema. The affective quality of the target concept can be used to disambiguate which slot a particular linguistic metaphor should map to.

We determine the affective quality of the LM by the interaction between the source and target concepts similar to Strzalkowski et al. (2014). This involves two questions: (1) Is the source viewed positively, negatively, or neutrally? (2) How do the target concept, and particularly the features of the target concept that are made salient by the LM, interact with the affective quality of the source concept? This is determined by the use of a set of rules that define the affective quality of the metaphor through the target concept, its semantic relation with the source concept, and the valence of the source concept. E.g., in the linguistic metaphor "cure gun control", we can assign "gun control" to the Threat slot in Health and Safety because "gun control" is seen as negative here.

4. *Semantic representation of the source lexeme.* This is used to identify semantically similar lexical items. The semantic representation is composed of three sub-components:

(a) *Dependency-based distributional vector space.* Each word's vector is derived from its neighbors in a large corpus, with both the dependency relation and the lexical neighbor used to represent dimensions for each vector, e.g., NSUBJ_Cure (Mohler et al., 2014). Prepositions collapsed and added to the dependency relation, e.g., PREP_TO_Moon. The distributional representation space ensures that words are only closely related to words that are semantically and grammatically substitutable. (This is in contrast to document- or sentence-based representations, which do not strictly enforce grammatical substitutability.) We do not apply dimensionality reduction. While dimension reduction assists the representation of low-occurrence words, it also forces words to be represented by their most frequent sense, e.g., "bank" would only be similar to other financial institutions. By using an unreduced space, rarer senses of words are maintained in the distributional space.

(b) *Sense-disambiguation* provides an enhancement of the distributional vector. For lexical seed terms that can map to multiple concepts, we use vector subtraction to remove conceptual generalizations that are made to the wrong "sense" of the lexical item. E.g., the distributional vector for "path" contains a component associated with computational file systems; this component is removed by subtracting the neighbors of "filename" from the neighbors of "path". This type of adjustment is only possible because the vectors exist in a space where the alternative senses have been preserved. This step is currently done manually, using annotators to identify expansions along inappropriate senses.

(c) *Recognition of antonyms.* Antonyms generally have identical selectional preferences (e.g., direct objects of "increase" can also "decrease") and almost identical distributional vectors. While they generally support mapping into the same ontological category, they often imply mapping to opposite slots. E.g., the subjects of both "attack" and "defend" map into the Health and Safety schema, but the former is a Threat and the latter is Protection. We use WordNet (Fellbaum, 1998) to identify antonyms of the

lexical seed terms to ensure that generalizations do not occur to antonymous concepts.

Novel linguistic metaphors fitting these expansions are classified into that particular scenario and slot.

Two issues arose related to coverage: The first is with the lack of coverage for mapping linguistic metaphors into schemas due to the low recall of systems that link non-verbal predicates and their semantic arguments. Metaphors occur across a variety of parts of speech, so extending argument detection to non-verbal predicates is an area for future work. The second issue involves insufficient recall of slot values. The schemas allow for a rich population of role and property values. While some of these are metaphorical, many also occur as literal arguments to a predicate. They can also occur in the same sentence or in neighboring sentences.

4.3 Validation

This system gave us a noisy stream of classifications based on our initial seed set. For two iterations of the system’s output, three of the authors were able to quickly mark members of this data stream as correct or not. For the second round of output validation, three of the authors also identified the correct schema slot for erroneous classifications. These iterations could be repeated further for greater accuracy and coverage, providing an ever better source of annotations for rapid human validation. For the first round of validation, we used a binary classification: correct (1) or incorrect (0). For the second round of validation, system output was marked as completely right (1), not perfect (0), or completely wrong (−1). The counts for these validations are shown in Table 1.

Fewer of the annotations in the second round were marked as correct, but this reflects the greater variety of schema slots being distinguished in those annotations than were included in the initial output. One limitation on the accuracy of the classifier for both rounds is that it was not designed to specially handle closed-class slots. As such, the text-excerpt values output for these slots were rejected or were manually corrected.

