Computational Modeling of Metaphor in Discourse

Hyeju Jang

CMU-LTI-17-007

Language Technologies Institute
School of Computer Science
Carnegie Mellon University
5000 Forbes Ave., Pittsburgh, PA 15213
www.lti.cs.cmu.edu

Thesis Committee:

Carolyn Penstein Rosé, Chair
Eduard Hovy
Louis-Philippe Morency
Ekaterina Shutova

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
In Language and Information Technologies

Computational Modeling of Metaphor in Discourse

Hyeju Jang

August 25, 2017

Language Technologies Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Carolyn Penstein Rosé , Chair
Eduard Hovy
Louis-Philippe Morency
Ekaterina Shutova

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Keywords: Metaphor, Metaphor Detection, Computational Modeling of Metaphor, Metaphor in Conversation, Metaphor in Discourse, Frame

For my family

Abstract

Metaphor is used as a language resource/tool to better represent one's point in communication. It can help achieving social goals such as illustrating attitudes indirectly. This thesis aims to understand metaphor from this social perspective in order to capture how metaphor is used in a discourse and identify a broad spectrum of predictors from the discourse context that contribute towards its detection. We build computational models for metaphor detection that adopt the notion of *framing* in discourse, a well-known approach for conceptualizing discourse processes. I claim that developing computational models based on this view paves the way for metaphor processing at the discourse level such as extended metaphor detection, and ultimately contribute to modeling people's use of metaphor in interaction.

In order to model metaphor from this social perspective, we begin with corpus studies to observe people's use of metaphor in three distinct domains where people use different metaphors for different purposes. This foundational work reveals how the layperson conception of metaphor differs from the technical operationalization of linguists from past work. The focus of our subsequent work is on metaphorical language that is recognizable as such by laypersons.

Next, we perform two case studies, which illuminate the value of metaphor detection in discourse, to explore situational factors that affect people's use of metaphor. The first study investigates inner situational factors. We build logistic regression models to discover whether metaphor usage is influenced by three psychological distress conditions including PTSD, depression, and anxiety. Our annotation scheme allows separating effects on language choices of the three factors: contextual expectations, content of the message, and framing. Separating these factors gives us deeper insight into understanding people's metaphor choice, and necessitates consideration of these factors in our next studies. The second study examines external situational factors. We investigate the influence of stressful cancer events on people's use of metaphor. This study verifies the association between the cancer events and metaphor usage, and the effectiveness of the situational factor as a new type of predictor for metaphor detection.

Then, we build computational models for detecting metaphors that can be around related metaphors, not restricted in their syntactic positions. These models find topical patterns by leveraging lexical context, to explore how a metaphorical frame switch is distinguished from a literal one. We design, implement, and evaluate computational models of three kinds: (1) features of frame contrast, which capture lexical contrast around metaphorical frames; (2) features of frame transition, which capture topic transition patterns occurring around metaphorical frames; and (3) features of frame facets, which capture frame facet patterns occurring around metaphorical frames. We demonstrate that these three features in a nonlinear machine learning model are effective in metaphor detection, and discuss the mechanism through which the frame information enables more accurate metaphor detection in discourse.

Acknowledgments

This thesis was such a long journey. In this journey, it came with a lot of frustrating moments, a lot of feeling-stuck moments, and a lot of overwhelming moments, but it also had a lot of happy moments, a lot of proud moments, and a lot of fascinating moments. I am grateful that I had amazing travel companions and that I was able to meet many wonderful people throughout this journey.

My most loyal and supportive companion has been my advisor, Carolyn Rosé. She is a great mentor who might be the most dedicated advisor in the world. She was always able to find time in her busy schedule to answer my questions and help me work through my ideas. She was patient, understanding, and genuine throughout the process, and supported me through difficult times. Her perceptive feedback gave me insight on how to condense a big picture into a single story.From her, I have learned many specific skills, but I've also learned how to be a better researcher and advisor.

I am grateful for my amazing thesis committee: Eduard Hovy, Louis-Philippe Morency, and Ekaterina Shutova. They were always encouraging, and discussing my thesis with them always opened new doors and exiting new ideas to explore. I especially want to thank Eduard Hovy for sharing so much of his time with me and always giving me advice.

I cannot list all my friends' names here, who are all over the world. You kept me going. I would like to thank you all for always being there.

Lastly, I want to thank my family in Korea. I love you.

Contents

List of Figures

List of Tables

Chapter 1

Setting the Stage

"The metaphor is probably the most fertile power possessed by man."

– Jose Ortega y Gasset

It is not difficult to imagine how inefficient and cumbersome it would be if we used only literal language to express our meaning. Almost every concept we refer to is richly multi-faceted, and when we communicate, we tend to focus on some facets more than others. Fortunately, human language provides many linguistic tools for abstraction and deictic reference, which allows us to package meaning in ways that draw upon shared knowledge and suggest additional nuances that we do not fully articulate. These linguistic tools allow us to communicate many aspects efficiently without actually spelling them all out. This abstraction and deictic reference is supported by the human ability of symbolic reasoning, understanding something (*A*) in terms of something else (*B*).

Metaphor is a form of abstraction and manifestation of symbolic reasoning in language. As a type of figurative (nonliteral) language, metaphor highlights similarities between two unlike things for rhetorical effect. For example, in the sentence *Time is money* (Lakoff and Johnson, 1980), we compare *time* to *money*, which is a well-known limited resource, to represent that time is valuable. In other words, the metaphor is seeing *time* (*A*) in terms of *money* (*B*). Seeing *A* in terms of *B* suggests a wide variety of inferences, brought by our shared knowledge of *B*. This shared knowledge includes not only understanding the concept *B* in isolation, but also in its relationships to surrounding concepts. Put differently, shared knowledge includes the information about the connections *B* has and can evoke connections between parts of *A* and other related entities. For example, if we see *pursuing a Ph.D degree* in terms of a *journey*, we can easily make connections between *pursuing a Ph.D degree* and journey-related activities such as *arriving at a destination*, *encountering obstacles*, and *falling in a road*. Another example shows *He is a snake* suggests not only properties of a *snake*, but also how people feel about or judge *snakes* in the culture. In this way, these inferences enable a speaker to find shorter and more effective ways of saying *A* by bringing in the information of *B* the listener already knows. Thus, metaphor allows people to deliver thoughts, feelings, and ideas effectively that might otherwise be difficult to say only by using literal language.

At the same time, casting *A* in terms of *B* may convey the social implications of *B*. If *B* represents some elite knowledge, the metaphor can be a way of bestowing respect or signaling group membership. Similarly, *B* representing some personal knowledge can be a way of showing intimacy or solidarity. For example, when Computer Science Ph.D students talk about the romantic dating process in terms of a machine learning algorithm, they can feel a sense of belonging to a particular group. This kind of social effect occurs at the same time that propositional meaning is communicated. Thus, when considering metaphor, we need to address the two dimensions of meaning that are communicated within a discussion: first, the propositional meaning of what a speaker is saying, and second, the social meaning of who the speakers are, their personal identities, their relationships with one another, and the communities in which the interaction is situated.

Metaphor is a unique language tool in that a metaphor invites participants to contribute to the meaning of a conversation in new ways, because it provides facets and perspectives of a new domain. The new domain affects how other participants might respond and become involved in the communication. For example, EX(1)–EX(4) from the same thread in the breast cancer discussion forum shows how conversational participants repeat and expand one another's metaphors by linking facets of *wagon* such as *falling*, *part of a journey*, and *on weagon*.

EX(1) "**falling off the wagon** is no big thing in my opinion, the psychological good feelings of enjoyment weigh in big for feeling good."

EX(2) "**** **falling off is part of this journey**, it is stupid to deny yourself everything."

EX(3) "I am **on the wagon** so far today ...ongoing battle."

EX(4) "**** - hope you **stay on the wagon**, or at least **get back on after you fall**!"

Metaphor has fascinated scholars in a wide variety of fields including philosophy, cognitive science, sociology, and computational linguistics. Broadly speaking, conceptions of metaphor in all of these fields embrace some version of the ideas communicated above. However, each one approaches metaphor from its own distinct methodological perspective and emphasizes some parts of the picture over others, depending on the shared values and goals of that scholarly community.

Philosophers, for example, seek to understand how people work in terms of discrete logical formalizations that can be manipulated using prescriptive rules of inference. This places philosophers at the high abstraction end of the continuum. They posit formalizations that match their subjective experience and reason about whether the implications are similarly consistent with their experience. They seek examples as illustrations, but are not strongly empirical in their approach. For instance, within language pragmatics, the Gricean model of the cooperative principle (Grice, 1975) and the Relevance Theory model (Sperber et al., 1986) offer formalizations consistent with our subjective experience about language, based on a philosophical approach. These models attempt to describe how metaphor works as a linguistic phenomenon. The Gricean model explains that flouting Grice's maxims invites a metaphorical interpretation. The Relevance Theory model looks at metaphors as general examples of loose talk. While both models address some important questions in metaphorical language use, they do not offer direction in terms of computation or empirical work since such approaches to metaphor fall outside the purview of their approach.

In contrast, cognitive science, while still placing a value on general principles that can be used to make causal claims, places a greater value on empiricism and has a greater aversion to over-generalization. Though a propositional understanding of metaphor can reach into the social sphere, cognitive scientists shy away from these aspects of meaning because the complexity of social interaction would force them to simplify more than they are willing to in applying their methodology to their object of study. For example, the Conceptual Metaphor Theory model (Lakoff and Johnson, 1980) of cognitive linguistics, the most recognizable view of metaphor, claims that metaphor occurs as a way of thinking beyond the level of language. The main premise of this approach is that "the essence of metaphor is understanding and experiencing one kind of thing in terms of another" (Lakoff and Johnson, 1980). The model depicts metaphor as linking two conceptual domains: *source* and *target*. The source domain is where a metaphorical expression resides and the target domain is the domain we try to understand. For example, in the sentence "*He attacked every weak point in my argument*" where *attacked* is metaphorically used, the source domain is *war* and the target domain is *argument*. According to this theory, we conceptualize argument in terms of *war* (where people attack, shoot, defend, and win) based on our culture, which affects the way we talk about *argument*. From empirical work that uses the Conceptual Metaphor Theory we can understand which concepts have typically been used to communicate with others, and what nuances of meaning are communicated through a mapping between structures. However, the theory does not offer guidance in understanding metaphor from a social perspective because its scope is limited to the cognitive aspects of communication.

Unlike both philosophical and cognitive science approaches, sociological approaches to metaphor embrace the social. With a similar aversion to over-simplification, sociologists who study metaphor limit the extent to which they are willing to abstract instead of limiting the scope of their interest. Using a sociological flavor of empiricism, they seek to preserve complexity, and so give up the opportunity to make causal claims or formalize generalizable principles. For example, from discourse studies, the Figured World model (Gee, 2014) explains metaphor as a reflection of master figured-worlds, based on a sociological approach. According to Gee (2014), a *figured-world* models in our minds what is 'normal' or 'typical'. A master figured-world is a way of seeing aspects of the world shared widely across a number of significant domains in our society. Metaphors are sometimes connected to "master models" in that what they imply is used widely in a given culture or social group to organize a number of significant domains. This model gives deep and nuanced insight into how metaphor is used and how it functions from a social perspective. However, it does not provide the formalization that is important for computationalization.

In the computational fields such as Computational Linguistics and Natural Language Processing, researchers have commonly viewed metaphor based on the Conceptual Metaphor Theory, and endeavored to operationalize source and target domains. Since metaphor brings two different domains into correspondence, we can often observe words from both domains in one metaphorical expression, which is distinct from literal language. For example, in EX(5), the adjective *thirsty* usually requires live creatures such as humans or animals as its subject in literal language. Therefore, when *thirsty* takes *car* as its subject, it becomes a good candidate for a metaphorical expression. Similarly, in EX(6), the verb *pour* is metaphorically used by looking at its object *funds* as *liquid*. People do frequently use the "money is a liquid" metaphor in everyday language. Many computational approaches find, by approximating the domains in different ways, these kinds of patterns appearing in source and target domain mappings (refer to Chapter 2 for more details). The approaches based on the Conceptual Metaphor Theory have been useful for operationalizing metaphors at a sentence level when both source and target domains appear within the same sentence.

EX(5) I have been driving for 3 hours without stopping at a gas station. My car is *thirsty*.

EX(6) If the venture is successful, investors will *pour* funds in.

While no existing theory/approach is sufficient, all reflect some aspects that are important. Since one must always grapple with the tension between respecting the inherent complexity in our object of study and seeking a formalization, there can be many different models from different perspectives. Yet, there is no complete model for the purpose of modeling metaphor in discourse. To address this gap, this thesis provides a theme-based model that is more comprehensive and also empirically tested in a domain, using a computational model that allows for making predictions by embracing the social aspect of metaphor empirically. Our model is simpler than existing models discussed above, but it does provide useful insight, which can support computation.

1.1 Definition of Metaphor

Metaphor is a type of nonliteral language – metaphorically used words mean something beyond the exact meaning of the words. In our definition, we call an expression a metaphor when the expression satisfies the following three conditions: (1) the expression, whether a single word or a group of words, needs to have an original established meaning, (2) the expression needs to be used in context to mean something significantly different from that original meaning, and (3) the difference in meaning should not merely be a difference in intensity or polarity. (1) and (2) apply for any forms of nonliteral language, but (3) is distinct in metaphor. Explanations for each condition are following.

(1) **Have an original meaning**

The expression or the words within the expression need to have original established meanings. For example, in the phrase "kick the bucket" the words "kick" and "bucket" have clear and commonly known original meanings. The phrase "kick the bucket" could easily be used according to those meanings; in other words, to mean, "strike the bucket with ones foot".

(2) **Alter the original and established meanings of the words**

The usage needs to change the original meaning of the expression in some way. For the same example, "he kicked the bucket," the nonliteral meaning of "he died" is far from the literal meaning of "he struck the bucket with his foot." In many idioms such as this, this meaning will be non-compositional, meaning that it cannot be determined from the meanings of the individual words.

Another example would be "keep your eyes open" which has an obvious original meaning. An ophthalmologist might use it in that way, to keep your eyelids open while examining your eyes. But it is also used to mean, "to watch carefully for someone or something, often while you are doing something else." This meaning does not derive from the original meanings of "keep," "eyes" and "open" but it is more easily interpreted than idioms like "kick the bucket."

(3) **Should not merely alter the intensity and polarity of the meaning**

The usage needs to alter the original meaning of the expression but should not simply be a change in the intensity or the polarity of the meaning. Language uses like hyperbole and understatement may simply change the intensity of the meaning without otherwise altering it. If one says "I am starving," the original meaning would be that the person is "dying from a lack of food." But if used to mean, "I am very hungry" then it is nonliteral and an example of hyperbole. The meanings of "starving" and "hungry" are very similar, but with "starving" having a much more intense meaning. One could also be hyperbolic with simple use of intensifiers like "very." Conversely, understatement works in the opposite direction. For example, the sentence "I know a little about running a company" when said by a highly successful businessman would not mean that they actually know only a little bit about business. In that case, the true meaning is simply stronger than the literal meaning but not otherwise different.

Language uses like sarcasm instead change the polarity of the meaning. Sarcasm is when "you say the opposite of what you really mean in order to be rude to someone" [note for Collins dictionary]. As such, the polarity of the expression is changed. For example, a person can say "you are such a wise guy" when the person he or she was talking to did something stupid. In this context, what the expression actually means is "you are not wise," negating the literal meaning of the original sentence. Sarcasm is often difficult to detect in writing, which also would make it harder to code.

In this thesis, I do not deal with nonliteral language related to the intensity of the language such as hyperbole and understatement, which exaggerate or weaken the original meanings in the same direction of the meaning, or sarcasm, which negates the original meaning. Note that although we do not include hyperbole, understatement, and sarcasm in general, those linguistic paradigms can be used along with metaphor. For example, if a 20 year old referred to a 30 year old as "a dinosaur" the word "dinosaur" would be used nonliterally to mean "very old" as a metaphor but in that case it would also be a hyperbole. In other words, "dinosaur" is both a metaphor and a hyperbole.

The scope of metaphorical language in this thesis includes expressions such as metaphors and idioms that alter the meanings of the words. I also include similes that work similarly to metaphors but are explicitly marked with the words such as "like" or "as".

1.2 Our View on Metaphor for Computational Modeling

A useful way to model metaphor is adopting the concept of *framing* in discourse (Fauconnier and Turner, 1998; Fillmore, 1976; Gee, 2014; Minsky, 1975; Schank and Abelson, 1975; Tannen, 1993; Tannen and Wallat, 1987), a well-known theory to understand discourse used in various forms in many fields such as linguistics, cognitive psychology, and artificial intelligence.

According to Tannen (1993), a *frame* is a "structure of expectations".[1] For example, what we expect in a classroom involves the typical sequence of events: a teacher explains a concept, students look at the screen, students listen to the teacher, a student asks a question, and so on. We would also expect students to use more polite language with their teacher than with their friends.

Using the notion of *frame*, we now define metaphor as follows.

> Def. A metaphor occurs when a speaker brings one frame into a context/situation governed by another, and explicitly relates parts of each, so that the original frame's expectations are extended/enhanced according to the new frame.

For example, the *journey* metaphor is frequently used in breast cancer support discussion forums as seen in examples EX(7) and EX(8) (Refer to Appendix B for all the example posts). While people are talking about their cancer treatment based on their *cancer frame*, they use metaphors such as *joining you on your journey*, *road*, and *moves along* from the *travel* domain. The literal meaning of those metaphors are all related with the *travel* frame. We can see the metaphors as switching frames from *cancer* to *travel*.

EX(7) Hello Ladies! I was supposed to start chemo in January, ... I cant start tx until that is done. So I will be *joining you on your journey this month*. I AM SICK OF the ANXIETY and WAITING. Ug,. Sorry. BFF you are so right, Hurry up and wait. (Post 1 in Appendix B)

EX(8) The *road* seems long now but it really *moves along* fast. (Post 3-(13) in Appendix B)

$$Cancer \left\{ \begin{array}{l} \text{Patient} \\ \text{Start state} \\ \text{Treatment process} \\ \text{...} \end{array} \right\}$$

$$Travel \left\{ \begin{array}{l} \text{Traveler} \\ \text{Start date} \\ \text{Path} \\ \text{...} \end{array} \right\}$$

Typically, we maintain multiple frames actively in our mind and jump from one frame to another. Switching frames can happen not only for metaphors but also for other reasons, like topic changes. Consider the following example EX(9). We can see that the post is about the poster's difficult cancer experience not only by where this post is from, but also by the words being used such as *survivorship* and *tough year*. Another thing we can see is that this person has a religious frame in her mind from the words such as *God*, *creation*, and *spiritual*. In addition, we can see that she represents her cancer experience using the word *journey*. We don't call the switch between *cancer experience* and *religion* a metaphor; however, we call the switch between *cancer experience* and *travel* a metaphor. This is because *travel* is used for representing *cancer experience* based on some similarities between the two frames, but *religion* is used for talking about another thing happening in connection to *cancer experience*. If someone describes his summer travel after talking about chemotherapy, the *travel* frame would not be metaphorical because *travel* here has another relation with *chemotherapy*, not connected by their similarity.

[1] *Frame* has been called different names such as *script*, *schema*, and *figured-world*, as discussed above.

EX(9) hi ladies thought I would come by and say hi... I am getting ready to celebrate my 1 yr survivorship... and I have decided to put together a video clip to express my gratitude and thanks to God for getting me through this very tough year. Here is the link... I hope you enjoy it. I have called this video, a new creation. *It is my spiritual journey over the past year.*

Feel free to pass on to anyone who you think may enjoy this thanksgiving. He really is an amazing God.

God bless each and every one of you ladies. (Post 5 in Appendix B)

$$Cancer \left\{ \begin{array}{l} \text{Patient} \\ \text{Duration} \\ \text{...} \end{array} \right\}$$
$$Religion \left\{ \begin{array}{l} \text{God} \\ \text{Spiritual} \\ \text{Bless} \\ \text{...} \end{array} \right\}$$
$$Journey \left\{ \begin{array}{l} \text{Traveler} \\ \text{Duration} \\ \text{...} \end{array} \right\}$$

What makes a certain frame switch metaphorical or literal? To be metaphorical, there should be associated mappings between the two frames that are involved in the frame switch, but the associated mappings could be either visible or invisible in discourse. For example, in EX(7), *journey* is mapped with *chemo*, and both words are visible in the text. In contrast, in EX(9), the target of the metaphor *journey* is not visible as a word or phrase in the text.

In other words, let's assume that there is a frame governing a section of discourse. The frame can have facet slots, and stereotypical slot filler types for each facet slot. For example, the *cancer treatment* frame can have facets such as *patient* and *treatment process*. When the governing frame has slot fillers that are typical such as the words "patient" and "chemo", the frame is literal. However, when the fillers are not typical such as the words "warrior" and "road", which are drawn from some other frames, then these fillers make the other frames metaphorical in the context of the governing frame. In this setting, the fillers constitute the mapping.

This formulation has some advantages in modeling metaphor. First, it allows the same frame to be sometimes metaphorical and other times literal. An example could be *grad school* as a metaphor for *life* and *life* as a metaphor for *grad school*. Second, a conventional or dead metaphor can be operationalized simply as when a frame has two possible sets of fillers, from two different domains, with one set predating the other.

When we look at metaphorical language in use as switching frames, we encounter several questions: *What* a frame looks like, *why* we want to switch a frame, *how* we make the switch, and *what effect* the switch has. First, we can see what a frame looks like from how people talk. For example, different frames such as *cancer frame*, *forum frame*, and *journey frame* can be observed from the groups of words related with cancer, the forum, and journey, represented in their posts, respectively. Second, by switching to the *journey* frame, people show aspects of the frame that are important to them. People using the *journey* metaphor might want to understand the experience of cancer treatment as a process of progressing along a path, struggling and learning, which allows for each person's experience to differ without judgment of personal success or failure (Reisfield and Wilson, 2004). They could focus on the cost of the cancer treatment, but then the *travel* frame does not really do so. If they focused on cost, they might use the *shopping* frame instead. Third, we can observe the switch in the words from the travel domain.

Finally, this frame switch could have an effect of increasing empathetic understanding between the participants (Ritchie, 2013). Additionally, reusing each other's metaphors reduces emotional distance and helps to build empathic understanding and bonding through a shared perception of their situations. These questions motivate my study of this thesis, in which I build computational models for metaphor in everyday language use, based on the *framing* view on metaphor. I try to understand and computationalize the ways that metaphors are used in actual discourse.

It is not the definition of metaphor per se that we challenge in this work, but rather the comparison between metaphorical language and literal language. While much prior work treats metaphor as deviant while literal language is the norm, in this framing view, we characterize metaphor as frequently unexpected in a context governed by another frame. However, while this unexpected flavor leads to interpretation by means of the recognition of the inserted frame as a flouting of a contextual expectation, we do not treat it as breaking a rule or violating a standard. Thus, our interpretation approach takes on a Gricean character. This view has advantages when it comes to modeling metaphors that are not, in fact, rarer than their literal counter-parts. A key focus of our work, therefore, is on modeling what makes certain transitions in discourse signal that text as part of a frame switch should be taken non-literally. This new perspective suggests experimentation with new computational approaches for detection of discourse level metaphors.

1.3 Task in This Thesis

Prior computational work on metaphor has mostly focused on metaphor detection at the level of an individual expression in isolation from its context, and has focused on a restricted set of predictors of metaphor available within the expression including selectional restriction, lexical concreteness, and source and target domain mappings. In this case, the scope of metaphor detection is usually limited to a metaphorical expression occurring within a sentence, frequently as a form of S+V+O (EX(10)) or A+N (EX(11). This approach can be extended to other syntactic relations within a sentence because a key clue to determine a metaphor can be found in another relation. This kind of metaphor might not be as popular as S+V+O or A+N. But selectional preferences can be observed in more diverse syntactic relations than S+V+O and A+N relations (Jang and Mostow, 2012) although selectional preferences strength could vary depending on a relation.

EX(10) This weekend trip will *eat up* my gloominess.

EX(11) He is releasing *dark* energy.

More recent approaches for metaphor detection use the context of a target expression based on the idea that metaphorically used words violate lexical cohesion of the context. For example, the words in "breaking the ice" are not semantically related to other words in their context as shown in EX(12). These approaches do not restrict syntactic positions of a target expression as in the approaches using local information within an expression. However, the text where a metaphor occurs should be topically cohesive around the target domain. In this case, the scope of metaphor is limited to an individual metaphorical expression, which is lexically isolated from the topic of its context. Therefore, detecting metaphors around which topically related metaphors co-exist (EX(13)) remains a daunting open challenge for metaphor detection in discourse.

EX(12) ... “Some of us may have have acted as critics at one point or another, but for the most part its just as film-goers,” he said. And, *breaking the ice* at a press conference, he praised his vice-president, French actress Catherine Deneuve. ...

EX(13) In a sense weve come to our nations capital to cash a check. When the architects of our republic wrote the magnificent words of the Constitution and the Declaration of Independence, they were signing a promissory note to which every American was to fall heir. This note was a promise that all men, yes, black men as well as white men, would be guaranteed the “unalienable Rights” of “Life, Liberty and the pursuit of Happiness.” It is obvious today that America has defaulted on this promissory note, insofar as her citizens of color are concerned. Instead of honoring this sacred obligation, America has given the Negro people a bad check, a check which has come back marked “insufficient funds.” – Martin Luther King, Jr. “I Have a Dream” speech, 1963

In this thesis, I address the problem of analyzing and detecting metaphors that are unrestricted in their syntactic positions and that can be around related metaphors. In other words, I do not restrict types of metaphors in terms of their syntactic positions and semantic neighbors. For analytic tasks, I use all metaphors appearing in text (Chapter 3, Chapter 4) and several target metaphors (Chapter 5). For detection tasks, I detect several target metaphors that are not restricted in their syntactic positions and that may be around related metaphors (Chapter 6, Chapter 7, Chapter 8). To do so, we need to deeply explore how topical patterns change around metaphorical frames to discover what makes a frame metaphorical or literal. In order not to confound the topical patterns, I restrict target and source domains into the cancer domain and “journey” related metaphors for developing models, and discuss issues appearing when applying the models to other domains.

1.4 Structure of This Thesis

This thesis is structured around three questions on metaphor: (1) what is metaphor, (2) why do people use metaphor, and (3) how is metaphor represented in discourse. Table 1.1 shows the questions and corresponding chapters.

Topic	Chapter
What is metaphor?	Chapter 1, Chapter 2. Chapter 3
Why do people use metaphor?	Chapter 4, Chapter 5
How is metaphor represented in discourse?	Chapter 6, Chapter 7, Chapter 8

Table 1.1: Structure of this thesis.

What is metaphor? There have been diverse discussions about what how to conceptualize metaphor. Chapter 1 and Chapter 2 talk about how prior work characterizes metaphor and how this thesis views metaphor. Chapter 3 reports a corpus study as a first step towards computational work on social and discourse functions of metaphor. To study the functions of metaphor in discourse, we need metaphor annotation reflecting what laypeople perceive as metaphor. The mechanism that gives the social effect of metaphor is not fully identified yet. Although all types of metaphor (e.g., novel metaphors and dead metaphors) would have effect to some degree, I am interested in more direct effects metaphor has, rather than indirect effects that originate from the fact that metaphor affects or reflects our cognition or psychology. I use Amazon Mechanical Turk (MTurk) to annotate data from three web discussion forums covering distinct domains, then compare these to annotations from our own annotation scheme which distinguishes levels of metaphor with the labels: *nonliteral*, *conventionalized*, and *literal*. The focus of the subsequent work in this thesis is on metaphorical language that is recognizable as such by laypersons.

Why do people use metaphor? We know that metaphor is creative at its core, and these linguistic regularities, though essential, are bounded in their ability to enable accurate metaphor detection in a broad sense. In contrast to previous approaches focusing on the linguistically inspired features, we begin to explore situational factors coming from a pragmatic perspective, related to the reasons why people choose to use metaphors. These situational factors may provide a complementary set of indicators to partner with tried and true linguistically inspired features in order to increase performance. Chapter 4 examines the influence of inner situational factors on people's use of metaphor. We build logistic regression models to discover whether metaphor usage is influenced by three psychological distress conditions including PTSD, depression, and anxiety. Our annotation scheme allows separating effects on language choices of the three factors: contextual expectations, content of the message, and framing. Separating these factors gives us deeper insight into understanding people's metaphor choice, and necessitates consideration of these factors in our next studies. Chapter 5 investigates the influence of external situational factors on people's use of metaphor. This study verifies the association between the cancer events and metaphor usage, and the effectiveness of the situational factor as a new type of predictor for metaphor detection.

How is metaphor represented in discourse? In language technologies, researchers have identified metaphor based on linguistic patterns observed in metaphors, mostly using local contextual cues within a single expression. Chapter 6 shows the limitation of the approach using local cues within a sentence in metaphor detection in discourse, and presents an approach that leverages lexical contexual information of a discourse to detect metaphors based on frame contrast patterns occurring metaphorical frames. Chapter 7 explores frame transition patterns around a target word. Chapter 8 builds metaphor frames using a semi-supervised bootstrapping approach, and apply the frame facet information to detect metaphors. The frames can be build manually, but it is not that easy, especially considering that metaphor is cultural. In addition, it is difficult to define the levels of frames, and make slots for all the levels, even only for most frequently used metaphors. This chapter also discusses issues appearing in generalizing the models presented in this thesis.

This thesis is in the beginning phase, working towards the goal of computational modeling of social and discourse uses of metaphor. In contrast to the large body of work on uncovering the intended propositional meaning behind metaphorical expressions, this thesis is most interested in the illocutionary and perlocutionary force of the same contributions. This thesis opens up a new direction of computational work on metaphor considering novel features for metaphor detection.

Chapter 2

Related Work

In this chapter, I introduce theoretical and computational work on metaphor in discourse, relevant to the scope of this dissertation. In Section 2.1, I give an overview of the research performed on metaphor as everyday language practice in the fields of linguistics and psychology, which provides the theoretical foundation for this dissertation. In Section 2.2, I discuss previous approaches to manual metaphor annotation, which supplies a comparative standard for computational modeling of metaphor. Lastly, in Section 2.3, I discuss computational approaches to modeling metaphor, focusing particularly on metaphor detection and the effect of metaphor on discourse.

2.1 Theoretical Work on Metaphor in Discourse

The Conceptual Metaphor Theory (Lakoff and Johnson, 1980) introduced in Chapter 1, has sparked an active line of research on metaphor in discourse, looking at metaphor as everyday language practice. For example, Cameron (2003) studied the use of metaphor in spoken discourse in classrooms, Liendo (2001) analyzed *war* metaphors in the business field, and Mio (1997) examined how metaphors are used as persuasive devices in politics. Additionally, Levitt et al. (2000) investigated the use of *burden* metaphors in relation to depression. Suárez Toste (2007) investigated how metaphors, specifically personification, are used inside the wine cellar. As we can see from these studies, metaphor is used in diverse domains of our life.

The studies on metaphor in discourse have revealed some interesting findings on metaphor use in everyday real world language. Previously, it was commonly considered that metaphor is mainly used in creative genres of writing such as poems and novels. However, Carter (2004) argued against this preconception. He claimed that participants in every day speech are as creative as writers of more poetic discourse registers, based on a 5-million word corpus of spoken conversations (Kaal, 2012). What metaphors look like in everyday language is also different from what people have commonly thought about metaphor. Cameron (2003)'s findings countered the assumption that typical metaphors are nominal, vivid and active. These findings include that the '*A is (like) B*' type of metaphor is rare, and that most metaphors are verbs and highly conventional. Cameron (2003) suggested that language users do not necessarily process metaphors that metaphor researchers identify as metaphors.

In addition to what metaphor looks like in everyday language, studies have paid attention to the effect metaphor brings to discourse. Metaphors can be used for a number of conversational purposes such as increasing or decreasing social distance or as a tactic of persuasion or manipulation (Ritchie, 2013). Carter (2004) demonstrated that creative metaphors are commonly used for humorous and interpersonal purposes rather than for more serious purposes such as explaining difficult abstract concepts in spoken language. Similarly, Cameron (2003), focusing on the affective dimension of metaphors, emphasized growing intimacy as a function of metaphor by observing metaphors used in classroom discourse. On the other hand, political discourses or press discourses are example genres in which metaphors are used in a calculated way. Charteris-Black (2004) discussed how metaphor arouses emotions and guides evaluation in political discourse. Similarly, Mio (1997) discussed how metaphors in political discourse can have persuasive power and guide the direction of people's opinion. Liendo (2001) showed how metaphors can create a reality rather than reflecting it by analyzing business-related metaphors. As a unique function of metaphor, Low et al. (2010) and Deignan (2010) showed an evaluative function of metaphors used to describe wine flavors.

This thesis contributes to this line of research on metaphor in everyday language use from the perspective of computational linguistics. We study metaphor use appearing in online discussion forums and clinical interviews, and observe how metaphor is used as a linguistic tool in the data. In addition, we attempt to quantify why people use metaphor and what effect metaphor brings to communication, and use this information to computational prediction tasks.

2.2 Metaphor Annotation

One of the main challenges in computational work on metaphor is the lack of annotated datasets. In contrast to common thinking, metaphor annotation is difficult because metaphor is fuzzy by its nature. In other words, it is difficult to decide the boundary between being metaphorical and being literal. This is mainly due to conventionalized metaphors, or dead metaphors, which are words that are widely used for a long time and are no longer recognized as metaphorical by laypeople. With use, some metaphorical senses become conventionalized and lose their metaphoricity, particularly if the original basic sense is no longer used. In this way, a word may acquire new senses because its metaphorical use has been conventionalized. Given the process of conventionalization, deciding the boundaries of a metaphor can be difficult. However, metaphor annotation is critical in computational work on metaphor because it provides the gold standard to which we will compare our computational model to see how well our model works. Thus, metaphor annotation provides a foundation for what counts as metaphorical, which is specific to the NLP application we are building a computational model for.

The most popular metaphor annotation scheme is the Metaphor Identification Procedure (MIP), presented by Pragglejaz-Group (2007), which introduces a systematic approach with clear decision rules. In this scheme a word is considered to be metaphorical if it is not used according to its most basic meaning, and if its contextual meaning can be understood in comparison with the most basic meaning (Figure 2.1). This method is relatively straightforward and can provide high inter-reliability among annotators. Additionally, the method is flexible, depending on the conditions used to identify the basic word meaning. However, defining the basic meaning of a

1. Read the entire text-discourse to establish a general understanding of the meaning.
2. Determine the lexical units in the text-discourse.
3. (a) For each lexical unit in the text, establish its meaning in context, that is, how it applies to an entity, relation, or attribute in the situation evoked by the text (contextual meaning). Take into account what comes before and after the lexical unit.
   (b) For each lexical unit, determine if it has a more basic contemporary meaning in other contexts than the one in the given context. For our purposes, basic meanings tend to be
      - More concrete [what they evoke is easier to imagine, see, hear, feel, smell, and taste];
      - Related to bodily action;
      - More precise (as opposed to vague);
      - Historically older; Basic meanings are not necessarily the most frequent meanings of the lexical unit.

   (c) If the lexical unit has a more basic current-contemporary meaning in other contexts than the given context, decide whether the contextual meaning contrasts with the basic meaning but can be understood in comparison with it.
4. If yes, mark the lexical unit as metaphorical.

Figure 2.1: Procedure and explication of the MIP (Pragglejaz-Group, 2007)

word is difficult. For Pragglejaz-Group (2007)'s purpose, basic meanings tend to be more concrete, related to bodily action, more precise, and historically older. This definition tends to result in a large proportion of words being annotated as metaphor that only linguistic experts identify as metaphorical. As a result, many of the annotated words would not be considered metaphorical by laypeople due to the words' long and widespread usage, i.e., dead metaphors. Despite this weakness, MIP has been used frequently as a basis for annotation in other research (Shutova and Teufel, 2010; Steen et al., 2010).

Steen et al. (2010), Shutova and Teufel (2010), and Shutova et al. (2013) have expanded upon the MIP. Steen et al. (2010) introduced the Metaphor Identification Procedure VU University Amsterdam (MIPVU), which uses the core protocol of MIP, but expands and refines it to be more useful for analyzing metaphor in discourse. Specifically, MIPVU does not allow word class boundaries to be crossed. For example, the contextual meaning of a verb is not compared with its basic meaning as a noun. When the word *cross* is used as a verb (e.g., *crossed* the Atlantic), it is not compared with its basic meaning as a noun (e.g., a gold *cross*) to decide its metaphoricity. In addition, MIPVU distinguishes different types of metaphors including direct metaphor, indirect metaphor, implicit metaphor, borderline cases of metaphor, metaphor signals, and metaphor due to personification. Shutova and Teufel (2010) and Shutova et al. (2013) have focused more on conceptual metaphors. The annotation procedure in these studies added to MIP a step for identifying underlying conceptual mappings between source and target domains, by comparing the contexts of a word when the word is literally vs. metaphorically used. These two extensions of the MIP better capture more nuanced understandings of metaphor in discourse. Their work, however, is still grounded in an identification scheme built on an expert perspective.

In addition to the annotation schemes described above, other annotations rely more on laypeople's intuition about metaphor. Wallington et al. (2003) conducted experiments to investigate various clues and patterns of metaphor usage. Two different teams annotated the same text with different instructions, one asked to label "interesting stretches" and the other "metaphorical stretches". They also asked annotators to tag words or phrases that indicated a metaphor nearby, in order to investigate signals of metaphoricity. This study is meaningful in that it contributed to identifying likely metaphors in the eyes of laypeople. However, for the purpose of their study, the form of metaphor in their annotation is not restricted but general and diverse including noun phrase, verb phrase, adjectival and adverbial phrases, prepositional phrase, and clause and sentence. Thus, it is difficult to be applied in the current level of computational modeling of metaphor where the scope of the metaphor form is usually limited to a word or a sentence.

Also investigating laypeople's understandings of metaphor, Beigman Klebanov and Flor (2013) built a metaphor annotation protocol for metaphors relevant to arguments in essays. They were interested in identifying metaphors that stand out and are used to support the writer's argument. Instead of giving a formal definition of a literal sense, the annotators were instructed to mark words they thought were used metaphorically and to write down the point being made by the metaphor, given a general definition of metaphor and examples. However, Beigman Klebanov and Flor (2013) target metaphors relevant to arguments in essays whereas our work focuses on conversational discussion data which are less well-formed and coherent.

Extending the previous work, this thesis looks at layperson perspectives of metaphor by testing the extent to when laypeople and experts recognize metaphors differently. In addition, I discuss why a layperson-based annotation scheme is necessary for computational work that is focused on how metaphors affect the ways people communicate in discourse.

2.3 Computational Work on Metaphor

Computational modeling of metaphor has been an active area of research for the past decades, mainly focusing on metaphor detection. More recently, however, research on more diverse topics has been carried out. Among them, this section discusses the prior work on metaphor detection, effect of metaphor in discourse, and metaphor property extraction, which are most relevant to how people use metaphor to communicate.

2.3.1 Metaphor Detection

Based on the metaphor annotation research described in 2.2, there has been a great deal of research on metaphor detection. Frequently people have drawn heavily from ideas that are reminiscent of the Conceptual Metaphor Theory (Lakoff and Johnson, 1980). The approaches can be categorized based on the type of metaphors they focus on. I define the metaphor type based on three dimensions of how the metaphor is situated. I explain the three dimensions below.

Dimension 1: restricted syntactic position vs. any position

The first dimension concerns the syntactic structure of the expression where a metaphor is located. Many researches detect metaphors in particular syntactic structures such as S+V+O or A+N whereas other researches detect metaphors regardless of their positions in sentence structures.

Dimension 2: within sentence vs. beyond sentence

The second dimension concerns the source of clues for deciding whether a certain expression is metaphorical or not, in other words, where the clues for detecting the metaphor come from. Some metaphors require clues only from the sentence where they are located whereas other metaphors cannot be detected by only using clues within the same sentences.

Dimension 3: individual vs. related The third dimension concerns the individuality of a metaphor. Some metaphors are individual in that they are not semantically or topically related to any other metaphors in the same discourse. In contrast, other metaphors are related under the same topic, and do span across the same discourse; these are called *extended metaphors*.

Detecting different types of metaphor requires different approaches, and there do not yet exist computational methods that can detect all types of metaphors within the three dimensions. Previous work has focused its attention on detecting metaphors in the dimensions 1 and 2. However, within the dimension 3, this related type of metaphor has not been addressed because it requires the processing of both the whole discourse and the metaphor topic frame.

In this section, I introduce metaphor detection approaches based on the categorization by the three dimensions, as shown in Table 2.1. Note that this table is two dimensional since all prior work focuses on individual metaphors; no prior approaches address the problem of extended metaphor.

		Dimension 1	
		Restricted syntactic position	Any position
Dimension 2	Within sentence	Section 2.3.1.1	Section 2.3.1.2
	Beyond sentence	N/A	Section 2.3.1.3

Table 2.1: Overview of Section 2.3.1 Metaphor Detection

2.3.1.1 Detecting metaphors in a restricted syntactic position within a sentence

Computational work on metaphor detection has largely focused on metaphor detection within individual sentences. This task requires approaches for modeling source and target domains appearing in the certain syntactic structures. These approaches can be classified into (1) work using selectional preference violation, (2) work using metaphorical domain mappings, and (3) work using lexical concreteness. Most of this work uses datasets that comprise grammatically restricted sentences (*e.g.,* ones with S+V+O or A+N structures) for their experiments, in order to test their hypotheses in controlled ways.

Selectional preferences violation:

One of the most popular approaches to metaphor detection using violation of lexical contextual constraints is the idea that metaphors violate selectional preferences. Selectional preferences relate to how semantically compatible predicates are with particular arguments. For example, the verb *drink* prefers *beer* as an object over *computer*. The idea behind using selectional preferences for metaphor detection is that metaphorical words tend to break selectional preferences and violate local lexical coherence within a sentence. In the case of "*the clouds sailed across the sky*", for instance, *sailed* is determined to be metaphorical because *clouds* as a subject violates its selectional restriction. The idea of using violation of selectional preferences as a cue for metaphors has been well studied in a variety of previous work.

One of the first models using the lexical contextual violation is (Fass, 1991). This model uses selectional preferences represented as a vector of types to distinguish literal vs nonliteral usage of language. After ruling out literal usage, the model implements explicit rules of discriminating metaphor vs. metonymy, and metaphor vs. anomaly, by using hand-crafted patterns for metonymy and analogy knowledge base for metaphor. This work is meaningful in that it offers explicit guidelines for identifying relationships between metaphor and metonymy, and metaphor and anomaly, and in that it provides many sentence examples. However, this work uses hand-crafted patterns, which have limitations of extension. Since hand-coding is laborious and expensive, it is difficult to apply to other domains.

Using statistical methods that learn selectional preferences automatically from text can be expanded easily. Krishnakumaran and Zhu (2007), Mason (2004), Shutova et al. (2010), and Huang (2014) inferred selectional preferences using corpus-based approaches for metaphor detection. Krishnakumaran and Zhu (2007), for example, learn selectional preferences from bigram frequencies on the Web. This work targets noun metaphors in three different relations: subject-verb(be)-object, subject-verb(non-be)-object, and adjective-noun. For the first type of metaphor, they determine a metaphor if the subject and the object are not in a hyponym relation in WordNet. For the second and third types of metaphor, they use both the WordNet hierarchy and bigram counts. A noun is identified as metaphorical if neither the noun itself nor its hyponyms or hypernyms co-occur frequently with the verb.

Mason (2004), Shutova et al. (2010), and Huang (2014) also use a statistical method such as the measure of selectional association (SA) proposed by Resnik (1997) to infer selectional preferences. While they take the perspective of metaphor as a violation of selectional preferences, they use it in a slightly different way. These studies focus on the strength of selectional preferences of verbs; verbs with weak selectional preferences are not likely to be metaphorical. Mason (2004) reflects the selectional preference strength of a verb when computing a predicate's salience to a domain. Shutova et al. (2010) and Huang (2014) filter out verbs with weak selectional preferences. In general, using corpus-based methods to learn selectional preferences works well to detect novel metaphors. However, these methods could be limited to detecting conventionalized metaphors since frequently used metaphorical word pairs would co-occur in text.

Very recently, Shutova et al. (2016) demonstrated another approach using violation of local lexical coherence expectation. They capture how source and target domains interact differently in a metaphor than in literal language. The idea is using embeddings to see whether words are from similar source and target domains by computing cosine similarity of the embeddings, assuming that similarity will be lower for metaphorical cases. In addition, they compared phrase embeddings and word embeddings. The embeddings of literal phrases will be more similar to the embeddings of individual words in the phrase than those of metaphorical phrases. They built linguistic and visual embeddings for individual words and then extend the methods to learn embeddings for phrases, using a set of arithmetic operations. Although this approach does not technically measure selectional preference violations, the idea is similar in that metaphorical instances would violate coherence.

Huang (2014) discusses the limitation of selectional preferences for metaphor detection. This study reports the results of metaphor detection by only using semantic outlier word detection and selectional association outlier detection, meaning that they report the result of metaphor detection when considering all the outliers as metaphors. The results show that many portions of outliers are not actually metaphors, indicating the limitation of the selectional preferences approach for metaphor detection. Because of this limitation, selectional preference violation by itself cannot detect all metaphors. However, it is still a very useful indicator in combination with other indicators.

Metaphorical mapping patterns:

One way to extend the selectional preferences violation idea is to investigate which domains are frequently mapped metaphorically, or, put differently, what target and source domains are frequently mapped in metaphors. This line of work is distinguished from looking at metaphor as a selectional preferences violation in that it investigates systematic mappings of domains beyond detecting violation. The approaches are varied based on how to approximate the domains and how to find the mappings and expand them.

Within the approaches that model frequent target and source domain mappings, the CorMet system (Mason, 2004) discovered mappings by finding systematic variation patterns in domain-specific selectional preferences. For example, *pour* frequently selects for *money* in the *FINANCE* domain and for objects of *liquid* in the *LAB* domain. By comparing these two domains, CorMet can infer a metaphorical mapping from money to liquids. In addition, by detecting asymmetric directionality of structure transfer between two concept domains, the system decides source and target domains. For example, the FINANCE domain is identified as target and the LAB domain is identified as source because while verbs in the LAB domain are used in the FINANCE domain, verbs in the FINANCE domain are not frequently used in the LAB domain, as shown in Figure 2.2. The domains are represented using WordNet concepts.

Taking another approach, Hovy et al. (2013) detected metaphors using certain semantic patterns appearing in metaphor manifestations. For example, “sweet” with *food* is literal, but is metaphorical with *people*. By finding these patterns on different levels, they extended the application of this mapping information from a narrow focus on verb relations to other syntactic relations.

Tsvetkov et al. (2013) and Tsvetkov et al. (2014) used semantic categories originating in WordNet to model domain mappings. They also use Vector space word representations to capture more fine-grained mapping patterns, which is the strongest feature type according to their experiments. The studies applied their models to cross-lingual metaphor detection, showing that the ideas can hold in different languages.