To measure inter-annotator agreement, 200 of the schema instantiations found by the system were doubly verified. (The extra annotations are not included in Table 1.) For those from Round 2, we collapse the

Rating	Round 1		Round 2	
1	515	67%	732	41%
0	253	33%	752	43%
−1			283	16%

Table 1: Ratings for two rounds of validation of automatic annotations. Correct = 1, incorrect = 0. For round two, ratings for incorrect annotations are split into 0 (not perfect) and −1 (completely wrong).

trinary rating to the original binary classification. The pairwise Cohen κ scores reflect good agreement in spite of the difficulty of the task:

Annotators	Cohen κ
1 and 2	0.65
1 and 3	0.42

4.4 Corpus Analysis

The resulting corpus is a combination of entirely manual annotations, automatic annotations that have been manually verified, and automatic annotations that have been manually corrected. The annotation and verification process is guided by the definitions of the scenarios and their schemas, as given in section 3. However, it also relies on the judgment of the individual annotators, who are native English speakers trained in linguistics or natural language processing. The creation of the corpus relies on our intuitions about what is a good metaphor and what are the likely meaning and motivation of each metaphor.

The result of these initial annotations and the manual validation and correction of the system output was a corpus containing 1,771 instantiations of metaphor schema slots, covering all 14 of the schemas, with more examples for schemas such as Health and Safety and Journey that occur more frequently in text. Statistics on the extent and distribution of annotations for the initial release of the corpus are given in Table 2.

The corpus is being publicly released and is available at (<http://purl.org/net/metaphor-corpus>).

5 Summary

Metaphors play an important role in our cognition and our communication, and the interpretation of metaphor is essential for natural language processing. The computational analysis of metaphors requires

Category	Scenarios	Sentences	Slots	Slot–Value Pairs	Sources	Targets
1. Health & Safety	1	416	4	429	30	11
2. Journey	1	116	6	132	28	11
3. Conflict	1	407	6	482	18	12
4. Power & Control	2	102	6	125	11	12
5. Engineering & Business	3	206	10	219	24	12
6. Morality & Justice	1	129	8	146	9	11
7. Systematic Explanations	1	15	5	19	4	5
8. Plants	1	10	3	12	2	5
9. Animals	1	11	2	11	2	3
10. Human Life Cycle. . .	1	16	4	32	4	5
11. Darkness & Light	1	9	3	16	2	5
12. High & Low	1	45	5	68	9	10
13. Nature	3	25	8	37	7	6
14. Embodied Experience	3	21	11	42	4	8
Total	20	1450	81	1770	68	12

Table 2: Statistics for the corpus of annotations being released. Each category consists of one or more scenarios. For a variety of sentences, the corpus gives instantiations (slot–value pairs) of the slots in the schema for each scenario. Each linguistic metaphor is also tagged as being about one or more source and target concepts. All counts are for unique entries, except for slot–value pairs, which includes duplicates when they occur in the data. Some sentences contain more than one metaphor, so the number of unique sentences is less than the sum of unique sentences for each schema.

the availability of data annotated in such a way as to support understanding. The ontological source categories described in this work provide a more insightful view of metaphors than the identification of source and target concepts alone. The instantiation of the associated conceptual schemas can reveal how a person or group conceives of a target concept—e.g, is it a threat, a force of oppression, or a hindrance to a journey? The schema analysis cannot capture the full meaning of metaphors, but it distills their essential viewpoints. While some types of metaphor seem resistant to this kind of annotation, they seem to be in the minority. We have annotated a diverse set of metaphors, which we are releasing publicly. This data is an important step toward the creation of automatic tools for the large-scale analysis of metaphors in a rich, meaningful way.

Acknowledgments

This work was supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Defense US Army Research Laboratory contract number W911NF-12-C-0025. The US Gov-

ernment is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoD/ARL, or the US Government.

References

- Rodrigo Agerri, John Barnden, Mark Lee, and Alan Wallington. 2007. Metaphor, inference and domain independent mappings. In Ruslan Mitkov, editor, *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP)*, pages 17–23. Borovets, Bulgaria.
- Antonietta Alonge. 2006. The Italian Metaphor Database. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC)*. European Language Resources Association, Genoa, Italy.
- Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin

- Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking. In *Proceedings of the Linguistic Annotation Workshop*, pages 178–86. The Association for Computational Linguistics.
- John A. Barnden and Mark G. Lee. 2002. An artificial intelligence approach to metaphor understanding. *Theoria et Historia Scientiarum*, 6(1):399–412.
- Julia Birke and Anoop Sarkar. 2006. A clustering approach for nearly unsupervised recognition of nonliteral language. In Diana McCarthy and Shuly Wintner, editors, *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 329–36. The Association for Computational Linguistics, Trento, Italy.
- BNC Consortium. 2001. The British National Corpus, v. 2. Distributed by Oxford University Computing Services. URL www.natcorp.ox.ac.uk.
- Dan Fass. 1991. met*: A method for discriminating metonymy and metaphor by computer. *Computational Linguistics*, 17(1):49–90.
- Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.
- Charles J. Fillmore, Christopher R. Johnson, and Miriam R. L. Petruck. 2003. Background to FrameNet. *International Journal of Lexicography*, 16:235–50.
- Jonathan Gordon, Jerry R. Hobbs, and Jonathan May. 2015. High-precision abductive mapping of multilingual metaphors. In *Proceedings of the Third Workshop on Metaphor in NLP*.
- Kevin Knight, Lauren Baranescu, Claire Bonial, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Daniel Marcu, Martha Palmer, and Nathan Schneifer. 2014. Abstract meaning representation (AMR) annotation release 1.0. Web download.
- George Lakoff, Jane Espenson, and Alan Schwartz. 1991. The master metaphor list. Technical report, University of California, Berkeley.
- George Lakoff and Mark Johnson. 1980. *Metaphors We Live By*. University of Chicago Press, Chicago, Illinois.
- Linlin Li and Caroline Sporleder. 2010. Using Gaussian mixture models to detect figurative language in context. In Jan Hajic, Sandra Carberry, and Stephen Clark, editors, *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 297–300. The Association for Computational Linguistics, Uppsala, Sweden.
- James H. Martin. 1994. MetaBank: A knowledge-base of metaphoric language conventions. *Computational Intelligence*, 10:134–49.
- Michael Mohler, Bryan Rink, David Bracewell, and Marc Tomlinson. 2014. A novel distributional approach to multilingual conceptual metaphor recognition. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING)*. Dublin, Ireland.
- Michael Mohler, Marc Tomlinson, and Bryan Rink. 2015. Cross-lingual semantic generalization for the detection of metaphor. In *Computational Linguistics and Intelligent Text Processing*. Springer.
- Srinivas Narayanan. 1997. *Knowledge-based action representations for metaphor and aspect (KARMA)*. Ph.D. thesis, University of California, Berkeley.
- Joakim Nivre, Johan Hall, and Jens Nilsson. 2006. Maltparser: A data-driven parser-generator for dependency parsing. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC)*. European Language Resources Association. Genoa, Italy.
- Ekaterina Ovchinnikova, Ross Israel, Suzanne Wertheim, Vladimir Zaytsev, Niloofar Montazeri, and Jerry Hobbs. 2014. Abductive inference for interpretation of metaphors. In *Proceedings of the Second Workshop on Metaphor in NLP*, pages 33–41. Association for Computational Linguistics.
- Samira Shaikh, Tomek Strzalkowski, Ting Liu, George Aaron Broadwell, Boris Yamrom, Sarah Taylor, Laurie Feldman, Kit Cho, Umit Boz, Ignacio Cases, Yuliya Peshkova, and Ching-Sheng Lin. 2014. A multi-cultural repository of automatically discovered linguistic and conceptual metaphors. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*. European Language Resources Association.
- Ekaterina Shutova. 2010. Automatic metaphor interpretation as a paraphrasing task. In *Proceedings*

- of Human Language Technologies – North American Chapter of the Association for Computational Linguistics (HLT-NAACL)*, pages 1029–37. The Association for Computational Linguistics.
- Ekaterina Shutova, Barry J. Devereux, and Anna Korhonen. 2013. Conceptual metaphor theory meets the data: A corpus-based human annotation study. *Language Resources & Evaluation*, 7(4):1261–84.
- Ekaterina Shutova, Lin Sun, and Anna Korhonen. 2010. Metaphor identification using verb and noun clustering. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING)*, pages 1002–10. The Association for Computational Linguistics, Beijing, China.
- Gerald J. Steen, Aletta G. Dorst, J. Berenike Herrman, Anna Kaal, Tina Krennmayr, and Trijntje Pasma. 2010. *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*. John Benjamins, Amsterdam/Philadelphia.
- Tomek Strzalkowski, Samira Shaikh, Kit Cho, George Aaron Broadwell, Laurie Feldman, Sarah Taylor, Boris Yamrom, Ting Liu, Ignacio Cases, Yuliya Peshkova, and Elliot Kyle. 2014. Computing affect in metaphors. In *Proceedings of the Second Workshop on Metaphor in NLP*, pages 42–51. Association for Computational Linguistics. Baltimore, MD.