While not exactly focusing on finding metaphorical “domain” mappings, Gedigian et al. (2006) also detect metaphors using the idea that there are patterns between predicates and their arguments. They learn metaphorical mapping patterns using a maximum entropy classifier. To represent arguments of a predicate, they use pronoun type, a named entity tag, and WordNet synsets. These categories could be seen as another level of domain representation. Similarly,

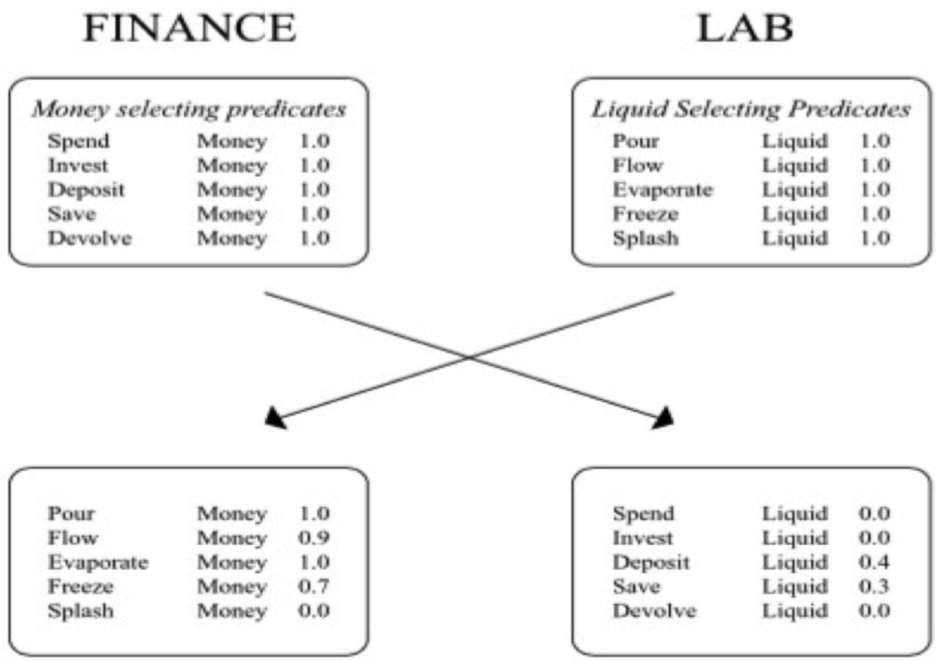

Figure 2.2: Predicates from LAB that select for liquids are transferred to FINANCE and select for money. On the other hand, predicates from FINANCE that select for money are transferred to LAB and do not select for liquids (Mason, 2004)

Birke and Sarkar (2006) proposed a nearly unsupervised clustering for determining literal and nonliteral usages of verbs by formulating the problem as a word disambiguation problem. While this model is not for processing "metaphor" and does not address domain mappings explicitly, the core idea behind this approach is clustering sentences based on their similarities to a literal seed set and nonliteral seed set. This method implicitly exploits the existing mappings appearing in the seed sets.

Other approaches have used clustering for approximating source and target domains. Shutova et al. (2010) identified new metaphors by expanding seed metaphors. The idea in this approach is that target concepts that are frequently used with the same source concept occur in similar lexico-syntactic settings. They cluster nouns (target domain) and verbs (source domain), and search the corpus for metaphors that use the verbs in the source domain lexicon to represent the target domain concepts. Extending Shutova et al. (2010), (Shutova and Sun, 2013) find metaphorical mappings by building and traversing a graph of concepts. Then, they generate lists of salient features for the metaphorically connected clusters, and search the corpus for metaphors that use the verbs in the salient features to represent the target domain concepts.

Li et al. (2013) use patterns obtained from similes as an initial set of metaphors, and expand this set using hyponym and hypernym relations. This study is similar to Krishnakumaran and Zhu (2007) in that it relies on hyponym and hypernym relations, but it goes beyond this by using existing mappings found in similes and expand these mappings using the relations.

As introduced above, there are many different approaches used to approximate source and target domains. Deciding the optimal level of abstraction to represent a domain is a critical issue in the approaches that find domain mappings to generalize metaphorical mechanisms. This is difficult because the semantic representations of a domain need to be general enough to capture patterns in mappings and specific enough to distinguish different mappings. Addressing this question, Bulat et al. (2017) compared different semantic representations to test whether attribute-based semantic representations provide better concept generalizations for metaphor modeling than the widely-used dense linguistic representations. The study results suggested that attribute-based representations provide a suitable means for generalization over the source and target domains in metaphor. This study experimentally shows the importance of attributes of a domain by looking at the metaphorical mappings at the attribute level. This idea is related to the view of this dissertation, which explores how people use metaphors by investigating facets (attributes) of a frame (domain).

	is_loud	has_keys	requires_air	is_long
ACCORDION	6	17	11	0
CLARINET	0	9	0	8
CROCODILE	0	0	0	6

Table 2.2: A subspace of the property-norm semantic space. Attribute-based vectors were built from this resource (Bulat et al., 2017)

These mapping approaches are effective for capturing frequent domain specific metaphorical mappings, and in appropriate contexts are helpful for metaphor detection. Nevertheless, they may overgeneralize in cases where frequent mappings are metaphorical when applied to an extended discourse. In addition, if these approaches consider "what" makes some domains have metaphorical mappings, they may perform better particularly at detecting novel metaphors from novel domains.

Lexical abstractness and concreteness:

Because metaphor often describes abstract concepts in terms of more concrete experiences, we might also imagine some violation of expectation at the level of concreteness of source and target terms in a metaphor. Following this assumption, the final line of work in detecting metaphor in restricted syntactic structures considers the discrepancy of lexical concreteness in source and target domains as a local indicator of metaphor. From the observations that metaphorical words (source domain) tend to use more concrete and imagination rich words than the target domain, the abstractness/concreteness approaches detect metaphors by computationally measuring the degree of abstractness of words. Take the following two phrases as examples that demonstrate this concept: *green idea* (metaphorical expression) and *green frog* (literal expression.) The former has a concrete word (*green*) modifying an abstract concept (*idea*), thus is more likely to be metaphorical.

This idea of leveraging abstractness/concreteness in detecting metaphors has been proposed and studied by several groups of researchers (Assaf et al., 2013; Broadwell et al., 2013; Neuman et al., 2013; Tsvetkov et al., 2013; Turney et al., 2011). Tsvetkov et al. (2014) also used this idea as a feature in her approach. Neuman et al. (2013) Gandy et al. (2013) extended the models by incorporating information about selectional preferences. Beigman Klebanov et al. (2015) also used this lexical concreteness idea to detect metaphors in certain syntactic structures appearing in naturalistic texts.

2.3.1.2 Detecting metaphors in any position within a sentence

While the approaches in the previous section require restricting syntactic structures to test hypotheses in controlled ways, there are approaches that detect metaphors regardless of their syntactic positions in a sentence. However, instead of releasing the syntactic constraint, this line of work put another restriction on the task – it detects whether a sentence contains a metaphor or not, rather than spotting which word in the sentence is metaphorically used.

The core idea in these approaches is that metaphors tend to be found in sentences where both source and target domains appear. Based on this hypothesis, the approaches approximate domains using external resources such as WordNet or Wikipedia articles. The target domain is controlled to test the hypothesis.

Mohler et al. (2013) presented a domain-aware semantic signature to capture source and target domains for a text. A semantic signature represents the placement of a text on a semantic space by using a set of related WordNet senses, and it includes source concept dimensions and target concept dimensions. The primary idea is that the signature of a known metaphor is used to detect the same conceptual metaphor. Heintz et al. (2013) learn LDA topics from Wikipedia, as proxies for concepts that can be potential source and target domains. Then, they use the LDA topics to identify words, which are associated with the source domain and target domain, in a sentence.

These approaches can detect metaphors regardless of their syntactic positions in a sentence. However, because the approaches are based on the strong assumption that both source and target domains should appear in the same sentence, they cannot detect metaphors that do not satisfy the assumption (e.g., when a whole sentence is metaphorical).

2.3.1.3 Detecting metaphors in any position in a wider context

Another line of work on metaphor detection is detecting metaphors regardless of their position. An approach like this might use contextual information that comes from the wider context surrounding the sentence where the metaphors reside. Most approaches that belong to this category of work consider, as a cue for metaphor, lexical coherence violation of the surrounding text. The idea is motivated by the observation that metaphorical words are often semantically incoherent with context words even beyond a sentence.

Several approaches have been proposed to model lexical coherence. Broadwell et al. (2013), for instance, employed topic chaining whereas Sporleder and Li (2009) proposed using lexical chains and semantic cohesion graphs to detect metaphors. Broadwell et al. (2013) and Strzalkowski et al. (2014) have formulated the metaphor detection problem similarly to outlier detection or anomaly detection tasks, and proposed using topic signatures as lexical coherence features. Schulder and Hovy (2014) used TF-IDF to obtain domain term relevance, and applied this feature to detect metaphors. While these lexical coherence violation approaches are effective in capturing metaphorical words from different domains, they are less apt at identifying when related metaphors exist in the surrounding context.

Another approach to detecting metaphors that appear in any position in a wider context is using linguistic cues that suggest the presence of metaphors in text. For example, Goatly (1997) discovered that some words such as *metaphorically speaking* and so to speak signal metaphors nearby. Based on this finding, Shutova (2011) conducted a corpus study to examine whether the contextual linguistic cues are actually effective in metaphor detection using sentences from the BNC corpus. The average precision was 0.40 even though other clues were not used together. This shows the potential that these linguistic cues can help metaphor detection in combination with other features.

Focus of this thesis

In this thesis, I address the problem of detecting metaphors that are positioned in the three dimensions as follows.

- Metaphors that are not restricted in their syntactic positions.
- Metaphors that do not contain clues only within the same sentence.
- Metaphors that can be around related metaphors (extended metaphors).

The computational models for metaphor detection in this thesis are the first tries to model extended metaphors. Although the task in this thesis is not "find all extended metaphors", the models can detect metaphors when extended metaphors are around.

2.3.2 Effect of Metaphor in Discourse

Until now, most computational work on metaphor including metaphor detection has been performed merely at the level of an expression. However, in recent years, there have been novel attempts related to what effect metaphor provides in discourse.

Metaphorically rich language is often considered influential (Strzalkowski et al., 2014). Persuasive metaphors often provoke an attitudinal response in recipients – suggesting a certain perspective, arousing emotions, etc. For example, Stone (1988) described how people use different metaphors about government spending according to their view (EX(14)). Thus, exploring the affective functions of metaphor could be an initial step towards understanding the power of metaphor.

EX(14) A conservative may view government spending on the rich as a "partnership," spending on the middle class as "spending," and spending on the poor as a "giveaway." From a liberal point of view, spending on the rich is a "bailout," spending on the middle class is a "stimulus to the economy," and spending on the poor is a "moral duty." (as cited in (Mio, 1997))

To identify affect carried by metaphors, Kozareva (2013) modeled predicting polarity of metaphor as a classification task and predicting valence of metaphor as a regression problem. The work showed the importance of context in the prediction. Source and target information of a metaphor, and the Linguistic Inquiry and Word Count (LIWC) repository (Tausczik and Pennebaker, 2010) were also useful resources for polarity identification. Strzalkowski et al. (2014) computed affect appearing in metaphors, aiming to isolate the affect of a metaphor from its particular context. They showed that affect tends to be more polarized in metaphors than in literal expressions. These approaches sought to model the affect of metaphor that is considered to play an important role in the power of metpahor as a language tool.

In addition to the affect of metaphor, there have been more direct approaches to finding the *effect* of metaphor on discourse. Beigman Klebanov et al. (2014) examined the effect of metaphor in writing, specifically metaphors used for argumentation in student essays. The study showed a moderate-to-strong correlation between percentage of metaphorically used words in an essay and the writing quality score.

Mohammad et al. (2016) quantitatively studied the extent to which metaphorical language carries emotion. This study examined whether a metaphorical statement is likely to convey a stronger emotional content than its literal counterpart, and how this emotional content arises in the metaphor. For example, it may come from the source domain, the target domain, interaction of the source and the target. According to their study, metaphorical uses of words tend to convey more emotion than their literal paraphrases in the same context. Additionally, the metaphorical sense of a word tends to carry more emotion than the literal sense of the same word.

In addition, there has been an attempt to study the effect of metaphor by modeling extended metaphor using a game theoretic model. Beigman Klebanov and Beigman (2010) observed metaphor use in political communication, specifically, how opposite parties use metaphor as their strategy in turns, focusing on metaphor as a framing device. If one party uses an effective metaphor to maintain their argument, the public become locked in its framework. The other party tries to pick up different aspects of the same metaphor in order to refute the argument of the rival. Beigman Klebanov and Beigman (2010) showed that using an effective metaphor first is an advantage because the rival would be compelled to take up the same metaphor to convince a public that is already locked in the frame of the formerly used metaphor.

In this thesis, our goal is to quantify the motivation behind metaphor use and the effect metaphor has on communication, which is in line with the prior work discussed in this subsection. However, our work is the first attempt to study situational factors that affect metaphors in online discussion forums and clinical interviews in the computational linguistics field. These genres are challenging to study because the text and the speech are much less organized.

2.3.3 Extraction of Properties

So far very little computational work has focused on facets, or properties, of metaphor specifically. However, the Qadir et al. (2016) approach automatically infers implicit properties evoked by similes. They generate candidate properties from different sources using a vehicle and an event. Then, properties are evaluated based on the influence of multiple simile components: using PMI or similarity between a candidate property and the second component of a simile, and aggregate ranking of the properties from different sources. This work is similar to our work in that it extracts properties related to the source domain. However, this work only focuses on similes, which have more formulaic structural patterns compared to metaphors, e.g., *He's as cold as ice*. In addition, the grammatical patterns used in their work are fixed manually by human intuition whereas we automatically infer the patterns in our work.

Chapter 3

Metaphors of Interest Here

To build a computational model of metaphor in discourse that considers why people use metaphor and what effect it has on communication, we first need a corpus that is annotated with metaphors. However, defining the boundary between metaphorical language and literal language is not straightforward. Because metaphors fall on a spectrum, with some being more nonliteral than others, a cut-off point needs to be decided based on the purpose of the annotations. In this chapter, we discuss where to place the cut-off of metaphor on this spectrum of nonliteralness, in order to consider the social aspects of metaphor. The findings in this chapter will form the foundation of all the subsequent chapters in this thesis.

3.1 Introduction

Metaphors can vary in how conventionalized they are, from those which have lost their original concrete meanings (i.e., *dead metaphors*), to completely novel and vivid metaphors. Intuitively, it also makes sense that metaphors which are more conventional and less obviously metaphorical will be used with less conscious thought than more novel or vivid metaphors. There are thus reasons to suspect that distinguishing between levels of metaphoricity could give insight into patterns of use. Therefore, annotation of metaphor draws a line in the metaphor spectrum that spans from literal to metaphorical.

As a first step towards computational work on social and discourse functions of metaphor, this chapter reports a corpus study in three web discussion forums including a breast cancer support group, a Massive Open Online Course (MOOC), and a forum for street gang members, which cover distinctly different domains and have differing community structure. First, we investigate how laypeople intuitively recognize metaphor by conducting Amazon Mechanical Turk (MTurk) experiments. Second, we introduce a new annotation scheme for metaphorical expressions. In our annotation scheme, we try to map the metaphor spectrum of nonliteralness to three types of language: *nonliteral*, *conventionalized*, and *literal*. Next, we compare MTurk results with our annotations expecting that different people will place the dividing line between literal language and metaphorical language in different places. In this work we have the opportunity to gauge the extent to which everyday conceptions of metaphoricity diverge from theoretical perspectives and therefore the extent to which models of metaphoricity may need to be adapted in order to adequately characterize metaphors in strategic use.

The chapter is organized as follows. Section 3.2 describes the data used for this study. Section 3.3 illustrates the functions metaphor serves in discourse through a qualitative analysis of our data. Section 3.4 explains our annotation scheme. Section 3.5 explains the annotation procedures for both our annotation and MTurk annotation, and presents the comparison results from the both annotations. Section 3.6 discusses these results. Section 3.7 concludes the chapter.

3.2 Data

We conducted experiments using data from three different web forums including a Massive Open Online Course (MOOC), a breast cancer support group (Breastcancer), and a forum for street gang members (Gang). We randomly sampled 21 posts (100 sentences) from MOOC, 8 posts (103 sentences) from Breastcancer and 44 posts (111 sentences) from Gang.

We chose these three forums because they all offer conversational data and they all differ in terms of the social situation. The forums differ significantly in purpose, demographics and the participation trajectory of members as shown in Table 3.1. Therefore, we expect that people will use language differently in the three sets, especially related to metaphorical expressions.

	MOOC	Breastcancer	Gang
Data	21 posts 100 sentences	8 posts 103 sentences	44 posts 111 sentences
Purpose	Task-based	Task-based, Social	Social
Participation Trajectory	Participate for a course, leave when the course ends	Participate in the forum after they are diagnosed with cancer, may leave the forum when they recover	No clear endpoint

Table 3.1: Inter-reliability between two trained annotators for our annotation scheme.

MOOC: This forum is used primarily for task-based reasons rather than socializing. People participate in the forum for a course, and leave when the course ends. As a result, the forum does not have continuity over time; participants do not spend long time with the same people.

Breastcancer: People join this forum for both task-based and social reasons: to receive informational and emotional support. People participate in the forum after they are diagnosed with cancer, and may leave the forum when they recover. This forum is also used episodically by many users, but a small percentage of users stay for long periods of time (2 or more years). Thus, continuity allows shared norms to develop over years centered around an intense shared experience.

Gang: In this forum, members belong to a distinct subculture prior to joining, whereas Breastcancer and MOOC members have less shared identity before entering the forum. This forum is purely social. There is no clear endpoint for participation; members leave the forum whenever they are not interested in it any more. Users may stay for a week or two, or for years.

3.3 Qualitative Analysis

Metaphors can be used for a number of conversational purposes such as increasing or decreasing social distance or as a tactic of persuasion or manipulation Ritchie (2013). In this section, we perform a qualitative analysis on how metaphor functions in our data. We illustrate some examples from each domain with an analysis of how some functions of social positioning are observed.

The choice of metaphor may reflect something about the attitude of the speaker. For example, *journey* is a metaphor frequently used in the breast cancer support discussion forum[1] as seen in EX(15) – EX(18) from the Breastcancer forum. People compare chemotherapy to a *journey* by using metaphors such as *journey*, *road* and *moves along*. A *journey* has a beginning and a goal one travels towards, but people may take different paths. This conveys the experience of cancer treatment as a process of progressing along a path, struggling and learning, but allows for each person's experience to differ without judgment of personal success or failure Reisfield and Wilson (2004). By contrast, another common metaphor compares cancer treatment to battles and war. This metaphor instead conveys an activity rather than passivity, a struggle against a defined foe, which can be won if one fights hard enough. But it also creates negative connotations for some patients, as forgoing treatment could then be seen as equivalent to surrender (ibid.).

EX(15) Hello Ladies! I was supposed to start chemo in January, ... I cant start tx until that is done. So I will be *joining you on your journey* this month. I AM SICK OF the ANXIETY and WAITING.

EX(16) So Ladies, please add another member to this club. Looks like we well all be *leaning on* each other. But I promise to *pick you up* if you *fall* if you can *catch* me once in a while!

EX(17) The *road* seems long now but it really *moves along* fast.

EX(18) I split *this journey* into 4 stages and I only deal with one.

In addition, using metaphors can have an effect of increasing empathetic understanding between the participants Ritchie (2013). We can see this in EX(15) – EX(18), where participants in the same thread use similar metaphors relating chemotherapy to a *journey*. Reusing each other's metaphors reduces emotional distance and helps to build empathic understanding and bonding through a shared perception of their situations.

Metaphor also serves to suggest associations between things that one would not normally associate. EX(19) from the MOOC forum frames participation in discussions as stepping into an arena, which refers to an area for sports or competition. By making such an analogy, it conveys an environment of direct competition in front of a large audience. It suggests that a student may be afraid of contributing to discussion because they may make a wrong statement or weak

[1] http:breastcancer.org

argument and another person could counter their contributions, and they will be embarrassed in front of their classmates.

EX(19) Hi, ****, great *point* – I do wish that teachers in my growing up years had been better facilitators of discussion that allowed EVERYone to practice adn become skillful at speaking...I think in the early years some of us need some *handholding* in *stepping into the arena* and speaking

Metaphors can also be used simply as a form of wordplay, to display one's wit and creativity. This can be seen in the exchange in EX(20) – EX(22), from the Gang forum. A common metaphor used on that forum is to refer to someone as *food* to mean that they are weak and unthreatening. The writer in EX(20) expands on this metaphor to suggest that the other person is especially weak by calling him *dessert*, while the writer in EX(21) then challenges him to fight by exploiting the meaning of *hungry* as "having a desire for *food*". The first writer EX(22) then dismisses him as not worth the effort to fight, as he does not *eat vegetables*.

EX(20) So If She Is *Food* That Must Make U *Desert*

EX(21) if u *hungry* nigga why wait?

EX(22) I Dont *Eat Vegatables*.

3.4 Our Annotation Scheme

When we performed qualitative analysis as in Section 3.3, we found that more noticeable metaphors such as "journey", "pick you up", and "fall" in EX(15) and EX(16) seem more indicative of speaker attitude or positioning than metaphors such as "point" in EX(19). This might suggest the degree of metaphoricity affects how metaphors function in discourse. In this section, we describe our metaphor annotation scheme (refer to Appendix A for the full scheme), which tries to map this variation among metaphors to a simpler three-point scale of nonliteralness: *nonliteral*, *conventionalized*, and *literal*.

3.4.1 Basic Conditions

Our annotation scheme targets language satisfying the following three conditions:

1. the expression needs to have an original established meaning.
2. the expression needs to be used in context to mean something significantly different from that original meaning.
3. the difference in meaning should not be hyperbole, understatement, sarcasm or metonymy.

These conditions result in metaphorical expressions including simile and metaphorical idioms. We consider simile to be a special case of metaphor which makes an explicit comparison using words such as "like". We include metaphorical idioms because they are obviously nonliteral and metaphorical despite the fact that they have lost their source domains. For more detailed explanation of each condition, see Section 1.1.

No.	Question	Decision
1	Is the expression using the primary or most concrete meanings of the words?	Yes = L
2	Does the expression include a light verb that can be omitted without changing the meaning, as in “I take a shower” $\rightarrow$ “I shower”? If so, the light verb expression as a whole is literal.	Yes = L
3	Is the metaphor composed of a single compound word, like “painkiller”, used in its usual meaning?	Yes = L
4	Is the expression a conventional term of address, greeting, parting phrase or a discourse marker?	Yes = L
5	Is the expression using terminology or jargon very common in this domain or medium?	Yes = L
6	Is the expression merely hyperbole/understatement, sarcasm or metonymy?	Yes = L
7	Is the expression a fixed idiom like “kick the bucket” that could have a very different concrete meaning?	Yes = N
8	Is the expression a simile, using “like” or “as” to make a comparison between unlike things?	Yes = N
9	Is the expression unconventional/creative and also using non-concrete meanings?	Yes = N
10	Is there another common way to say it that would convey all the same nuances (emotional, etc.)? Or, is this expression one of the only conventional ways of conveying that meaning?	If yes to the latter = C
11	If you cannot otherwise make a decision between literal and nonliteral, just mark it as C.	

Table 3.2: Questions to annotate (N: Nonliteral, C: Conventionalized, L: Literal).

3.4.2 Decision Steps

To apply the basic conditions to the actual annotation procedure, we come up with a set of decision questions (Table 3.2). The questions rely on a variety of other syntactic and semantic distinctions serving as filtering questions. An annotator follows the questions in order after picking a phrase or word in a sentence he or she thinks might be nonliteral language. We describe some of our decisions below.

Unit: The text annotators think might be nonliteral is considered for annotation. We allow a word, a phrase, a clause, or a sentence as the unit for annotation as in (Wallington et al., 2003). We request that annotators include as few words as necessary to cover each metaphorical phrase within a sentence.

Category: We request that annotators code a candidate unit as *nonliteral*, *conventionalized*, or *literal*. We intend the *nonliteral* category to include nonliteral language usage within our scope, namely metaphors, similes, and metaphorical idioms. The *conventionalized* category is intended to cover the cases where the nonliteralness of the expression is unclear because of its extensive usage. The *literal* category is assigned to words that are literal without any doubt.

Syntactic forms: We do not include prepositions or light verbs. We do not consider phrases that consist of only function words such as modals, auxiliaries, prepositions/particles or infinitive markers. We restrict the candidate metaphorical expressions to those which contain content words.

Semantic forms: We do not include single compound words, conventional terms of address, greeting or parting phrases, or discourse markers such as "well". We also do not include terminology or jargon specific to the domain being annotated such as "twilight sedation" in healthcare, since this may be simply borrowing others' words.

3.5 Annotation Procedure

In this section, we present our comparative study of the MTurk annotations and the annotations based on our annotation scheme. The purpose of this experiment is to explore (1) how laypeople perceive metaphor, (2) how valid the annotations from crowdsourcing can be, and (3) how metaphors are different in the three different domains.

3.5.1 Procedure

We had two annotators who were graduate students with some linguistic knowledge. Both were native speakers of English. The annotators were asked to annotate the data using our annotation scheme. We will call the annotators *trained annotators* from now on.

In addition, we used Amazon's Mechanical Turk (MTurk) crowdsourcing marketplace to collect laypeople's recognition of metaphors. We employed MTurk workers to annotate each sentence with the metaphorical expressions. Each sentence was given along with the full post it came from. MTurkers were instructed to copy and paste all the metaphors appearing in the sentence to given text boxes. They were given a simple definition of metaphor from Wikipedia along with a few examples to guide them. Each sentence was labeled by seven different MTurk workers, and we paid $0.05 for annotating each sentence. To control annotation quality, we required that all workers have a United States location and have 98% or more of their previous submissions accepted. We monitored the annotation job and manually filtered out annotators who submitted uniform or seemingly random annotations.

3.6 Corpus Study

3.6.1 Comparison between Annotations from Our Scheme and MTurk

To evaluate the reliability of the annotations, we used weighted Kappa (Cohen, 1968) at the word level, excluding stop words. The weighted Kappa value for annotations following our annotation scheme was 0.52, and the percent agreement was 95.68%. To measure inter-reliability between two annotators per class, we used Cohen's Kappa (Cohen, 1960). Table 3.3 shows the Kappa values for each dataset and each class. Table 3.5 shows the corpus statistics.

To evaluate the reliability of the annotations by MTurkers, we calculated Fleiss's kappa (Fleiss, 1971). Fleiss's kappa is appropriate for assessing inter-reliability when different items are rated

Dataset	N	C	N+C	Weighted
all	0.44	0.20	0.49	0.52
breastcancer	0.69	0.20	0.63	0.71
Gang	0.26	0.28	0.39	0.34
MOOC	0.41	0.13	0.47	0.53

Table 3.3: Inter-reliability between two trained annotators for our annotation scheme (N: Nonliteral, C: Conventionalized).

by different judges. We measured the agreement at the word level, excluding stop words as in computing the agreement between trained annotators. The annotation was 1 if the MTurker coded a word as a metaphorical use, otherwise the annotation was 0. The Kappa values are listed in Table 3.4.

Dataset	Fleiss's Kappa
all	0.36
breastcancer	0.41
Gang	0.35
MOOC	0.30

Table 3.4: Inter-reliability among MTurkers.

Dataset	Posts	Sent.	Words	Content Words	N	C	N/Sent.	C/Sent.
MOOC	21	100	2005	982	23	59	0.23	0.59
Breastcancer	8	103	1598	797	27	41	0.26	0.4
Gang	44	111	1403	519	30	51	0.27	0.46

Table 3.5: Data statistics (N: Nonliteral, C: Conventionalized).

We also measured the agreement between the annotations based on our scheme and MTurk annotations to see how they agree with each other. First, we made a gold standard after discussing the annotations of trained annotators. Then, to combine the seven MTurk annotations, we give a score for an expression 1 if the majority of MTurkers coded it as metaphorically used, otherwise the score is 0. Then, we computed Kappa value between trained annotators and MTurkers. The agreement between trained annotators and MTurkers was 0.51 for N and 0.40 for N + C. We can see the agreement between trained annotators and MTurkers is not that bad especially for N.

Figure 3.1 shows the percentage of words labeled as N, C or L according to the number of MTurkers who annotated the word as metaphorical. As seen, the more MTurkers who annotated a word, the more likely it was to be annotated as N or C by our trained annotators. The distinction between Nonliteral and Conventionalized, however, is a bit muddier, although it displays a moderate trend towards more disagreement between MTurkers for the Conventionalized category. The vast majority of words (>90%) were considered to be literal, so the sample size for comparing the N and C categories is small.

Dataset	N	N+ C
all	0.51	0.40
breastcancer	0.64	0.47
Gang	0.36	0.39
MOOC	0.65	0.36

Table 3.6: Inter-reliability between trained annotators and MTurkers (N: Nonliteral, C: Conventionalized).

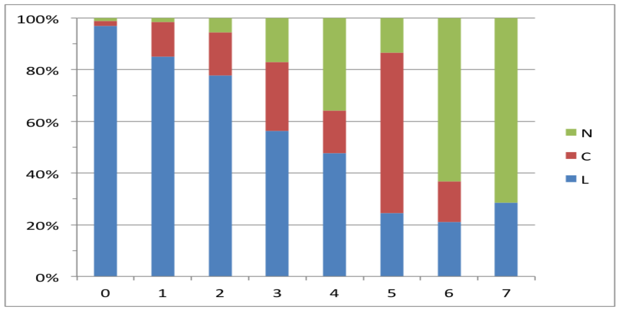

Figure 3.1: Correspondence between MTurkers and trained annotators. X-axis: the number of MTuckers annotating a word as metaphor. Y-axis: the relative percentage of each type.

3.6.2 Discussion

In this section, we investigate the disagreements between annotators. A problem inherent to the annotation of metaphor is that the boundary between literal and nonliteral language is fuzzy. Different annotators may draw the line in different places even when it comes to phrases they are all familiar with. It is also true that each person will have a different life history, and so some phrases which are uninteresting to one person will be strikingly metaphorical to another. For example, someone who is unfamiliar with the internet will likely find the phrase “surf the web” quite metaphorical.

Since we did not predefine the words or phrases that annotators could consider, there were often cases where one person would annotate just the noun and another might include the entire noun phrase. If it was part of a conventional multi-word expression, MTurkers seemed likely to include the entire collocation, not merely the metaphorical part. Boundaries were an issue to a lesser extent with our trained annotators.

One of our datasets, the Gang forum, uses a lot of slang and non-standard grammar and spellings. One of our trained annotators is quite familiar with this forum and the other is not. This was the set they had the most disagreement on. For example, the one annotator did not recognize names of certain gangs and rap musicians, and thought they were meant metaphorically. Similarly, the MTurkers had trouble with many of the slang expressions in this data.

Another issue for the MTurkers is the distinction between metaphor and other forms of nonliteral language such as metonymy and hyperbole. For example, in the Gang data, the term “ass” is used to refer to a whole person. This is a type metonymy (synecdoche) using a part to refer to the whole. MTurkers were likely to label such expressions as metaphor. Hyperbolic expressions like "never in a million years" were also marked by some MTurkers.

In a few cases, the sentence may have required more context to decipher, such as previous posts in the same thread. Another minor issue was that some data had words misspelled as other words or grammatical errors, which some MTurkers annotated as metaphors.

Certain categories of conventionalized metaphors that would be annotated in the original presentation of MIP (Pragglejaz-Group, 2007) were never or almost never annotated by MTurkers. These included light verbs such as "make" or "get" when used as causatives or the passive "get", verbs of sensation used for cognitive meanings, such as "see" meaning "understand", and demonstratives and prepositions in themselves. This may indicate something about the relevance of these types of metaphors for certain applications.

3.7 Conclusion

We annotated data from three distinct conversational online forums using both MTurks and our annotation scheme. The comparison between these two sets of annotations revealed a few things. First, although MTurkers did not show high agreement among themselves, they did show acceptable agreement with trained annotators in the Nonliteral category. Second, we found that domain-specific knowledge was important for accurate identification of metaphors. Even trained annotators had difficulty when they were not familiar with the domain, perhaps because they did not even understand the text's meaning.

We conclude that the metaphors laypeople understand are quite different from the metaphors researchers recognize as cross-domain mappings. MTurkers almost never recognized dead metaphors as metaphor, but showed acceptable agreement in the Nonliteral category of our annotation scheme. Based on these findings, we will annotate metaphors that laypeople recognize, for our subsequent studies.

Part I

Situational Factors and Metaphor Usage

Chapter 4

Case Study 1: Metaphor and Psychological Distress Conditions

In the previous chapter (Chapter 3), we examined metaphor use in three different contexts: a breast cancer support group, a Massive Open Online Course (MOOC), and a forum for street gang members. Through the corpus studies, we observed that people use different metaphors for different purposes in different contexts. In this chapter, we hone in on the metaphor use in one particular context to investigate people's metaphor use more closely – interviews for determining psychological distress conditions such as depression, anxiety, and PTSD. In this context, we are able to study how distress affects people's metaphor usage. In this chapter specifically, we investigate the case of when the source of distress may reside inside oneself.

4.1 Introduction

People tend to use metaphors when it is difficult to describe or explain something with only literal language. For example, metaphor is useful when describing feelings and emotions because they are abstract and subjective, and thus difficult to verbalize for other people. In addition, it may be difficult to share emotions with others depending on one's level of intimacy, especially if the emotion is negative (e.g., complaint, distress, etc.). Metaphors can mask one's personal details. Hence, metaphor is a useful linguistic tool for symbolic representation, providing people with ways to convey emotions without revealing personal details.

The importance of metaphor as a language tool has long generated interest in the role of metaphors in psychotherapy. Researchers have found that patients use metaphors extensively, and looking at these metaphors can increase understandings of patients' psychological states. For example, McMullen (1985) studied the *target* of metaphors. This study revealed that patients who responded to treatment in psychotherapy used metaphors primarily for their own inner thoughts and personal experiences, and other patients who did not used metaphors primarily for external circumstances. On the other hand, Levitt et al. (2000) investigated the *source* of metaphors. They analyzed patients' "burden" metaphors in treatments of depression, and showed that metaphor can be used to track therapeutic change. The results indicate that metaphors of "being burdened" were converted to metaphors of "unloading the burden" in therapy that showed good progress, but not in therapy that showed poor progress. Sharpe (1940) also analyzed the source of patients' metaphors qualitatively, and claimed that metaphors provide rich knowledge of the speaker's suppressed ideas or emotions, or even the speaker's underlying problems. This prior research has indicated that metaphor can be used to reveal someone's inner psychological state.

While previous works usually focused on one psychological condition, this chapter examines how metaphors are related to three different psychological distress conditions: anxiety, depression, and post-traumatic stress disorder (PTSD). These three distinct conditions allow us to not only attribute what we observe to distress, but also allow us to examine the source of the distress by looking at differences across conditions. In addition, our annotation scheme allows us to separate effects on language of the *contextual expectation* (i.e., properties of the interview questions), the *content of the message* (i.e., the target of the metaphors), and the *framing* (i.e., the source of the metaphors). These factors were confounded in the prior analyses. In our analysis, we break these three aspects of language apart, and investigate metaphor usage at these different levels across the three different psychological conditions.

To explore the possibility that metaphor would help assess psychological distress conditions, we propose two research hypotheses based on known characteristics of these psychological conditions. These hypotheses test whether the signs of the distress conditions are observable in the metaphors that interview participants use. The hypotheses are listed below:

H1. Psychologically distressed people use more self-focused metaphors than non-distressed people.

H2. Psychologically distressed people use more negative metaphors than positive metaphors.

First, we hypothesize that distressed people will use more metaphors to describe themselves (or "self-focused" metaphors) since many psychologically distressed people have been associated with self-focused attention. Many forms of psychological distress, from depression (Johnson, 2003) and trauma (Startup et al., 2007) to social anxiety (Spurr and Stopa, 2002; Wells and Papageorgiou, 1998), have been associated with self-focused attentions like self-loathing, isolation, and self-blame. In our study, we investigate if this self-focused trait is observable in participants' metaphor use. Using metaphors to describe oneself, rather than external things, might reveal a self-focused trait.

Second, we hypothesize that distressed people would use more negative metaphors since negative thinking and mood are representative symptoms of psychological distress conditions. Teasdale (1983) discusses the reciprocal relationship between negative cognition and depression, which produces a vicious cycle that aggravates depression. Wells and Papageorgiou (1998) talk about negative belief in anxiety, and Watson et al. (1988) show that negative affect is correlated with both depression and anxiety. Weems et al. (2007) show that pre-disaster negative affect can be a predictor for disaster-related PTSD and anxiety symptoms. Our study investigates if this negativity trait is observable in people's metaphor usage, since metaphors may reflect the way people frame or feel about what they are referring to.

The remainder of the chapter is organized as follows: Section 4.2 describes the data we use for our study. Section 4.3 describes our research hypotheses and annotation scheme for the data. Section 4.4 elaborates the analysis performed on the data. Section 4.5 concludes the chapter.

4.2 Data

Our study is performed on the Distress Analysis Interview Corpus (DAIC), a collection of semi-structured clinical interviews (Figure 4.1) devised to assist in the diagnosis of psychological distress conditions (Gratch et al., 2014). The participants are coded for anxiety, depression, and PTSD based on established psychiatric questionnaires.

Interviewer Let's see, is there someone in your life that's been a really positive influence for you?
Participant Uh yeah, my husband, yeah.
Interviewer Yeah.
Interviewer What kind of values did you take away from him?
Participant Uh he's always uh thinking ahead and looks at the big picture and doesn't uh mull over trivial things so that's something that helped me.
Interviewer Mhm yeah, those are good traits to have.
Participant Yeah, yes.
Interviewer Um how did you guys meet?

Figure 4.1: Sample excerpt from interviews (Gratch et al., 2014).

The interviews used for our study, which are part of the DAIC corpus, were conducted by humans, and participants were interviewed by one of two female interviewers. The interviews took 30-60 minutes and all were conducted in English. Interview questions started with questions that aimed to make the participant comfortable. The interviews ended with a "cool-down" phase that aimed to recover the participant's state from more specific questions about his/her life related to psychological distress. Table 4.1 lists interview questions given to one participant.

This corpus is suitable for studying the relationship between language and psychological distress because we can group people depending on their psychological distress conditions. In addition, several common questions were asked to different people, and these questions have different levels of affective polarity and intimacy. This controlled but naturalistic conversation situation allows us to observe the way people use metaphor potentially differently depending upon their conditions and asked questions.

4.3 Annotation Procedure

Metaphor provides a particular lens through which people can use to interpret and express certain events or objects. This may allow us to glimpse their world view, through their metaphor use. In order to analyze people's metaphor use in relation to psychological distress conditions, we annotate verbal metaphors appearing on their interview transcripts. In this section, we describe how we annotate our data to study our research hypotheses.

#	Question
1	Do you do like a lot of these experiments?
2	Do you have a favorite one that you've done so far?
3	What kind of video games do you like?
4	Would you say that you're an introvert or an or an extrovert?
5	Mhm was there something that like made you change your mind and you're like i should be an extrovert?
6	Um let's see so what you s what did you say that you do now you do gaming right
7	Yeah i can tell you're really passionate <laughs>so how long have you been doing that
8	Are you happier at this one?
9	Can you think of someone that's um been a really strong influence in your life?
10	Can you think of any things that like piss you off or tend to make you angry
11	How do you tend to like handle stress if you get really stressed out about something?
12	That's great do you prefer like the treadmill inside or like going outside and
13	I saw that you also said in the survey that you were like previously diagnosed with depression right how long ago was that
14	Did you feel apprehensive at all about like seeing a therapist before you like went to one
15	If you could change something about yourself what would it be?
16	Has the craving come back to you though in like a midlife crisis at all no
17	Yeah totally do you remember the last time you were upset about something or really angry
18	Let's see when was the last time that you uh got into an argument with like a loved one
19	If you could give yourself advice like ten years ago or maybe even twenty years ago like what what would you have told yourself

Table 4.1: Example of interview questions asked to one participant.

The rest of this section is structured as follows. We begin with selecting particular interview questions for annotation (Section 4.3.1). Next, we annotate metaphors appearing in the interview responses to the selected questions at two levels. In the first stage of the annotation process, we identify metaphors occurring in the interview responses (Section 4.3.2). In the second stage, we annotate the characteristics of the metaphors that are more relevant to our hypotheses, which can reveal more about people's choice of metaphor use, such as what they refer to and what sentiment they convey (Section 4.3.3).

4.3.1 Question Selection

In order to annotate the data, we first selected certain interview questions for which we would annotate the responses. We decided to annotate only part of the data since manual annotation of data takes considerable time and effort. Among 40 main questions we collected from the entire data, we selected 7 questions that would lead to more metaphors, based on selection criteria that consider affective polarity, intimacy, frequency, and the target of the questions.

To collect questions that were commonly asked, we considered questions as main questions only when they have a question ID, and intimacy and polarity ratings. This procedure led us to find 44 types of main questions in total. Among them four types of questions (Table 4.2) were excluded from the main question list because they refer to the previous question (although they have a unique question ID). This left us with 40 main questions. Since responses to the main questions are often too short, we consider all interactions between main questions as belonging to the preceding main question.

#	Question
200	asking for elaboration e.g., you said somewhat so can you elaborate
201	e.g., how do you feel about it?
202	how come?
203	do you remember that do you remember the last time that happened?

Table 4.2: Example of questions that are excluded for annotation.

We used intimacy and polarity annotations of the questions in order to select responses we would annotate. Fainsilber and Ortony (1987) showed that descriptions of feelings contain more metaphorical language than descriptions of behavior. Given the personal nature of our dataset, we expect that questions with high affective polarity and intimacy will lead to more metaphor use. Specifically, we chose questions both with high intimacy and with either high positive or high negative affective polarity. For example, the question "What makes you angry" is considered intimate and negative because it asks more about an interviewees personal experience and feelings, especially negative experience.

In addition, we also considered the frequency of a question as a selection criterion because more responses to the same question would lead to an opportunity for more valuable comparison under the same condition. The average frequency of a question being asked is 17.4 times. Therefore, we selected questions with frequencies higher than the average.

Lastly, we took into account whether the question is targeted only for participants with a specific condition. For example, a question such as “so can you tell me a little more about when you were first diagnosed with PTSD” was excluded since this question had been only asked to participants who were diagnosed with PTSD. These four criteria resulted in the six questions listed in Table 4.3, and 114 interview responses for 80 interviewees.

#	Question
26	if you could think over the last month what was the happiest time the can you think of something that made you really happy
27	was there anyone when you were growing up that was kinda like a positive influence for you or a mentor
23	so after going through all of this like do you ever regret going into the military
22	yeah what about what was the last time that you got in an argument with someone
10	um can you tell me some things that tend to make you angry or piss you off
28	do you remember the last time you felt uh really angry about something

Table 4.3: Example of questions selected for annotation.

4.3.2 Metaphor Identification

To investigate the relationship between people’s metaphor use and their psychological distress conditions, we first identify metaphorically used words appearing in the interview transcripts. In this step, the annotators are asked to find all metaphors in the interviewees’ responses. Since we are interested in exploring people’s choices to use metaphor, we discouraged annotating dead metaphor for which imagery has almost disappeared (as described in Chapter 3). The phrase “a black cloud hanging over his head” is annotated as metaphorical in the example below.

> the veterans have a nickname for me get away from he’s bad luck he’s got <Metaphor>*a black cloud hanging over his head* </Metaphor>and it’s it’s i have to agree with ’em i don’t i i my psychologist said that’s all crazy that’s all nonsense but wherever i turn around something’s always happening to me like uh when the irs came after me

The annotation was performed by two graduate students based on metaphor guidelines described in Chapter 3. The inter-agreement recorded 0.77 in Cohens kappa (Cohen, 1968). Because we study metaphor use, we excluded interview responses that do not contain any metaphors from our dataset after this annotation step. This process resulted in 314 metaphor annotations in total for 41 participants. We excluded participants who did not use any metaphors in their interview. We also excluded the question “do you remember the last time you felt uh really angry about something” from our question list because only participants with no psychological conditions used metaphors. Figure 4.2 shows the distribution of metaphor use in the data. The graph shows that most people only used metaphor a few times, and there are large personal differences in metaphor use, which needs to be considered in further analysis.

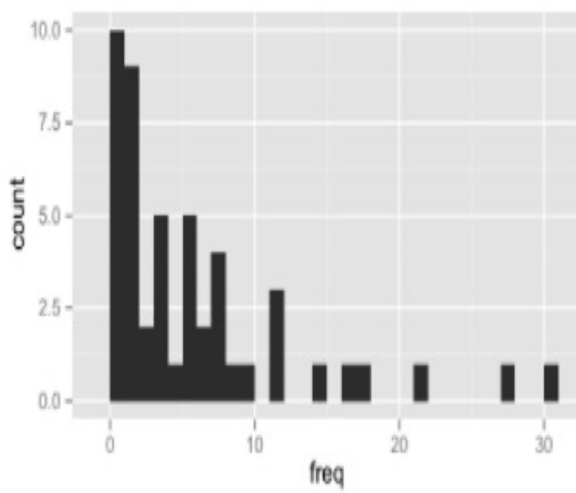

Figure 4.2: Distribution of metaphor use; The x-axis represents the number of metaphors a person used, and the y-axis represents the number of people.

4.3.3 Metaphor Characteristic Annotation

Metaphor represents how people perceive an event or object they refer to. We want to look at the way people use metaphor in their framing potentially differently depending on their psychological distress condition. Our research hypotheses motivate us to annotate characteristics that each metaphor has – mainly what a metaphor refers to, and how it is referred to.

In our annotation scheme, we have four annotation categories: *target*, *target sentiment*, *source sentiment*, and *source sentiment negation*. The "target" category is devised for H1, and the other three categories are for H2.

4.3.3.1 H1: internal vs. external

Psychologically distressed people use metaphors more for speaking in a self-focused way.

To verify this hypothesis, the "target" category annotates what the metaphor refers to at two levels, oneself or other. This annotation specifies whether the speaker describes himself/herself or something external. For example, "a black cloud hanging over his head" in EX(23) describes someone else whereas "hot under the collar" in EX(24) describes the speaker's status.

EX(23) he's got a black cloud hanging over his head.

EX(24) it got me hot under the collar.

4.3.3.2 H2: negative vs. positive

Psychologically distressed people use more negative metaphors than positive metaphors.

To verify this hypothesis, we annotate three categories: "target sentiment", "source sentiment", and "source sentiment negation". The "target sentiment" category annotates the sentiment of what a metaphorical expression refer to. When the metaphorical expression refers to an external thing, we code three scales of "positive", "neutral", and "negative" depending on the sentiment of the external thing. For example, "an angel" in EX(25) can be annotated "neutral" because the metaphor "angel" is describing "my mom", which is neutral without further explanation in its context. When the metaphorical expression refers to oneself, we code "self" referring to the "target" category described above.

The "source sentiment" category annotates the sentiment of a metaphorical expression itself. It has three scales of "positive", "neutral", and "negative". For instance, EX(25) can be annotated positive, EX(26) can be annotated neutral, and EX(27) can be annotated negative.

EX(25) My mom is an angel.

EX(26) when we leave the earth,..

EX(27) i crashed and burned because i was overwhelmed.

The "source sentiment negation" category annotates the sentiment of a metaphorical expression as the 'source sentiment" category, but takes into account whether the expression is negated in the sentence. When the expression is negated, this category annotates the opposite sentiment of the "source sentiment." When the expression is not negated, this category value is the same as the "source sentiment." This category also has three scales of "positive", "neutral", and "negative." For instance, the "source sentiment" of "an angel" in EX(28) is positive, but the "source sentiment negation" becomes negative because it is negated in the sentence. On the other hand, the "source sentiment negation" of "crashed and burned" in EX(30) is positive because the negativity of the expression is negated. However, the "source sentiment negation" of EX(29) is still neutral because we assume that the sentiment of "neutral" is still the same when it is negated.

EX(28) My mom is not an angel.

EX(29) if we do not leave the earth,..

EX(30) i didn't crashed and burned.

4.4 Analysis

The purpose of our analysis is to investigate whether people's metaphor use can be a signal for their psychological distress conditions. Therefore, throughout our analysis, we compare metaphor use between two groups: people who have psychological distress conditions and people who do not. For instance, we compare those with PTSD to those without PTSD, those with anxiety to those without anxiety, and those with depression to those without depression. We also compare those with any of the three conditions to those without any conditions. Note that we do not make a comparison between different distress conditions because most distressed people in this dataset have multiple conditions simultaneously.

In addition, we account for question variables for each analysis. Since we perform our analysis on this interview corpus, people respond to the questions they are asked, which means the responses are affected by the questions as well as people's distress conditions. Hence, we include question variables that describe the questions in each analysis. The question variables include "intimacy" and affective "polarity", as explained in Section 4.3.1 as criteria for selecting questions for annotation. "Intimacy" describes how intimate the question is and "polarity" describes how positive or negative the affect of the question is. Both are continuous numbers.

4.4.1 H1: internal vs. external

To test whether distressed people use more self-focused metaphors, we factor out the effect of the question first. The first logistic regression model has "target" as a response variable, and "intimacy" and "polarity" as explanatory variables. When predicting "target" from question factors of "intimacy" and "polarity", the logistic regression model analysis revealed that "polarity" has a marginally significant effect on Target (p=0.0535). Its log odd-ratio is 0.1414 which means that metaphors refer more frequently to internal things related to themselves when people respond to questions with more positive affect. The most positive questions in our analysis include "if you could think over the last month what was the happiest time, can you think of something that made you really happy?" and "was there anyone when you were growing up that was kinda like a positive influence for you or a mentor?" We observe that participants describe themselves as in EX(31)

EX(31) He is a dog. All men are dogs. I am a dog.

Next, we test the effect of psychological conditions. A logistic regression model analysis was performed for each psychological distress condition to see whether they predict the "target" of metaphors. We controlled for the question variables. However, we found no significant effect of the psychological conditions on the target of metaphors. Thus, we can see that a question has an effect on what people talk about using metaphors, but psychological distress conditions do not.

4.4.2 H2: negative vs. positive

To verify this hypothesis, we first test whether distressed participants use metaphors more frequently to refer to negative things. Then, we test whether they use more negative metaphors when the target sentiment is controlled. In other words, we examine whether distressed people view things from a more negative perspective.

Since question factors might affect what participants talk about through metaphors, we first test whether "intimacy" and "polarity" of a question have an effect on "target sentiment" using a nominal logistic regression model analysis. We found that polarity has a significant effect ($p<0.05$). Table 4.4 displays the model parameter estimate results. Participants's metaphors tend to refer more frequently to negative things when questions with negative affect are given ($p<0.05$, log odd-ratio=-0.3090).

Next, we test whether each distress condition has an effect on "target sentiment" conditioned on "intimacy" and "polarity." We employed a nominal logistic regression model analysis, and found that PTSD has a highly significant effect (p=0.0003) and depression has a marginally significant effect (p=0.0536), according to an effect likelihood ratio test. Table 4.5 displays the model parameter estimate results for PTSD. Participants who have PTSD has a tendency to use metaphors more frequently to refer to themselves ($p<0.005$, log odd-ratio=1.7751), and to refer to negative ($p<0.001$, log odd-ratio=1.9570), compared to neutral things. 4.6 elaborates the model parameter estimate results for depression. Participants who have depression is inclined

Term	Log Odd-Ratio	P-value
intimacy	-1.1227	0.7583
polarity	-0.0948	0.5208
intimacy	-3.2905	0.3719
polarity	-0.3090	0.0432*
intimacy	-1.1354	0.7750
polarity	-0.2092	0.2042

Table 4.4: Nominal logistic regression model of "intimacy" and affective "polarity" of questions on "target sentiment". The odds ratio is defined as P(me) / P(positive) (top three), P(negative) / P(positive) (middle three), and P(neutral) / P(positive) (bottom three).

to use metaphors more frequently to refer to negative things ($p<0.05$, log odd-ratio=1.0718), compared to neutral things.

Term	Log Odd-Ratio	P-value
intimacy	-0.1193	0.9614
polarity	0.0810	0.4783
PTSD	1.7751	0.0014*
intimacy	-2.1278	0.3949
polarity	-0.1308	0.2755
PTSD	1.9570	0.0006*
intimacy	1.4769	0.7156
polarity	0.2050	0.2432
PTSD	1.0831	0.1409

Table 4.5: Nominal logistic regression model of PTSD on "target sentiment" controlled on "intimacy" and "polarity". The odds ratio is defined as P(me) / P(neutral) (top three), P(negative) / P(neutral) (middle three), and P(positive) / P(neutral) (bottom three).

To test whether distressed people frame things more negatively, we perform analyses for two different cases: when "target sentiment" is coded as "self" and when "target sentiment" is coded as "positive", "neutral", or "negative." The first case tests whether distressed participants use negative metaphors when describing themselves. The second case tests whether they use more negative metaphors when referring to external things. These sub-hypotheses are listed below.

H2.1. When psychologically distressed people use metaphors for themselves, they use more negative metaphors.

H2.2. When psychologically distressed people use metaphors for external things, they use more negative metaphors.

For H2.1., we fitted a logistic regression model using a psychological distress condition as a dependent variable, and "source sentiment" as an independent variable. "Source sentiment" did not show any effect on PTSD and depression, but has a relationship with anxiety. The results for

Term	Log Odd-Ratio	P-value
intimacy	0.5487	0.8185
polarity	0.1239	0.2445
Depression	0.7398	0.0903
intimacy	-1.3391	0.5829
polarity	-0.0853	0.4511
Depression	1.0718	0.0170*
intimacy	1.2426	0.7509
polarity	0.2107	0.2016
Depression	0.1158	0.8648

Table 4.6: Nominal logistic regression model of depression on "target sentiment" controlled on "intimacy" and "polarity". The odds ratio is defined as P(me) / P(neutral) (top three), P(negative) / P(neutral) (middle three), and P(positive) / P(neutral) (bottom three).

anxiety are displayed in Table 4.7. Using more negative metaphors for expressing oneself has a positive effect on being diagnosed with anxiety.

Term	Log Odd-Ratio	P-value
source sentiment [negative]	0.5135	0.0329*
source sentiment [positive]	0.0039	0.9885

Table 4.7: Logistic regression model of "source sentiment" on anxiety when "target" is self. The odds ratio is defined as P(anxiety = 1) / 1-P(anxiety = 1). Source sentiment [neutral] is a baseline dummy variable.

We also tested for the effect of "source sentiment negation" on each distress condition, but there were no findings.

For H2.2., we fitted a nominal logistic regression model using "source sentiment" as a dependent variable, and a psychological distress condition as a independent variable conditioned on "target sentiment." However, in this case, psychological distress conditions did not show any effects on "source sentiment".

We additionally tested the effect of "source sentiment negation." We fitted a nominal logistic regression model using "source sentiment negation" as a dependent variable, and kept all other variables the same as the previous analysis for "source sentiment." The results suggested that anxiety and PTSD have significant effects on "source sentiment negation" when "target sentiment" is controlled.

Table 4.8 shows that PTSD has a positive effect ($p<0.05$, log odd-ratio=2.1478) when comparing the ("source sentiment negation" = "negative") case to the ("source sentiment negation" = "positive") case. This indicates that participants who are diagnosed with PTSD use negative metaphors more frequently than non-PTSD participants when describing external things. It also implies that they often use positive metaphors in a negative way by negating the metaphors.

Table 4.9 shows that anxiety has a positive effect ($p<0.05$, log odd-ratio=1.6202) when comparing the ("source sentiment negation" = "negative") case and the ("source sentiment negation"

Term	Log Odd-Ratio	P-value
target sentiment [negative]	1.7135	<.0001*
target sentiment [neutral]	1.1209	0.0314*
PTSD	2.1478	0.0192*
target sentiment [negative]	-0.0557	0.9123
target sentiment [neutral]	2.3825	<.0001*
PTSD	1.2582	0.2171

Table 4.8: Nominal logistic regression model of PTSD on “source sentiment negation” when “target” is others. The odds ratio is defined as P(negative) / P(positive) (top three) and P(neutral) / P(positive) (bottom three). Target sentiment [positive] is a baseline dummy variable.

Term	Log Odd-Ratio	P-value
target sentiment [negative]	1.7014	<.0001*
target sentiment [neutral]	0.9263	0.0678
Anxiety	1.2614	0.0526
target sentiment [negative]	-0.0979	0.8489
target sentiment [neutral]	2.4061	<.0001*
Anxiety	1.6202	0.0262*

Table 4.9: Nominal logistic regression model of Anxiety on “source sentiment negation” when “target” is others. The odds ratio is defined as P(negative) / P(positive) (top three) and P(neutral) / P(positive) (bottom three). Target sentiment [positive] is a baseline dummy variable.

= “neutral”) case to the (“source sentiment negation” = “positive”) case. This indicates that participants who are diagnosed with anxiety use negative or neutral metaphors more frequently than participants without anxiety when describing external things.

4.5 Conclusion

In this chapter, we demonstrated that people’s metaphor usage is influenced by their psychological distress conditions. The analysis revealed differences between distressed and non-distressed people at two levels: how they frame things and what they talk about through metaphors. These results show that distressed participants’ use of metaphor suggest negativity at both levels; they use more negative metaphors when speaking in general, and the topics they discuss metaphorically are more negative.

This study is meaningful in that it supports previous literature that suggested a correlation between distress conditions and negativity in general. However, a more important contribution is that the patterns revealed are slightly different among the three distress conditions, depending on the level of analysis: *contextual expectation*, *content of message*, and *framing*. These three levels affect people’s language choices differently, but are often conflated in the majority of language technologies research as we usually look at features that are related to all of these levels. The findings in this chapter emphasize the importance of separating out the three levels that affect people’s language choices.

Chapter 5

Case Study 2: Metaphor and Stressful Cancer Events

In the previous chapter (Chapter 4), we examined how psychological distress conditions affect people's metaphor usage. Psychological distress conditions such as depression, anxiety, or PTSD can be considered as internal situational factors that could cause people's distress. In this chapter, we discuss how external situational factors could affect metaphor usage. Specifically, we examine the relationship between major cancer events (e.g., diagnosis or chemotherapy) and metaphor use.

5.1 Introduction

Describing an experience metaphorically is an effective conversational strategy for achieving social goals that are relevant within an online medical support community. For example, a metaphor may be useful for drawing the listener closer by revealing not just what has been experienced, but how the speaker is personally engaged with the event, such as *journey* and *battle* Jang et al. (2014). For example, the *journey* metaphor conveys the experience of cancer treatment as a process of progressing along a path in which the cancer patient is a traveler, whereas the *battle* metaphor conveys a more active attitude towards cancer treatment by comparing cancer treatment to conflict and war where the speaker is positioned as a warrior. In this way, metaphors may be used to build solidarity or a sense of camaraderie as they increase insight into the speaker's personal experience and thus facilitate empathetic understanding between the participants Ritchie (2013).

Beyond the social implications of using a metaphor, there are implications at the cognitive level as well. In particular, metaphor is a type of linguistic tool used to express an abstraction. As such, usage of metaphor requires somewhat more cognitive effort than the equivalent literal description. Usage of a metaphor may thus reflect the effort the speaker has invested in making sense out of the associated experience.

Both cognitive and social factors may contribute towards an elevated level of usage of specific metaphors that are associated with the experience of a stressful cancer event in the recent past of a speaker. Specifically, speakers experience a need for more social support during and soon after a stressful event, and thus may engage in behaviors that are useful for building closeness and drawing others in. Additionally, as part of the coping process, experiencers of stressful cancer events are faced with the need to adjust to a new reality after the experience, and this adjustment process may be reflected in linguistic mechanisms that are associated with abstraction and reasoning. Leveraging this insight, we hypothesize that for ambiguous terms (those that can be used either in a literal or metaphorical sense), the concentration of metaphorical use will be elevated within a short window of time following the experience of the associated cancer events. We thus hypothesize that a context variable associated with these events will be a useful clue for increasing accuracy at disambiguating the interpretation of these terms.

In this chapter, we present a corpus analysis of data extracted from an online medical support community, where technology has been deployed to extract mentions of specific cancer events (e.g., diagnosis, chemotherapy, etc.). First, we investigate how popular metaphors we find to be unambiguous in our data from the discussion forum are used in connection with major cancer events. This validates the proposed association between cancer events and metaphor usage. Second, we evaluate the extent to which event information can be helpful for a computational metaphor disambiguation task over more ambiguous candidate metaphor words. In this work, we quantitatively verify the effectiveness of considering situational features in metaphor detection.

The remainder of the chapter is organized as follows. Section 5.2 describes the data used for our experiment. Section 5.3 explains the event extraction method we adopted. Section 5.4 illustrates popular metaphors related to cancer events in our data through a statistical analysis. Section 5.5 presents our metaphor disambiguation experiments. Section 5.6 presents the experiment results. Section 5.7 concludes the paper with a discussion of limitations and next steps in the work.

5.2 Data

We conduct experiments using data from discussion boards for an online breast cancer support group. Participants in the discussion forums are mainly patients, family members, and caregivers. People use the discussion for exchanging both informational support and emotional support with each other by sharing their stories, and through questioning and answering. Some people begin participating in this forum immediately after being diagnosed with cancer, while others do not make their first post until a later event in the cancer treatment process, such as chemotherapy Wen and Rosé (2012).

The data contains all the public posts, users, and profiles on the discussion boards from October 2001 to January 2011. The dataset consists of 1,562,459 messages and 90,242 registered members. 31,307 users have at least one post, and the average number of posts per user is 24.

We picked this dataset for our study of relationship between metaphor and situational factors for two reasons. First, people in this community have a common set of events (e.g cancer diagnosis, chemotherapy, etc.) that are frequently discussed in user posts. Second, people use metaphorical expressions quite frequently in this domain. Thus, the dataset is suitable for a study about metaphor use related with user events. Below is an example post containing metaphors. Some parts in the post have been changed for private information.

> Meghan, I was diagnosed this pst 09/02/07. I was upset for a day when I realized after I had two mammograms and the ultrasound that I had cancer-I didn't have a diagnosis, but I knew. After the ultrasound came the biopsy and then the diagnosis, I was fine. I did research. I made up my mind about what treatement I thought I wanted. I was good...I really was fine up to my visit with the surgeon last week. That made it really real for me. I am waiting for my breast MRI results, and I have to have an ultrasound needle guided auxillary node biopsy before I even get to schedule my surgery. My PET showed other issues in the breast, thus the MRI and the biopsy. Be kind to yourself. It will be a *roller coaster ride* of emotions. Some days really up and strong, other days needing lots of hugs and kleenex. Melody

5.3 Extracting Cancer Event Histories

The cancer events investigated in this paper include *Diagnosis*, *Chemotherapy*, *Radiation Therapy*, *Lumpectomy*, *Mastectomy*, *Breast Reconstruction*, *Cancer Recurrence* and *Metastasis*. All these eight events induce significant physical, practical and emotional challenges. The event dates are extracted from the users' posts as well as the "Diagnosis" and "Biography" sections in their user profiles. 33% of members filled in a personal profile providing additional information about themselves and their disease (e.g., age, occupation, cancer stage, diagnosis date).

We apply the approach of Wen et al. Wen et al. (2013) to extract dates of cancer events for each of the users from their posting histories. A temporal tagger retrieves and normalizes dates mentioned informally in social media to actual month and year referents. Building on this, an event date extraction system learns to integrate the likelihood of candidate dates extracted from time-rich sentences with temporal constraints extracted from event-related sentences.

Wen et al. Wen et al. (2013) evaluate their event extraction approach in comparison with the best competing state-of-the-art approach and show that their approach performs significantly better, achieving an 88% F1 (corresponding to 91% precision and 85% recall) at resolution of extracted temporal expressions to actual calendar dates, and correctly identifies 90% of the event dates that are possible given the performance of that temporal extraction step.

We adopt the same method to extract all users' cancer event dates in our corpus. Note that even were we to use a perfect event extraction system, we can only extract events that the users explicitly mention in their posts. Users may experience additional events during their cancer treatment process, and simply choose not to mention them during their posts.

5.4 Investigation into the Connection between Metaphor and Events

As users continue to participate in the cancer community we are studying, over time they experience more and more significant cancer events. Earlier work Wen and Rosé (2012) shows elevated levels of participation frequency and posting frequency around the time of and immediately after experiencing one of these stress-causing events. This pattern suggests that one way users work to process their traumatic experience is by participating in the forum and obtaining support from other people who are going through similar experiences. Since using metaphorical language suggests elevated levels of cognitive effort related to the associated concept, it is reasonable to expect that users may also engage in a higher concentration of metaphorical language during this time as well as an additional reflection of that processing. In this section, we investigate how the use of metaphor changes with respect to specific traumatic cancer events. We examine a set of common metaphors to see whether situational factors, i.e. cancer events, affect their use. We use cancer event dates extracted in Wen et al. (2013)

5.4.1 Before and After Events

As our first analysis of the relationship between metaphor use and events, we pick eight unambiguous metaphor words in our data – *journey*, *boat*, *warrior*, *angel*, *battle*, *victor*, *one step at a time*, and *roller coaster ride* – and consider the distribution of these metaphors around each event. We categorized these metaphors as unambiguous based on their usage within a small sample of posts we analyzed by hand. Since these are unambiguous, we can be sure that each time we detect these words being used, the speaker is making a metaphor. For each metaphor-event pair, we construct a graph showcasing the frequency of the metaphor usage both before and after the event. We center each user's post dates around the month of the event, so times on the x-axis are relative dates rather than absolute dates (the center of the graph corresponds to the actual event month). The graphs for *journey* and *warrior* paired with the diagnosis event are shown in Figure 5.1 and Figure 5.2, respectively.

Certain metaphor/event pairs show a peak around the event, or at 1 year after the event, for example on the anniversary of diagnosis, which is a significant event in the life of a cancer patient. However, the pattern does not hold across all such pairs, making it difficult to generalize. For example, in Figure 5.1, we see a peak of metaphor frequency occurring at the time of the event, but in Figure 5.2, we do not see such a peak at the time of the event, but see other peaks both before and after the event date. Another complicating factor is that different users experience different cancer treatment timelines. For instance, one user might experience these events over a long period of time, whereas another user may encounter these events in quick succession Wen and Rosé (2012). These factors motivated us to consider other methods, including hierarchical mixed models, for more in-depth analysis.

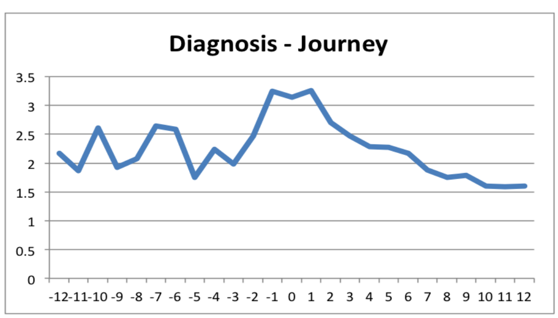

Figure 5.1: Distribution of *journey* metaphor centered around diagnosis event (x-axis: months from event, y-axis: average frequency of metaphor usage)

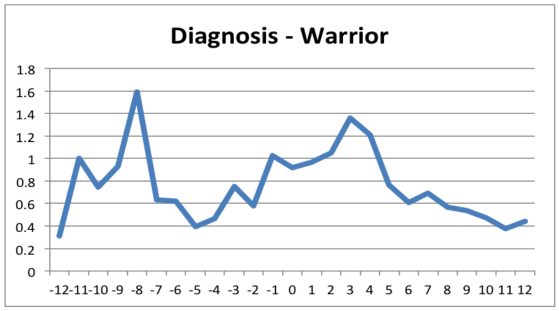

Figure 5.2: Distribution of *warrior* metaphor centered around diagnosis event (x-axis: months from event, y-axis: average frequency of metaphor usage)

metaphor	#		%	
	M	**L**	**M**	**L**
journey	5,787	1,329,560	0.43	99.57
boat	21,398	1,313,849	1.60	98.40
warrior	3,462	1,331,785	0.26	99.74
angel	16,025	1,319,222	1.20	98.80
battle	6,347	1,328,900	0.48	99.52
victor	3,540	1,331,707	0.27	99.73
one step at a time	1,554	1,333,693	0.12	99.88
roller coaster ride	536	1,334,711	0.04	99.96
all	64,755	1,270,492	4.85	95.15

Table 5.1: Corpus-wide unambiguous popular metaphor use statistics (among posts where the user used the metaphor at least once) (**M**: posts that contain each metaphor, **L**: posts that do not contain each metaphor).

5.4.2 Associated Events Analysis

Hierarchical mixed models enable us to model the effect of the experience of a cancer event in the history of a user while controlling for other important factors, such as time and personal tendency. We prepared data for analysis by sampling users. We identified the list of users who used any of our target metaphors at least once, and extracted all the posts of those users. In our models, we treat the message as the unit of analysis, and the dependent measure is always either the presence or absence of a specific metaphor, or the presence or absence of metaphorical language more generally, in all cases indicated by a dichotomous variable. Independent variables including dichotomous indicators of the experience of a specific cancer event in the recent past. We treat each user post as being in the critical period of a cancer event if the post date falls within a time window of two months prior to the event month to two months after the event month, which we selected based on informal observation. Data statistics are shown in Table 5.1.

We tested the association between each dependent variable and the set of independent variables. These hierarchical mixed models were built using the Generalized Linear Latent and Mixed Models (GLLAMM) add-on package in STATA Rabe-Hesketh and Skrondal (2008);

candidate	associated events
journey	diagnosis, recurrence, mastectomy
boat	diagnosis, reconstruction
warrior	chemo
angel	chemo, rads, mets
battle	diagnosis, rads, lumpectomy
victor	chemo, rads, reconstruction
one step at a time	diagnosis
roller coaster ride	diagnosis, reconstruction

Table 5.2: Metaphor candidates and their associated events

Rabe-Hesketh et al. (2004), using maximum likelihood estimation to estimate the models. A random intercept is included for each poster, which is necessary for avoiding obtaining biased estimates of the parameters since there were multiple data points for each user, and users varied in their tendency to use metaphorical language or not. We also experimented with time as an independent variable to control for potential consistent increases in usage of metaphorical language over time, but we did not find any such strong effect, and so we dropped this variable from our models.

We did not find significant effects with a dependent measure that indicated that any of the set of metaphors were used, however, we did find significant associations between metaphors and events when we used dependent variables associated with specific metaphors. Our finding was that the subset of events associated with a metaphor varied by metaphor in a way that made sense given the connotation of the metaphor. For instance, *warrior* is associated with chemo, and *journey* is associated with diagnosis, recurrence, and mastectomy. Associations for all metaphors used for analysis are listed in Table 5.2.

5.5 Experiment on Metaphor Disambiguation

Knowing that there is a significant association between the experience of a cancer event and the usage of a metaphor opens up the possibility for using knowledge of a user's experience of cancer events in the interpretation of their language choices. In particular, if they use a word that may or may not be metaphorical, and the metaphorical usage is associated with a cancer event that occurred in their recent past, then the model should be more likely to predict the metaphorical interpretation. Conversely, if the user is not within the critical period of the event associated with the potential metaphorical interpretation, the metaphorical interpretation should be correspondingly less preferred. We hypothesize that usage of this contextual information might improve the accuracy of disambiguation of potentially metaphorical language. In this section, we test that hypothesis in a corpus based experiment conducted this time on a set of ambiguous, potentially metaphorical words.

5.5.1 Task

Our task is metaphor disambiguation: given a candidate word, decide whether the word is used metaphorically or literally in a post. For example, *road* in EX(32) is used metaphorically, and *road* in EX(33) is used literally. The task is to classify *road* into metaphor and literal use.

EX(32) Great hobbies! ... My hobbie that I love is *road* bike riding. My husband and I both have bikes and we love to ride. ... That's the beauty of living in the south is that you can ride all year long.

EX(33) Another thing to consider is cosmetic outcome. ... If you have a recurrence of cancer and have to do a mast down the *road*, reconstruction is more difficult after having radiation. ...

5.5.2 Data Annotation

We picked six metaphor candidates that appear either metaphorically or literally in the breast-cancer corpus: *candle*, *light*, *ride*, *road*, *spice*, and *train*.

We employed MTurk workers to annotate metaphor use for candidate words. A candidate word was given highlighted in the full post it came from. MTurkers were instructed to copy and paste the sentence where a given highlighted word is contained to a given text box to make sure that MTurkers do not give a random answer. They were given a simple definition of metaphor from Wikipedia along with a few examples to guide them. Then, they were questioned whether the highlighted word is used metaphorically or literally. Each candidate word was labeled by five different MTurk workers, and we paid $0.03 for annotating each word. To control annotation quality, we required that all workers have a United States location and have 98% or more of their previous submissions accepted. We filtered out annotations whose the first task of copy and paste failed, and 18 out of 11,675 annotations were excluded.

To evaluate the reliability of the annotations by MTurkers, we calculated Fleiss's kappa Fleiss (1971). Fleiss's kappa is appropriate for assessing inter-reliability when different items are rated by different judges. The annotation was 1 if the MTurker coded a word as a metaphorical use, otherwise the annotation was 0. The kappa value is 0.80.

We split the data randomly into two subsets, one for analysis of related events, and the other for classification. The former set contains 803 instances, and the latter contains 1,532 instances. The unusual number of instances within each subset arises from the fact that some posts contain multiple metaphors, and we specifically chose to set aside 1,500 posts for classification.

5.5.3 Analysis on Associated Events

We performed a statistical analysis on the six metaphor candidate words as in Section 5.4.2. We combined the users from all the six metaphor candidates, and extracted posts of these users. Independent variables for the model were binary values for each event, where the value is 1 if a

candidate	#		%	
	N	**L**	**N**	**L**
candle*	4	18	18.18	81.81
light	503	179	73.75	26.25
ride	234	185	55.85	44.15
road	924	129	87.75	12.25
spice*	3	21	12.50	87.50
train	94	41	69.63	30.37
all	1762	573	75.46	24.54

Table 5.3: Metaphor use statistics of data used for MTurk (* indicates metaphor candidates for which the literal usage is more common than the non-literal one, **N**: nonliteral use **L**: literal use).

candidate	**associated events**
candle	none
light	diagnosis, rads, mast
ride	diagnosis
road	diagnosis, rads
spice	none
train	mast

Table 5.4: Metaphor candidates and their associated events

post was written in the critical period (defined previously in Section 5.4.2), and 0 otherwise. The dependent variable is a binary value regarding the usage of a metaphor candidate within a post. If a particular post does not include a metaphor candidate or if a post includes a literally used metaphor candidate, the binary dependent value is set to 0. Otherwise, it is set to 1.

The results of conducting the hierarchical mixed model analysis on the data similar to the one conducted above on non-ambiguous metaphors suggest that some candidate words show an association with different cancer events as shown in Table 5.4.

model	Accuracy	Kappa	Precision	Recall	F1 score
(1) word	0.8133	0.3493	0.8105	0.9827	0.8884
(2) context unigram	0.8094	0.4701	0.8651	0.8860	0.8754
(3) context unigram + event	0.8127	0.4777	0.8657	0.8903	0.8778
(4) context unigram + associated event	0.8146	0.4729	0.8612	0.8998	0.8801
(5) context unigram + fs	0.8277	0.5155	0.8731	0.9033	0.8879
(6) **context unigram + event + fs**	**0.8336**	**0.5325**	**0.8772**	**0.9067**	**0.8917**
(7) context unigram + associated event + fs	0.8244	0.504	0.8695	0.9033	0.8861

Table 5.5: Performance on metaphor disambiguation evaluation. (6) is significantly better than (5) [p=0.013] (fs.: used feature selection)

5.5.4 Classification

We used the LightSIDE Mayfield and Rosé (2010) toolkit for extracting features and classification. For the machine learning algorithm, we used the support vector machine (SVM) classifier provided in LightSIDE with the default options. We used basic unigram features extracted by LightSIDE.

To see the effect of event information for classification, we defined two sets of event features. One is a feature vector over all the events, consisting of both binary variables to indicate whether or not a post belongs to the critical period of each event, and numerical variables to indicate how many months the post is written from a known event. We will refer to these features as *event* in Table 5.5. The other is a binary variable to indicate whether or not a post belongs to the critical period of *any* of the associated events for the given metaphor (defined in Section 5.5.3). We will refer to this feature as *associated event* in Table 5.5.

We used multilevel modeling for the features when including *associated event*. We also used the FeatureSelection feature in LightSIDE, where a subset of features is picked on each fold before passing it to the learning plugin. We performed 10-fold cross validation for these experiments.

Because we want to see the effect of event information, we compare our model with a unigram model that uses only the word itself as in Beigman Klebanov et al. (2014), and the context unigram model which uses all the context words in a post as features as baselines.

5.6 Results and Discussion

Table 5.5 displays the results for our experiments. First, we observe the strong performance of the unigram baseline. As in Beigman Klebanov et al. (2014), our evaluation also shows that just using the word currently being classified gives relatively high performance. This result suggests that our candidate words are popular metaphors repeatedly used metaphorically in this domain, as precision is above 80%.

Second, surprisingly, we do not see improvement on accuracy from adding the context words as features. However, we do observe that this addition results in a higher kappa value than just using the candidate words themselves.

Finally, we can see both *event* and *associated event* features show promising results. Both additions give higher result when added to the context unigram model, and the *event* features continue to show improvement when considering models with feature selection. The best model, using *event* features with feature selection, shows significant improvement ($p < 0.05$) over the next best model of context unigram with feature selection.

5.7 Conclusion

In this chapter, we studied how external situational factors affect people's metaphor use. We conducted a study in an online medical support community, which contains a variety of external distress factors (e.g., diagnosis, chemotherapy, etc.). First, we investigated how popular unambiguous metaphors in the discussion forum are used in relation to major cancer events. In addition, we demonstrated that event information can be helpful in a computational metaphor disambiguation task with ambiguous candidate metaphor words. In this work, we quantitatively verified the effect of external situational features.

The major contribution of this work from a computational perspective is to introduce novel types of features for automatic metaphor detection. Metaphor is not a purely linguistic phenomenon, but it is language in action. It can be affected by a variety of situational factors, including the speaker's inner distress, mood, audience, identity, and situational context. Our hope is that this work opens a door for more diverse situational features to be used for metaphor detection, together with linguistically inspired features. In addition, our work reinforces and extends earlier insights into the social and cognitive factors that influence conversational uses of metaphor.

Part II

Metaphor Detection in Discourse

Chapter 6

Metaphor Detection Using Frame Contrast

In the previous chapters (Chapter 4 and Chapter 5), we examined how internal and external situational factors influence people's metaphor usage by building statistical models. From this chapter onward, we describe how we detect metaphors in discourse by using lexical contextual information, based on the framing view on metaphor. In this chapter, we detect frame contrast between a target word and its overall surrounding lexical context.

6.1 Introduction

Detecting metaphors in text is an active line of research which has attracted attention in recent years. To date, most of the previous literature has looked at local information within a sentence, such as selectional restriction violation or domain mapping patterns between verb and its arguments (refer to Chapter 2 for more details). While these approaches have been shown to be successful in detecting metaphors given a single sentence, metaphor detection in discourse brings a new dimension to the task. Consider the following excerpt from an online *Breast Cancer* discussion forum as an example:

> *welcome, glad for the company .... just sad to see that there are so many of us. Here is a thought that I have been thinking since I was diagnosed. This disease should be called the "Hurry up and Wait" illness. Since the day I heard the dreaded words "you need to have a biopsy", I feel like I am on a speeding train. <u>It rushes into every station</u> where you have to make instant decisions, while this ominous clock is ticking. Wait for test results, wait for appointments, wait for healing.*

In the example above, it is difficult to identify *"rushes into every station"* as a metaphorical expression using the previous approaches because it does not violate selectional restrictions or have any notable contrast in lexical concreteness and abstractness. The reason for this is clear: the action of *rushing into stations* itself makes perfect sense literally when it is viewed *locally* as an isolated phrase. However, the contextual cues for this metaphor are embedded globally throughout the discourse (*e.g., diagonsed*, *disease*, *biopsy* are semantically contrasted with *train*, *rushes*, and *station*). This clearly demonstrates the need for a new set of computational tools to represent context beyond a single sentence, in order to better detect metaphorical expressions that have contextual connections outside of the sentence in which they are used.

As we already know, metaphor is a semantic phenomenon that describes objects, often with a view borrowed from a different frame. As such, it is natural that metaphors inherently break the lexical coherence of the current frame. Beigman Klebanov et al. (2009), for example, showed in their study that words related to the topic of discussion are less likely to be metaphorical than other words in text, implying that contextual incoherence might serve as a cue for detecting metaphors. Based on this observation, the idea of leveraging textual context to detect metaphors has been recently proposed by some researchers (Broadwell et al., 2013; Sporleder and Li, 2009).

We extend the previous approaches for detecting metaphors by explicitly addressing the *global* discourse context of a entire text, as well as by representing the *local* context of a sentence in a more robust way. Our contribution in this chapter is thus twofold: first, we propose several textual descriptors that can capture global contextual shifts within a discourse, such as semantic word category distribution obtained from a frame-semantic parser, homogeneity in topic distributions, and lexical chains. Second, we show how global and local contextual information is complimentary in detecting metaphors, and that leveraging syntactic features is crucial for better describing lexico-semantic information in a local context. Our method achieves higher performance on a metaphor disambiguation task than state-of-the-art systems from prior work (Beigman Klebanov et al., 2014; Tsvetkov et al., 2014) on our newly created dataset from an online discussion forum.

The rest of the chapter is organized as follows. Section 6.2 explains our method in detail, specifically in regards to how we use global context and local context for metaphor detection. Section 6.3 describes the *Breast Cancer* dataset annotated and used for our experiment. In Section 6.4, we present our experimental results and show the effectiveness of our method with the task of metaphor disambiguation. Section 6.5 analyzes the results and identifies potential areas of improvement, and we give our concluding remarks in Section 6.6.

6.2 Our Method for Context Frame Representation

In this section, we describe our method to measure nonliteralness of an expression in context. Specifically, we describe how we use contextual information as features for metaphor classification in discourse.

We first define *lexical cohesion* before we introduce our motivation and method for utilizing global contexts as features for detecting metaphor. A text is said to be lexically cohesive when the words in the text describe a single coherent topic. Specifically, lexical cohesion occurs when words are semantically related directly to a common topic or indirectly to the topic via another word. Figure 6.1 illustrates the lexical cohesion among words shown as a graph.

The intuition for our main idea is that metaphorically-used words would often break lexical cohesion of text, while literal expressions would maintain a single connected graph of topically or semantically related words. Therefore, we identify that these incohesive words may serve as cues for nonliteral expressions. The following two examples illustrate the described phenomenon, both of which contain the same phrase *"break the ice"*.

> *... Meanwhile in Germany, the cold penetrated the vast interior of Cologne cathedral, where worshippers had to* <u>*break the ice*</u> *on holy water in the font. The death toll from the cold also increased ...*

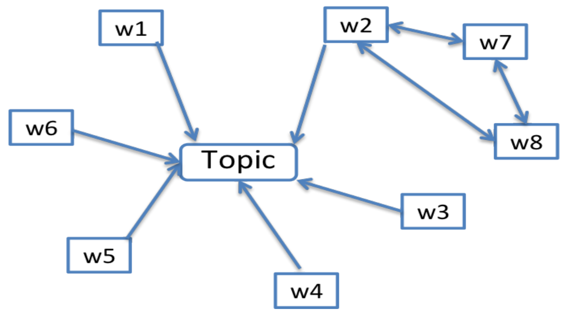

Figure 6.1: Graph representation depicting lexical cohesion among words in a given text. Edges represent lexical relatedness between a topic and a word or between words. For example, w_1 is directly related to the topic of discussion, whereas w_7 is only indirectly related to the topic through w_2.

> *... "Some of us may have have acted as critics at one point or another, but for the most part its just as film-goers," he said. And, breaking the ice at a press conference, he praised his vice-president, French actress Catherine Deneuve ...*

The phrase *"break the ice"* in the first example is used with words such as *cold* and *water* which are semantically coherent with its literal meaning, whereas in the second example, the phrase is used with *press conference*, *praised*, and *vice-president*, which are far from the literal meaning of *break* and *ice*.

Note that this contextual information lies in different parts of a discourse, sometimes locally in the same sentence as the target word or globally throughout multiple surrounding sentences in a discourse. Given this observation, we categorize contextual information into two kinds depending on the scope of the context in text: *global* and *local*. Global contexts range over the whole document, whereas local contexts are limited to the sentence that contains the expression of interest. Section 6.2.1 explains how we represent global contexts. Section 6.2.2 describes the features we use for local contexts, and how we leverage syntactic information to make a more robust use of the semantic features in local context.

6.2.1 Global Contextual Features

We use the following features to represent global contexts of a given text.

Semantic Category: Lexico-semantic resources (*e.g.,* `FrameNet`, `WordNet`) provide categorical information for much of the English lexicon. If a target word is used literally, the document may have a high proportion of words in the same semantic category. If the word is used metaphorically, the document may contain more words that share different semantic categories. To implement this intuition, we use `SEMAFOR` (Das et al., 2014) to assign each word to one of the categories provided by the `FrameNet 1.5` taxonomy (Baker et al., 1998). Then, we compute the relative proportion of the target word's category with regards to categories appearing in the document to measure the alignment of categories of the target word and the surrounding contexts. Formally, we define the value of the *global word category feature* as

$$\frac{\sum_{w \in d} \mathbb{1}(c_w = c_{tw})}{N_d},$$

where c_w is the category of word w, c_{tw} is the category of the target word, and N_d is the number of words in document d. $\mathbb{1}(.)$ is an indicator function that equals 1 when the expression inside is true and 0 otherwise.

We have also used `WordNet`[1]'s 44 lexnames in our preliminary experiment to obtain word categories. However, we have found that its coarse categorization of words (44 categories as opposed to `FrameNet`'s 1204) led to poorer performance, thus we have used `FrameNet` here instead.

Topic Distribution: Our intuition for using topic distributions is that non-literal words tend to have a considerably different topic distribution from that of the surrounding document (global context). To implement this idea, we run a topic model to obtain a word-topic distribution ($=$ P(topic|word)) and document-topic distribution ($=$ P(topic|document)). We use Latent Dirichlet Allocation (LDA) (Blei et al., 2003) to find 100 topics from the entire corpus, and calculate the topic distribution per document and the topic distribution per word from the trained topic model. Specifically, we begin by training our model for 2,000 iterations on a large data set. Then, for the estimation on test documents we apply this model to our test data set for 100 iterations of Gibbs sampling.

The original LDA computes P(word|topic) instead of P(topic|word). In order to compute P(topic|word), the first 20 iterations out of 100 are used as a burn-in phase, and then we collect sample topic assignments for each word in every other iteration. This process results in a total of 40 topic assignments for a word in a document, and we use these topic assignments to estimate the topic distributions per word as in (Remus and Biemann, 2013). We used the `GibbsC++` toolkit (Phan and Nguyen, 2007) with default parameters to train the model.

Finally, we use the cosine similarity between P(topic|document) and P(topic|word) as features that represent the global alignment of topics between the target word and the document.

Lexical Chain: We use *lexical chains* (Morris and Hirst, 1991) to obtain multiple sequences of semantically related words in a text. From the intuition that metaphorical words would not belong to dominant lexical chains of the given text, we use the lexical chain membership of a target word as a cue for its non-literalness. Because each discourse instance of our dataset tends to be short and thus does not produce many lexical chains, we use a binary feature of whether a target word belongs to the longest chain of the given text. In our implementation, we use the `ELKB` toolkit (Jarmasz and Szpakowicz, 2003) to detect lexical chains in text which is built on `Roget's thesaurus` (Roget, 1911). Note that a similar approach has been used by Sporleder and Li (2009) to grasp topical words in a text.

Context Tokens: In addition, we use unigram features to represent the global context. Specifically, we use binary features to indicate whether the context words appeared anywhere in a given discourse.

[1]https://wordnet.princeton.edu/man/lexnames.5WN.html

6.2.2 Local Contextual Features

The local contextual information within a sentence is limited because it often contains fewer words, but the information could be more direct and richer because it reflects the immediate context of an expression of interest. We represent local contextual information using the semantic features listed below, combined with grammatical dependencies to induce relational connections between a target word and its contextual information.

Semantic Category: We follow the same intuition as using semantic categories to represent global features (Section 6.2.1), and thus compare the target word's semantic category and that of other words in the same sentence to induce local contextual information. However, since a sentence often has only a small number of words, the proportion of the target word's category in one sentence depends too much on the sentence length. Therefore, we instead look at the words that have dependency relations with the target word, and create nominal features by pairing word categories of lexical items with their dependency relations. The paired dependency-word category features specifies *how* local contextual words are used in relation to the target word, thus providing richer information. We also specify the target word's category as a categorical feature, expecting that the interplay between the target word's category and other words' categories is indicative of the non-literalness of the target word.

Semantic Relatedness: If the semantic relatedness between a target word and the context words is low, the target word is likely to be metaphorically used. From the observation that the words that are in grammatical relation to the target word are more informative than other words, we use the dependency relations of a target word to pick out the words to compute semantic relatedness with. To represent the semantic relatedness between two words, we compute the cosine similarity of their topic distributions.

We use the semantic relatedness information in two different ways. One way is to compute average semantic relatedness over the words that have dependency relations with a target word, and use it as a feature (AvgSR). The other is to use semantic relatedness of the words in grammatical relations to the target word as multiple features (DepSR).

We use the same techniques as in Section 6.2.1 to compute topic distribution using an LDA topic model.

Lexical Abstractness/Concreteness: People often use metaphors to convey a complex or abstract thought by borrowing a word or phrase having a concrete concept that is easy to grasp. With this intuition, Turney et al. Turney et al. (2011) showed that the word abstractness/concreteness measure is a useful clue for detecting metaphors.

To represent the concreteness of a word, we used Brysbaert's database of concreteness ratings for about 40,000 English words (Brysbaert et al., 2014). We use the mean ratings in the database as a numerical feature for the target word. In addition, we also use the concreteness ratings of the words in grammatical relations to the target word as local context features.

Grammatical Dependencies: We use the `stanford-corenlp` toolkit (Manning et al., 2014) to parse dependency relations of our data and apply grammatical dependencies as described above for each semantic feature. We use grammatical dependencies only between content words (*e.g.,* words with syntactic categories of noun, verb, adjective, and adverb).

6.3 Data

We conduct experiments on data acquired from discussion forums for an online breast cancer support group. The data contains all the public posts, users, and profiles on the discussion boards from October 2001 to January 2011. The dataset consists of 1,562,459 messages and 90,242 registered members. 31,307 users have at least one post, and the average number of posts per user is 24.

We built an annotated dataset for our experiments as follows. We first picked seven metaphor candidates that appear either metaphorically or literally in the *Breast Cancer* corpus: *boat*, *candle*, *light*, *ride*, *road*, *spice*, and *train*. We then retrieved all the posts in the corpus that contain these candidate words, and annotated each post as to whether the candidate word in the post is used metaphorically. When the candidate word occurs more than once in a single post, all occurrences within a post were assumed to have the same usage (either metaphorical or literal).

Note that our annotation scheme is different from the VU Amsterdam metaphor-annotated dataset (Steen et al., 2010) or the essay data used in (Beigman Klebanov et al., 2014), where every word in the corpus is individually labeled as a metaphor or a literal word. Our approach of pre-defining a set of metaphor candidate words and annotating each post as opposed to every word has several practical and fundamental benefits. First, metaphors often have a wide spectrum of “literalness” depending on how frequently they are used in everyday text, and there is a continuing debate as to how to operationalize metaphor in a binary decision (Jang et al., 2014). In our work, we can circumvent this metaphor decision issue by annotating a set of metaphor candidate words that have a clear distinction between metaphorical and literal usages. Second, our annotation only for ambiguous words ensures to focus on how well a model distinguishes between metaphorical and literal usage of the same word.

We employed Amazon Mechanical Turk (MTurk) workers to annotate metaphor use for candidate words. A candidate word was highlighted in the full post it originated from. MTurkers were asked to copy and paste the sentence where a highlighted word is included to a given text box to make sure that MTurkers do not give a random answer. We gave a simple definition of metaphor from Wikipedia along with a few examples to instruct them. Then, they were asked whether the highlighted word is used metaphorically or literally. Five different MTurk workers annotated each candidate word, and they were paid $0.03 for annotating each word. For annotation quality control, we requested that all workers have a United States location and have 98% or more successful submissions. We excluded annotations for which the first task of copy and paste failed. 18 out of 13,348 annotations were filtered out in this way.

To evaluate the reliability of the annotations by MTurkers, we calculated Fleiss’s kappa (Fleiss, 1971), which is widely used to evaluate inter-annotators reliability. Using a value of 1 if the MTurker coded a word as a metaphorical use, and a value of 0 otherwise, we find kappa value of 0.81, suggesting strong inter-annotator agreement.

We split the data randomly into two subsets, one as a development set for observation and analysis, and the other as a cross-validation set for classification. The development set contains 800 instances, and the cross-validation set contains 1,870 instances. Table 6.1 shows the metaphor use statistics of the annotated data.

candidate	#		%	
	N	**L**	**N**	**L**
boat*	54	281	16.12	83.88
candle*	4	18	18.18	81.82
light	503	179	73.75	26.25
ride	234	185	55.85	44.15
road	924	129	87.75	12.25
spice*	3	21	12.50	87.50
train	94	41	69.63	30.37
all	1816	854	68.01	31.99

Table 6.1: Metaphor use statistics of data used for MTurk (* indicates metaphor candidates for which the literal usage is more common than the non-literal one, **N**: nonliteral use **L**: literal use).

6.4 Evaluation

We evaluate our method on a metaphor disambiguation task detailed in Section 6.4.1. Section 6.4.2 lists the metrics we used for the evaluation on this test set. Section 6.4.3 describes the baselines we compare our method against on these metrics. We detail our classification settings in Section 6.4.4 and report our results in Section 6.5.1.

6.4.1 Task

The task for our experiment is metaphor disambiguation: given a candidate word, decide whether the word is used as a metaphor or as a literal word in a post. For example, *boat* in EX(34) is used metaphorically, whereas *boat* in EX(35) is used literally. The task is thus to classify each of the seven candidate metaphors defined in Section 6.3 into either a metaphor or a literal word.

EX(34) *Just diagnosed late November. Stage I and with good prognosis. ... Now I am having to consider a hysterectomy and am really scared and don't know what to do. I have no children and don't really think I want to. I really want to do what is best for me but it is so hard to know. Anyone else been in the same <u>boat</u> with the endometriosis?*

EX(35) *Good Morn Girls, It is 52 this morn. WOW! there is a bad storm rolling in at this time and tornado watches but those are pretty common. ... Hubby started his truck driving school today. We use to have ski <u>boats</u> so he and I could both drive a semi. Backing is the hardest part cause the trailer goes opposite of the direction you turn but once you get use to it, it's not hard. ...*

6.4.2 Evaluation Metrics

We report four evaluation metrics: accuracy, precision, recall, and F-score.

Accuracy: Accuracy is the percentage of correctly classified instances among all instances.

Precision: Precision is the percentage of correctly classified instances among instances assigned to a particular class (metaphor or literal) by the model.

Recall: Recall is the percentage of correctly classified instances among all nonliteral or literal instances. Precision and recall are recorded for both metaphorical and literal labels.

F-score: F-score is the harmonic mean of precision and recall.

6.4.3 Baselines

We compare our method to a context unigram model as well as two other baselines from recent work on metaphor detection: Beigman Klebanov et al. (2014), and Tsvetkov et al. (2014).

Context unigram model uses all the context words including the target word in a post as features.

Tsvetkov et al. (2014) use local contextual features (such as abstractness and imageability, supersenses, and vector space word representations), and targets for two syntactic constructions: subject-verb-object (`SVO`) and adjective-noun (`AN`) tuples. Note that the output of this system is a sentence level label rather than a word (*e.g.,* they output a binary label that indicates whether the target sentence contains any metaphorical phrase). Thus, we take the output of their sentence level label on the sentence that contains our target word, and treat their output as a label for our target word disambiguation task. Although it is therefore not a fair comparison, we included this system as a baseline because this is a state-of-the-art system for metaphor detection tasks. In addition, we can make this comparison to contextualize results with regards to how a state-of-the-art non-discourse model (i.e. not using global context) will perform in more general discourse contexts.

Beigman Klebanov et al. (2014) use target word lexical features such as part-of-speech tags, concreteness rating, and topic score. Their approach does not use any contextual information as our method does. As a result, the same words are most likely to obtain the same features. Note that Beigman Klebanov et al. (2014) evaluated their approach for each content word in a given text, but in our paper we evaluate how their method performs on ambiguous words in particular.

6.4.4 Classification

We used the `LightSIDE` toolkit (Mayfield and Rosé, 2010) for extracting features and performing classification. For the machine learning algorithm, we used the logistic regression classifier provided by `LightSIDE` with L_1 regularization. We used basic unigram features extracted by `LightSIDE`, and performed 10-fold cross validation for the following experiments. Instances for each fold were randomly chosen.

6.5 Results and Discussion

6.5.1 Results

The classification results on the *Breast Cancer* corpus are shown in Table 6.2 and in Table 6.3.

Type	Model	A	P-M	R-M	P-L	R-L	F1
Baseline	T	0.245	0.857	0.168	0.236	0.991	0.207
	K	0.833	0.830	0.984	0.866	0.340	0.694
	U	0.836	0.867	0.929	0.697	0.535	0.751
Global	U+GWC	0.842	0.869	0.934	0.716	0.541	0.759
	U+GT*	0.843	0.873	0.931	0.711	0.557	0.763
	U+LC	0.839	0.866	0.934	0.709	0.530	0.753
	U+GWC+GT+LC*	0.845	0.871	0.936	0.724	0.546	0.762
Local	U+LWC	0.849	0.874	0.939	0.735	0.557	0.634
	U+SR(AvgSR)	0.852	0.873	0.965	0.563	0.243	0.628
	U+SR(DepSR)	0.858	0.880	0.943	0.756	0.580	0.783
	U+AC	0.853	0.880	0.936	0.735	0.582	0.778
	U+LWC+SR+AC*	0.862	0.885	0.942	0.759	0.598	0.791
Global+Local	ALL*	0.860	0.882	0.943	0.761	0.589	0.788
	ALL-LC*	0.863	0.886	0.941	0.759	0.605	0.793

Table 6.2: Performance on metaphor disambiguation evaluation. (**Models**) T: Tsvetkov et al. (2014), K: Beigman Klebanov et al. (2014), U: context unigram, GWC: global word category, GT: global topic dist., LC: lexical chain, LWC: local word category, SR: semantic relatedness, AC: abstractness/concreteness. (**Metrics**) A: accuracy, P-M: precision on metaphors, R-M: recall on metaphors, P-L: precision on literal words, R-L: recall on literal words, F1: Average F1 score over M/L., *: statistically significant improvement over baselines

Target word	A	P-M	R-M	P-L	R-L	F1
boat	0.843	0.886	0.935	0.500	0.351	0.843
light	0.831	0.857	0.920	0.738	0.594	0.773
ride	0.843	0.847	0.888	0.836	0.782	0.838
road	0.926	0.936	0.983	0.823	0.543	0.806
train	0.711	0.759	0.887	0.429	0.231	0.559

Table 6.3: Performance on metaphor disambiguation task per target word with the best setting ALL-LC. Note that the performance results on target words *candle* and *spice* are not reported because of their small number of instances.

Note that both our global context features (*e.g.,*U+GWC+GT+LC, U+GT) and local context features (*e.g.,*U+LWC+SR+AC) perform significantly better than all of the baselines ($p < 0.05$). This indicates that our contextual features successfully capture additional information from discourse both locally and globally. In general, it can be seen that local features are more powerful indicators of metaphors than global features. Note also that Tsvetkov et al. (2014) performs poorly on this task, probably due to the reasons mentioned in Section 6.4.3. It is interesting to note that Beigman Klebanov et al. (2014) performs poorly at recall on literal words. We conclude that our methods significantly outperform the baselines in detecting metaphors in discourse.

6.5.2 Discussion

The results of our methods on the metaphor disambiguation task are promising, indicating that both global features and local features can serve as strong indicators of metaphor.

Note that the combined global+local features did not show significant improvement over the local features on this task in Table 6.2. We had believed that local and global features (aside from unigram features) would provide synergistic predictions, however we found that the local features provided stronger predictions and drowned out the effect of the global features.

We identify the following possible sources of errors of our method: first of all, the low performance of lexican chain (LC) features is noticeable. This might be due to errors originating from the output of the `ELKB` toolkit which we employ to obtain lexical chains. More specifically, `ELKB` builds lexical chains using a standard thesaurus, which is extremely vulnerable to noisy text such as our online discussion forum (which contains typos, abbreviations, medical terms, etc.).

Secondly, the semantic relatedness scores obtained from LDA gives high scores to frequently co-occurring words, thus inevitably reducing effectiveness in disambiguating frequently used metaphors. While this is an issue inherent in any distributional semantics approach, we find that our LDA-based features do improve overall performance.

6.6 Conclusion

We identified that both global and local contextual features can serve as powerful indicators of metaphor, and proposed several methods to represent contextual features in discourse. We also extended previous literature that considers local contextual information by explicitly incorporating the syntactic information, such as dependency relations, into local contextual features, resulting in an improved performance. The performance was evaluated on our newly built *Breast Cancer* dataset, which provides examples of metaphors in a discourse setting. We showed that our method significantly outperforms the systems from recent literature on a metaphor disambiguation task in discourse. Our method can be easily applied to disambiguate all the content words in text once we have correspondingly labeled data.

Chapter 7

Metaphor Detection Using Frame Sentence-Level Transition

In the previous chapter (Chapter 6), we identified metaphors by detecting frame contrast between a target word and its global and local contexts. In this chapter, to exploit a more detailed level of contextual information, we specifically look at the topic of a sentence that contains a target word and how the topic changes around the target word at a sentence level.

7.1 Introduction

Our earlier computational model detected topical discrepancy between a target word and the dominant theme of its lexical context. This approach was effective at capturing some aspects of the governing context of a metaphor, but it tended to overclassify literal words as metaphorical if it found semantic contrast with the governing context. This manifested as high recall but low precision for metaphorical instances.

We now present an approach that uses a more fine-grained level of lexical and topical context to address the problem of low precision on metaphor detection. To better capture the relevant context surrounding a metaphor, we hypothesize that topic transition patterns between metaphorical sentences and their contexts are different from that of literal sentences. To this end, we incorporate several indicators of sentence-level topic transitions as features, such as topic similarity between a sentence and its neighboring sentences, measured by Sentence LDA. This updated metaphor detection approach expands upon our previous approach by exploring topic transitions between a metaphor and its context, rather than only detecting lexical discrepancies.

7.2 Our Approach

To better capture the distinctions between metaphorical and literal usages of the same word (target word), we approach the task in two directions. First, we model how topics in context change for both metaphorical and literal instances of a target word (Section 7.2.1). Second, We use multi-level modeling to combine these features with the specific target word to model interactions between the features and a particular metaphor (Section 7.2.2).

7.2.1 Topic Transition

In writing, cohesion refers to the presence or absence of explicit cues in the text that allow the reader to make connections between ideas (Crossley and McNamara, 2010). For example, overlapping words and concepts between sentences indicate that the same ideas are being referred to across these sentences. Metaphorically used words tend to be semantically incohesive with the governing context. Therefore, determining semantic or topical cohesion is important for metaphor detection.

However, even if a text is literal and cohesive, not all words within the text are semantically related. In EX(36), a human could easily determine that "pillows", "music", "flickering candles", and "a foot massage" share the theme of relaxation. But it is difficult to define their relatedness computationally – these terms are not synonyms, hypernyms, antonyms, or in any other well-defined lexical relation. Additionally, even if the whole sentence is correctly interpreted as ways of indulging oneself, it is still semantically contrasted with the surrounding sentences about medicine. In this example, the target word "candle" is used literally, but the contrast between the sentence containing the target word and its context makes it computationally difficult to determine that it is not metaphorical:

EX(36) ... yet encouraged to hear you have a diagnosis and it's being treated. Since you have to give up your scented stuff you'll just have to figure out some very creative ways to indulge yourself. *Soft pillows, relaxing music, flickering <u>candles</u>, maybe a foot massage*. Let's hope your new pain relief strategy works and the Neulasta shot is not so bad . I never had Taxotere, but have read it can be much easier than AC for many people. ...

EX(37) also shows semantic inconsistency between the candidate metaphor "boat" and the surrounding sentences about medicine. However, in this example, "boat" is metaphorically used. Thus, it is difficult to determine whether a word is metaphorical or literal when there is semantic contrast because both EX(36) and EX(37) show semantic contrast.

EX(37) When my brain mets were discovered last year, I had to see a neurosurgeon. He asked if I understood that my treatment was pallative care. *Boy, did it rock my <u>boat</u> to hear that phrase!* I agree with ****, pallative treatment is to help with pain and alleviate symptoms.....but definitely different than hospice care.

The primary difference between these two examples is in the nature of the semantic contrast. In EX(36), the topic of the sentence containing "candle" is relaxation, while the topic of the previous and following sentences is medicine. The transition between medicine and relaxation tends to be more literal, whereas the transition between the topic in the sentence containing "boat" and the surrounding medical topic sentences tends to be more metaphorical.

We use these differences in the topic transition for metaphor detection. We consider topic transitions at the sentence level, rather than the word level, because people often represent an idea at or above the sentence level. Thus, topic is better-represented at the sentence level.

To model context at the sentence level, we first assign topics to each sentence using Sentence Latent Dirichlet Allocation (LDA) (Jo and Oh, 2011). Sentence LDA has two main advantages over standard LDA for our work. First, while standard LDA assumes that each word is assigned a topic derived from the topic distribution of a document, Sentence LDA makes the constraint that all words in the same sentence must be assigned the same topic. Due to this property, the generated topics are better aligned with the role or purpose of a sentence, compared to topics generated from LDA. Additionally, having each sentence assigned to one topic helps us avoid using heuristics for representing the topic of each sentence. [1]

Using Sentence LDA, we modeled four features to capture how the topic changes around the sentence where a target word resides. We refer to this sentence as the *target sentence*.

Target Sentence Topic (TargetTopic): We hypothesize that sentences containing a metaphor may prefer topics that are different from those of sentences where the same word is used literally. Hence, TargetTopic is a T-dimensional binary feature, where T is the number of topics, that indicates the topic assigned to the sentence containing the target word.

Topic Difference (TopicDiff): We hypothesize that a metaphorical sentence is more likely to be different from its neighboring sentences, in terms of topic, than a literal sentence. Therefore, TopicDiff is a two-dimensional binary feature that indicates whether the topic assigned to the target sentence is different from that of the previous and next sentences.

Topic Similarity (TopicSim): Under the same hypothesis as TopicDiff, TopicSim is a two-dimensional feature that represents the similarity between the topic of the target sentence and its previous and next sentences. Unlike TopicDiff, which is binary, TopicSim has continuous values between 0 and 1, as we use the cosine similarity between each topic's word distributions as topic similarity. Note that in Sentence LDA, all topics share the same vocabulary, but assign different probabilities to different words as in LDA although all tokens in a sentence are assigned to the same topic in Sentence LDA.

Topic Transition (TopicTrans): The topic of a metaphorical sentence may extend over multiple sentences, so a topic transition may occur a few sentences ahead or behind the target sentence. TopicTrans looks for the nearest sentences with a different topic before and after the current target sentence and encodes the topics of the different-topic sentences. Hence, TopicTrans is a $2T$-dimensional feature, where T is the number of topics, that indicates the topics of the nearest sentences that have a different topic from the target sentence.

Topic Transition Similarity (TopicTransSim): The topics before and after a transition, even in the extended case for TopicTrans, are still expected to be more different in metaphorical cases than in literal cases, as we assume for TopicSim. Therefore, TopicTransSim is a two-dimensional continuous feature that encodes the cosine similarity between the topic of the target sentence and the topics of the nearest sentences that have a different topic before and after the target sentence.

7.2.2 Multi-Level Modeling

Our sentence-level topical transition features are general across target words. However, the specific features that are informative for metaphor identification may depend on the target word. To account for the specificity of target words, we use multi-level modeling (Daume III, 2007). The idea of multi-level modeling is to pair each of our features with every target word while keeping one set of features independent of the target words. There are then multiple copies of each topic

[1] We also tried standard LDA for assigning topics to sentences, by representing each sentence as a topic distribution over its words. However, this representation was not as informative as Sentence LDA in our task, so we leave out the LDA topics in further discussion.

transition feature, all paired with a different target word. Thus, if there are N target words, our feature space becomes $N + 1$ times larger.

7.3 Experiments

We evaluate this approach on the same dataset as in Chapter 6. To display the experiment results, we use the following two baselines: the feature set of Chapter 6 and a context unigram model.

7.3.1 Settings

We ran Sentence LDA, setting the number of topics to 10, 20, 30, 50, and 100. α and β determine the sparsity of the topic distribution of each document and the word distribution of each topic, respectively; the lower the sparser. Following convention, we set these parameters to 0.1 and 0.001, respectively, to enforce sparsity. We also removed the 37 most frequent words in the corpus, drawing the threshold at the point where content words and pronouns started to appear in the ranked list. The models with 10 topics performed the best on the development set, with performance degrading as the number of topics increased. We suspect that poorer performance on the models with more topics is due to feature sparsity.

We used the support vector machine (SVM) classifier provided in the `LightSIDE` toolkit Mayfield and Rosé (2010) with sequential minimal optimization (SMO) and a polynomial kernel of exponent 2. For each experiment, we performed 10-fold cross-validation. We also trained the baselines with the same SVM settings.

Model	κ	F1	P-L	R-L	P-M	R-M	A
Unigram	.435	.714	.701	.434	.845	.943	.824
C	.575	.786	.758	.587	.882	.943	.859
C + AllTopic*	.619	.809	.784	.630	.893	.947	.873

Table 7.1: Performance on metaphor identification task. (**Models**) C: Frame Contrast model from Chapter 6, (**Metrics**) κ: Cohen's kappa, F1: average F1 score on M/L, P-L: precision on literals, R-L: recall on literals, P-M: precision on metaphors, R-M: recall on metaphors, A: accuracy, *: statistically significant ($p < 0.05$) improvement over corresponding baseline by Student's t-test.

7.3.2 Results

The results of our classification experiment are shown in Table 7.1. We tested our sentence-level frame transition features in combination with lexical features from our baselines: unigram and frame contrast features from Chapter 6.

Adding the sentence-level topical transition features to the baselines improved performance in predicting metaphor detection. When these transition features are combined with the frame contrast features from Chapter 6, we see large gains in performance.

	Topic	Top Words	Example Sentences
0	Disease/ Treatment	get, chemo, if, they, as, out, can, like, now, she, feel, did, up, know, think, been, good, time, or, when	I'm scared of chemo and ct scans because it makes cancer come back and you become more resistance to treatment with drugs like these later.
1	Food	good, they, gt, can, like, eat, fat, or, if, some, one, as, them, get, up, fiber, think, more, what	*Martha's Way* Stuff a miniature marshmallow in the bottom of a sugar cone to prevent ice cream drips.
2	Emotions	love, great, laura, good, hope, like, debbie, amy, up, happy, too, everyone, day, glad, look, fun, mary, what, kelly, how	Too funny. / You're so cute! / ene23...the photo in the locket idea sounds great!
3	Time	chemo, week, go, last, then, next, weeks, taxol, good, done, treatment, first, start, one, more, rads, after, today, 'll, now	I am now 45, and just had my ONE year anniversary from finishing chemo last week!!
4	Greetings/ Thanks	thanks, hugs, hi, here, carrie, thank, welcome, love, us, glad, know, greg, good, everyone, thread, ladies, there, how, sorry, mags	Thank you so much for the story!! / Big Hugs!
5	People	she, he, they, out, get, up, her, when, like, one, as, from, there, our, time, did, if, can, go, what	She has three children and her twin sister has taken her and her 3 children in.
6	Support	good, hope, well, happy, everyone, doing, glad, luck, hear, better, take, jen, care, great, liz, birthday, hugs, lol, watson, feeling	YAY! / lol. / I wish you all good luck and peace.
7	Relation	what, know, she, as, can, her, cancer, if, there, has, think, been, how, like, our, who, when, they, would, us	She knows that she has BC but does not know that it has spread. / I just read your message and I wondered about you.
8	Religion	god, love, lord, us, prayers, our, bless, dear, her, lu, may, day, patti, thank, know, comfort, amen, xoxo, he, pray	Dear Lord, I come to you with a friend that is not doing well, Please bless her that her hands will reach for you threw the last part of her breast cancer fight.
9	Diagnosis	diagnosed, when, chemo, she, breast, years, stage, cancer, dx, now, found, nodes, no, after, lump, they, age, then, year, mastectomy	I was 64 when diagnosed wtth pure DCIS.....I had my ninght radiation treatment today. / I was diagnosed Nov 2007 at age 45.

Table 7.2: Topics learned by Sentence LDA.

7.4 Discussion

Metaphorical instances tend to have personal topics. An author was more likely to use target words metaphorically when the target sentence relates more closely to their own experience of disease and treatment. Specifically, metaphors were relatively frequent when people shared their

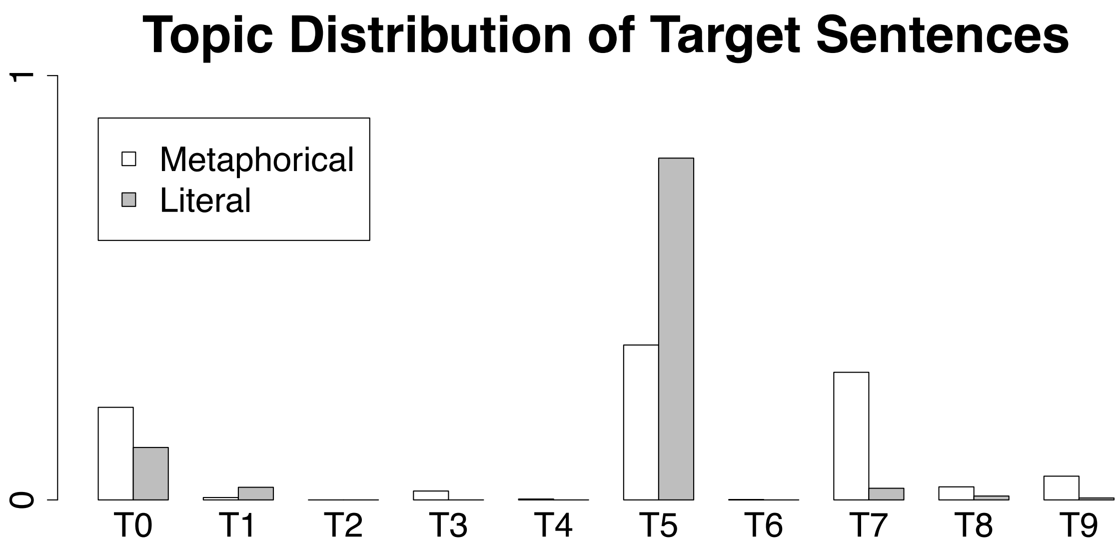

Figure 7.1: Proportions of topics assigned to target sentences, when target words were used metaphorically vs. literally. The proportions of metaphorical and literal cases are different with statistical significance of $p < 0.01$ by Pearson's chi-square test.

own disease experience (Topic 0, Topic 9) or sympathized with other people's experiences (Topic 7), but were more infrequent when they simply talked about other people in Topic 5 (Figure 7.1). According to our closer examination of sample sentences, Topic 0 had many personal stories about disease and treatment, and Topic 7 was about learning and relating to other people's experiences. Example metaphorical expressions include "There is *light* during chemo." (Topic 0) and "Hi Dianne - I am glad I found your post as I am sort of in the same *boat*." (Topic 7).

The topics of the surrounding context (TopicTrans) were also informative for metaphor detection (Figure 7.2). However, the topics of the surrounding sentences followed an opposite pattern to the topics of the target sentence; talking about other people (Topic 5) in the context of a target sentence led to more metaphorical usage of target words. Similarly, writers used target words more literally before or after they shared their personal stories (Topic 0). This pattern could be because the topic of the target sentence differs from the topics of the surrounding sentences in these instances, which would mean that the target sentence is a topic that is more likely to be literal. Topic 9, however, does not follow the same pattern. One possible reason is that Topic 9 and Topic 0 tend to frequently co-occur and be metaphorical. Thus, if a target word comes after or before Topic 9 and it is Topic 0, then this word may more likely be metaphorical.

Topic transitions are effective indicators of metaphor. Metaphorical instances accompanied more drastic topic transitions than literal instances. This tendency, which matched our hypothesis, was shown in all our topic features. The immediately neighboring sentences of metaphorical instances were more likely to have a different topic from the target sentence than those of literal instances (Figure 7.3). Additionally, differences in topic between the target sentence and the neighboring sentences were greater for metaphorical instances (Figure 7.4). The

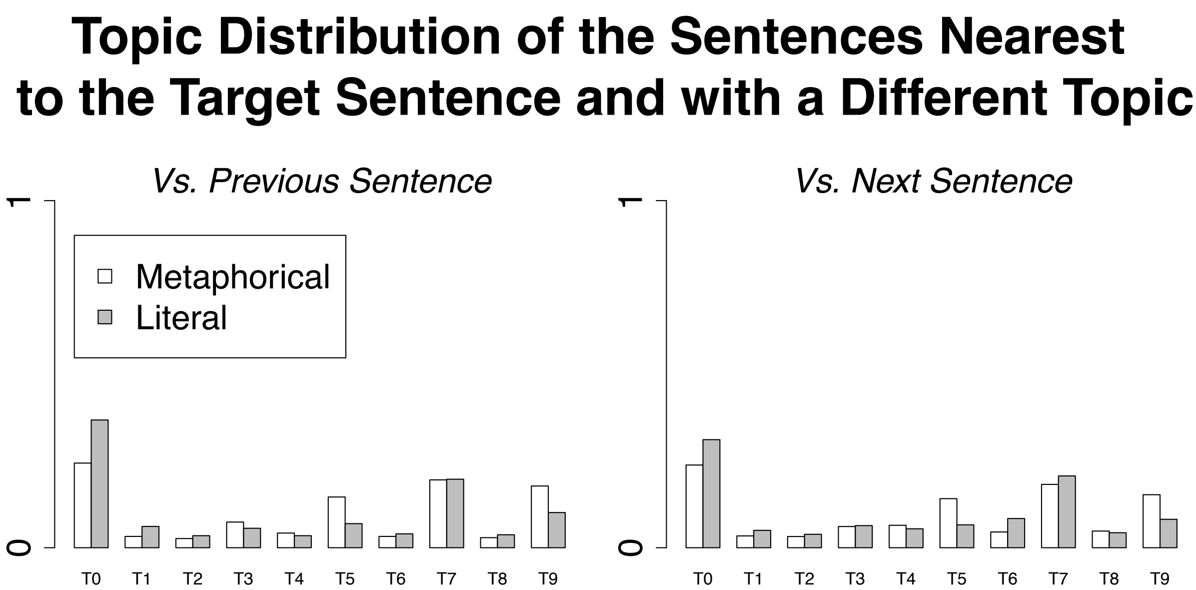

Figure 7.2: Proportions of the topics of the sentences that are nearest to the target sentence and have a different topic from the target sentence. The proportions of metaphorical and literal cases are different with statistical significance of $p < 0.01$ by Pearson's chi-square test.

nearest sentences with topics different from the target sentence (TopicTransSim) also showed this pattern (Figure 7.5). An interesting finding was that a topic transition after the target sentence was more indicative of metaphor than a transition before.

A multi-level model captures word-specific effects. Our features in context helped recognize metaphors in different ways for different target words, captured by the multi-level model. The paucity of general trends across metaphorical terms does not mean a limited applicability of our method, though, as our features do not suppose any specific trends. Rather, our method only assumes the existence of a correlation between metaphors and the theme of their context, and our multi-level model effectively identifies the interaction between metaphorical terms and their contexts as useful information.

7.5 Conclusion

In this chapter, we proposed a new, effective method for metaphor detection using sentence level topic transitions between target sentences and surrounding contexts. In particular, our system made significant gains in solving the problem of overclassification in metaphor detection, which was shown in the model that uses only frame contrast features (Chapter 6).

Our proposed features in this chapter can be expanded to other domains. Though in other domains, the specific topic transition patterns would likely be different, these features would still be relevant to metaphor detection.

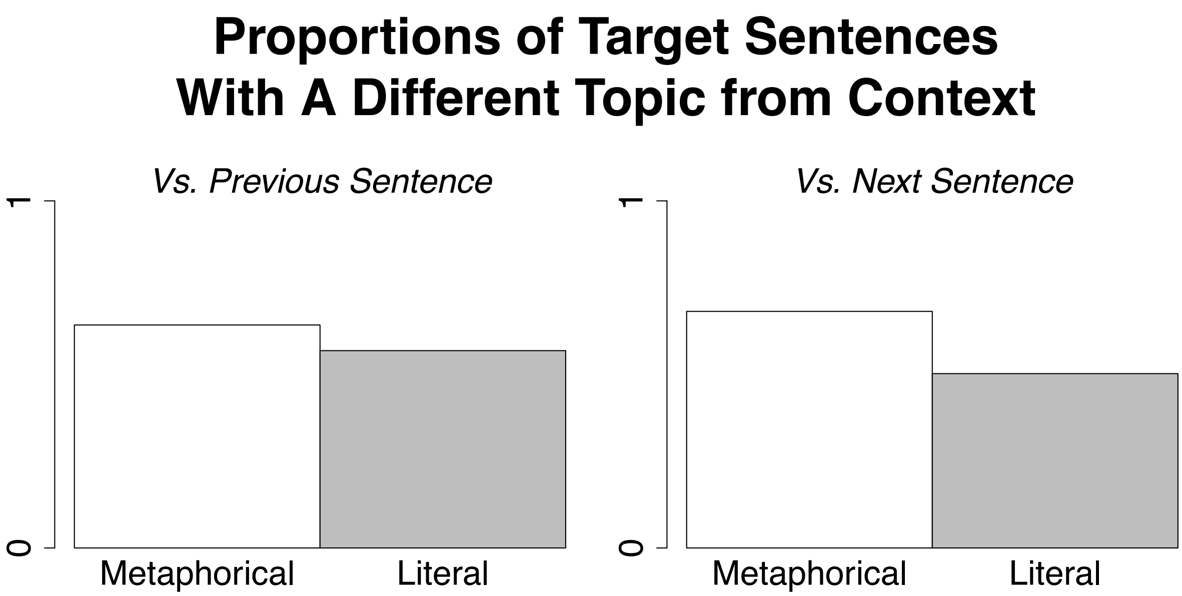

Figure 7.3: Proportions of target sentences whose topic is different from that of the previous/next sentence, when target words were used metaphorically vs. literally. The proportions of metaphorical and literal cases are different with statistical significance of $p < 0.01$ by Pearson's chi-square test.

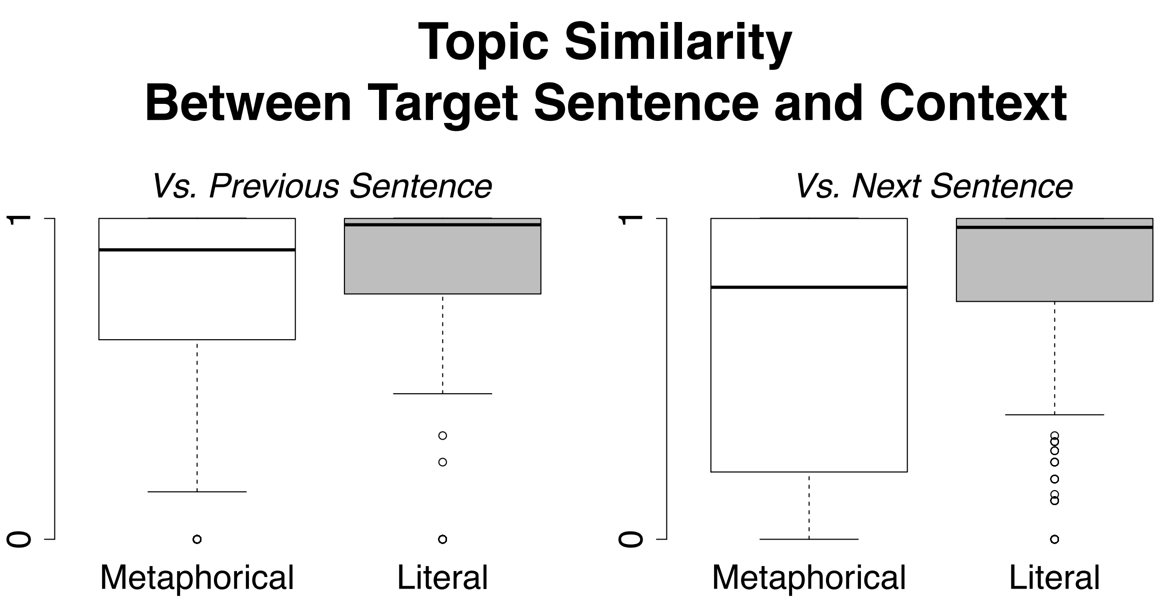

Figure 7.4: Cosine similarity between the topic of a target sentence and the topic of its previous/next sentence, when target words were used metaphorically vs. literally. The means of the metaphorical and literal cases are different with statistical significance of $p < 0.01$ by Welch's t-test.

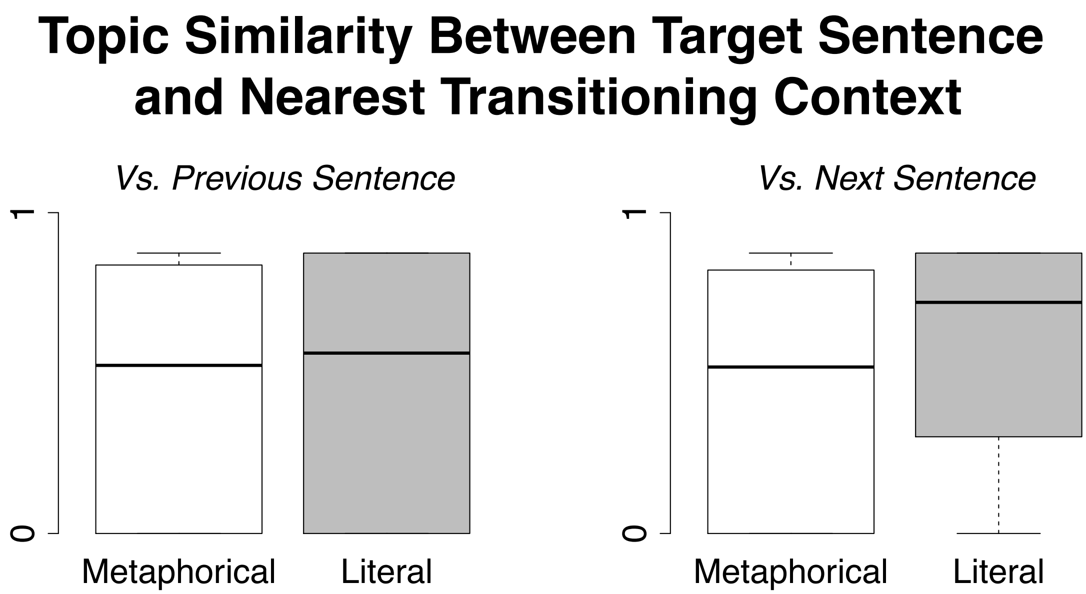

Figure 7.5: Cosine similarity of the topic of a target sentence and the topic of the sentences that are nearest to the target sentence and have a different topic from the target sentence. The means of metaphorical and literal cases are different with statistical significance only for the next sentence, with $p < 0.01$ by Welch's t-test.

Chapter 8

Metaphor Detection Using Frame Facets

In the previous chapters, we observed that detecting frame contrast and sentence-level topic transition using lexical context of an expression helps to predict whether the expression is metaphorical. However, frame contrast or transition do not guarantee an occurrence of metaphor. In addition, the approaches are still limited to identifying metaphor within a single sentence. In this chapter, we move one step forward to address the question of what makes a frame switch metaphorical. We infer implicit facets of a given metaphor frame using a semi-supervised bootstrapping approach on an unlabeled corpus. Our model in this chapter applies this frame facet information to metaphor detection, and achieves the state-of-the-art performance on our social media dataset when building upon the frame contrast and transition features in a nonlinear machine learning model.

8.1 Introduction

Language allows people to present and deliver information and ideas to others. In addition to this basic function, people use language to accomplish other interpersonal communication goals such as apologies, declarations, requests, threats, and consolations. In this way, language enables people to establish closeness with or distance from others, if used effectively in social situations. As we show in this chapter, metaphor is a useful linguistic resource for this purpose.

Metaphors are not used randomly. A metaphor invites participants to contribute to the meaning of a conversation in new ways, because it introduces facets and perspectives of a new domain. The new domain influences how another person might respond and become involved in the communication. For example, texts about people's experiences with cancer commonly use the following two metaphors: *journey* and *battle*. When people want to emphasize the process of cancer treatment, comparing it with progressing along a path, the journey metaphor introduces the role of a guide who can be part of the journey. When people want to emphasize the struggle, the battle metaphor introduces the notions of becoming a warrior and using a weapon. People choose a metaphorical domain that resonates with their communicative goals, for instance, seeking a good doctor (a guide) or seeking a good treatment method (a weapon).

Metaphors also serve a second function, related but different. In addition to introducing specific facets for specific communicative purposes, the speaker may introduce a metaphor as a general vehicle for creating an emotional or social connection between participants. It does so by sharing communication at a personal level, putting the topic in a very different domain and creating different associations. For example, literal cancer-related words invoke the medical domain, where one talks about details of the treatment. The medical domain tends to be mechanical and technical, so it feels more like a scientific topic than a personal topic about one's life experience. Using journey or battle metaphors, which are not technical or medical, can trigger a whole new set of associations. These associations allow conversation participants to think of cancer as more than a medical, technical, and scientific topic, and reframe it as a life experience. By making conversation more personal in this way, metaphor can more effectively support social functions such as sympathizing, consoling, and so on.

In order to discover how multiple metaphors can interact in this social perspective, we have to observe how people use metaphor back and forth in conversation. This requires study of extended (multi-sentence) metaphor, at the discourse level. However, most computational work on metaphor has focused on detecting one metaphor appearing within a single phrase or sentence. There are presently no approaches to computationally modeling metaphor at the discourse level, to our knowledge.

In this chapter, we aim to computationally model a metaphor frame (refer to Section 1.2) that operates across sentences, as an initial step toward understanding metaphor from a social perspective. To infer the implicit facets of a metaphor frame, we propose template induction using a semi-supervised bootstrapping approach. A template needs to have the different slots for facets of the corresponding frame. Each slot is filled with linguistic manifestations (in the text) of a facet. By applying these templates to a metaphor detection task, we can explore how frame switching occurs when people use metaphor. We first apply the templates to a metaphor detection task on the journey-related metaphors in the breast cancer domain. Then, we evaluate our approach's performance in two other domains (illness-related metaphors and battle-related metaphors in the political domain) to evaluate its generalizability and reveal some of its limitations. We hope that this work will bring about a better understanding of the structure behind metaphorical associations, and thus become an initial step toward detecting metaphor in an extended discourse, especially for modeling the use of metaphor in social interactions.

The remainder of this chapter is organized as follows. Section 8.2 shows how adopting the concept of a frame may be useful for studying metaphor in discourse from a social perspective. Section 8.3 explains our semi-supervised approach of template induction to model a metaphor frame in detail. Section 8.4 tests the effectiveness of the frame information through metaphor detection experiments. Section 8.5 analyzes the results and identifies when the frame information is beneficial. Section 8.6 presents experiments that apply the frame facet model to other domains, and discusses the limitation of the model. Section 8.7 concludes the chapter.

8.2 Metaphor Frames

According to our view on metaphor (refer to Section 1.2), a metaphor occurs when a speaker brings one frame into a context/situation governed by another. In this section, we offer a qualitative analysis of the data from this perspective, and the technical approaches described in forthcoming sections will build on this understanding.

The same or related metaphors from the frame may be used repeatedly. For example, EX(38) compares people to a gun and bullets, and EX(39) compares the world and people to a stage and players. Related metaphors can be used not only within a sentence, but also beyond a sentence. For instance, EX(40) compares the author's imagination to a circus and imagination-related things to circus-related things throughout the paragraph.

EX(38) "He is the pointing gun, we are the bullets of his desire."

EX(39) "'All the world's a stage and men and women merely players."

EX(40) "Bobby Holloway says my imagination is a three-hundred-ring circus. Currently I was in ring two hundred and ninety-nine, with elephants dancing and clowns cart wheeling and tigers leaping through rings of fire. The time had come to step back, leave the main tent, go buy some popcorn and a Coke, bliss out, cool down." (Dean Koontz, Seize the Night. Bantam, 1999)

In the breast cancer discussion forum we use in our work, community participants frequently bring in *journey* and *battle* frames when talking about their cancer experiences. Depending on what aspects of the cancer experience they choose to focus on, they invoke different frames accordingly even within the same text. For example, in EX(41), the *journey* and *road* metaphors are used to say that the speaker is having a similar experience with the listener. Further on, *weapons* from the *battle* frame are used to emphasize the power of faith and prayer in cancer treatment. In this way, metaphor introduces specific facets for specific communicative purposes.

EX(41) "I know, the age thing struck me too when I read about **your bc journey** - **we have been going down the same road at the same time, only in another part of the country!** It does help to know you are not alone! How amazing with the size of your tumor, that you did not have positive nodes. That is a miracle in itself. I do believe faith and prayer are our most powerful **weapons against this disease**. It is what gets me thru each day."

While metaphor provides resources for the speaker to use in communication, it also creates corresponding resources for the hearer. For example, EX(42)–EX(45) from the same thread in the breast cancer discussion forum shows how conversational participants repeat and expand one another's metaphors. The speaker in EX(42) starts using the *falling off the wagon* metaphorical idiom to convey her opinion that failing to stay on a controlled diet is okay. EX(43) relays the *falling off* part, and connects it to *journey*. EX(44) and EX(45) carry the *wagon* part of the initial post, and use *on the wagon* to describe her status (EX(44)) and her well-wishes for the other person with the extension of *get back on after you fall*. Although *falling off the wagon* and *on the wagon* are metaphorical idioms, *get back on after you fall* is a novel metaphor created by the following speaker. This novel metaphor is drawn from the *wagon* frame that has been brought into this conversation. In this way, a metaphor that is taken up by multiple speakers may increase empathetic understanding as well as add creative opportunities (e.g., for "fun") to the conversation.

EX(42) "**falling off the wagon** is no big thing in my opinion, the psychological good feelings of enjoyment weigh in big for feeling good."

EX(43) "**** **falling off is part of this journey**, it is stupid to deny yourself everything."

EX(44) "I am **on the wagon** so far today ...ongoing battle."

EX(45) "**** - hope you **stay on the wagon**, or at least **get back on after you fall**!"

As shown in the above examples, metaphor performs social functions through the switching of frames. In other words, observing frame switches offers insight into the ways in which people use metaphor to achieve social goals. The goal of our work is to lay a computational foundation for detection of such switches so that social strategies regarding metaphor use in interaction can be accomplished as follow-up work. Thus, in this chapter, we empirically construct a metaphor frame, and model the linguistic signals of frame switches.

8.3 Our Approach to Building a Metaphor Frame

To investigate how a metaphor frame appears in discourse, we computationally model frames that can be either metaphorically or literally used. A frame characterizes a conceptual domain, a "world" that is defined by a number of co-occurring facets. For example, the *journey* domain in *"life is a journey"* or *"he took a journey to Sweden"* could have facets such as *origin*, *destination*, *path*, *vehicle*, *companion*, and *guide*. Using a *journey*-related metaphor activates this domain and its facets, which become available as conversational resources in communication. In our work, we identify facet "slots" of a frame such as the *origin* and *destination* of the *journey* frame, and discover linguistic manifestations of the facets that fill the slots. We later use this frame information for metaphor detection, and observe how the same frame is used metaphorically or literally depending on its facets. We will call the facet slots *facets*, *facet categories*, or *facet slots*, and the linguistic manifestations *facet instances*.

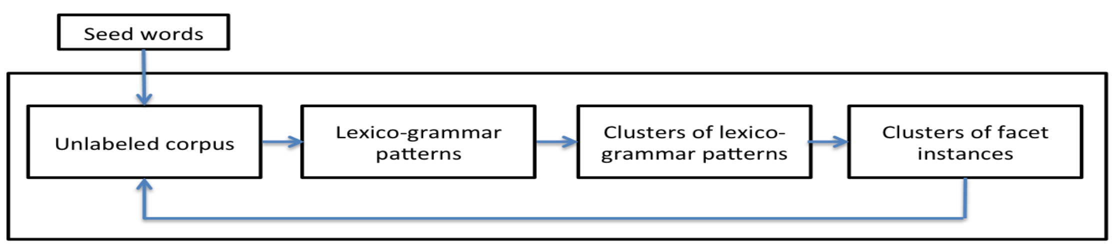

Figure 8.1: System flow diagram.

In order to obtain both facets (template slots) and facet instances (slot instances), we propose a simple bootstrapping algorithm (Figure 8.1) which expands on the number of the facet instances, inspired by earlier bootstrapping approaches such as Riloff et al. (1999, 2003) and Qadir and Riloff (2013). In our model, we assume that a sentence tends to contain more than one important facet of a metaphor frame. In other words, if a sentence contains one facet of a metaphor frame, the sentence is likely to contain additional facets. Additionally, we assume that facets and dependency relations have some relationship. There are certain grammatical patterns that represent semantic relations that connect facets in context. Note that we disregard frame facet instances that do not co-occur with a keyword (e.g., *journey*) within the same sentence. This can be considered a limitation of this approach.

1.	Harvest sentences containing the seed words from the unannotated texts.
2.	Parse the harvested sentences, and obtain lexico-grammatical patterns of the sentences.
3.	Cluster the lexico-grammatical patterns.
4.	Extract candidate facet instances from the lexico-grammatical patterns in each cluster.
5.	Compute the score of each candidate facet instance.
6.	Top ranked candidate facet instances of each cluster are added to the original seed words.
7.	Repeat starting with step 1.

Table 8.1: The bootstrapping process.

Our bootstrapping process begins with several seed words (Section 8.3.1) that specify the domain and provide seed facet instances. Using the seed words, we collect lexico-grammatical patterns (Section 8.3.2) in unannotated texts and cluster them to find facets (template slots) (Section 8.3.3). Next, the induced patterns are used for identifying facet instances which comprise a facet cluster (Section 8.3.4). Then, the most representative facet instances for each cluster are identified and added to the seed word set. Repeating this process expands the seed facet instances and lexico-grammatical patterns into larger sets. The overall sequence is illustrated in Table 8.1.

8.3.1 Seed words

The mutual bootstrapping process begins with predefined seed words and a text corpus. The seed words are the frame related words including the domain (e.g., *journey*) and a few examples of representative facet instances (e.g., *train*, *long*) for one or more unspecified facets. The corpus is then filtered for sentences that contain the frame (e.g., *journey*) and at least one example seed facet instance. Note that the sentences in the corpus are not annotated as metaphorical or literal. Since we are building a frame that can be used either metaphorically or literally, we do not require sentences where the seed words are used in a desired sense. For this reason, any general corpus that contains a sufficient amount of sentences with frame-related words can be used.

8.3.2 Collect Lexico-Grammar Patterns

We collect lexico-grammatical patterns using the seed words to represent relations between the domain and its facets. Representing relations in this way is a common approach in event extraction where relations often appear in text within a verb relation. For example, in a bombing event, *perpetrator* can be represented as a *person/org* who detonates, blows up, plants, hurls, stages, launches, or is detained, suspected, or blamed for the bombing (Chambers and Jurafsky, 2009). However, representing a relation for a domain and its facets for our purpose is not as straightforward as it is in event extraction because facets appear in more diverse ways than merely as verb relations. In particular, facets appear in a diversity of syntactic contexts.

As a solution, we propose using lexico-grammatical patterns generated from dependency paths between a domain word and facet words via the *ROOT*. The lexico-grammatical patterns are defined as the shortest path that passes through the *ROOT* in dependencies between the domain name and seed facet instances. For example, StanfordCoreNLP (Manning et al., 2014) outputs the dependencies in Table 8.2 for the sentence “She resumed her journey through the city.” The lexico-grammatical pattern that connects *journey* with other candidate property words such as *she* and *city* is defined as the reverse path from *journey* to *ROOT* combined with the path from *ROOT* to *journey*. The paths for the example are shown in Table 8.3. Words are lemmatized to reduce sparsity.

Sentence	She resumed her journey through the city.
Dependencies	nsubj(resumed-2, She-1) root(ROOT-0, resumed-2) nmod:poss(journey-4, her-3) dobj(resumed-2, journey-4) case(city-7, through-5) det(city-7, the-6) nmod:through(resumed-2, city-7)

Table 8.2: Dependencies from parsed result

origin	destination	pattern
journey	she	dobj_r(*origin*, resume), root_r(resume, root), root(root, resume), nsubj(resume, *destination*)
journey	city	dobj_r(*origin*, resume), root_r(resume, root), root(root, resume), nmod:through(resume, *destination*)

Table 8.3: Examples of lexico-grammar patterns. _r represents a reverse dependency.

This lexico-grammatical pattern representation has advantages. First, it allows us to represent patterns connecting pairs of words in a position-invariant manner. For example, in our baseline bootstrapping model, it is difficult to represent the pattern *reach __ of my journey* because *reach* is not located between the slot for a property instance and *journey*. However, using the lexico-grammatical pattern enables formalization of this pattern. Second, the lexico-grammar pattern is not affected by modifiers in the path. For example, the patterns representing the relationships between *journey* and *she*, and between *journey* and *city* do not change even for the sentence “She resumed her long journey through the city,” in which *long* has been added.

8.3.3 Cluster Lexico-Grammar Patterns

Using the idea that lexico-grammar patterns can approximate semantic relations, we first cluster collected lexico-grammar patterns so that each cluster may represent a different relation (facet slot).

The feature representation of each pattern is based on all arguments (e.g., *origin* and *destination* in Table 8.3) the pattern has in the corpus. For example, the pattern “dobj_r(*origin*, resume), root_r(resume, root), root(root, resume), nmod:through(resume, *destination*)” in Table 8.3 may have other *origins* and *destinations* in the corpus in addition to many occurrences of “city”. We use all arguments appearing with the pattern as features for the pattern, with the feature space size of the whole vocabulary. This is based on the idea that patterns with similar arguments would have similar roles that can be facet slots, which is similar to the distributional hypothesis (Harris, 1954).

For the clustering algorithm, we use Nonnegative Matrix Factorization (NMF) (Lin, 2007). We adopt this algorithm because our feature space is greatly sparse and NMF is effective for sparse data. We use the scikit-learn (Pedregosa et al., 2011) implementation of NMF.

8.3.4 Identify Representative Facet Instances

After obtaining pattern clusters, we extract tokens that match patterns in each pattern cluster. Tokens extracted for each pattern cluster are facet instance candidates.

Although we have clusters of similar facet instance candidates, there are many noisy instances in each cluster. To determine which instances are most reliable, we score each instance based on how far its generating patterns are from the center of the cluster. Specifically, an instance is scored high if it is found in more patterns in the cluster, and in patterns with higher within-cluster scores. We also take into account how semantically close each instance is to the other words in the same cluster. We use the GloVe vector representations (Pennington et al., 2014) to compute cosine similarity between two words. The scoring formula is shown below, where N_i is the number of different patterns that extracted $word_i$, Sim is the average cosine similarity with all other words in the same cluster, $score\_pattern_k$ is within-cluster score computed by NMF.

$$score(word_i) = Sim * \sum_{k=1}^{N_i} 1 + (.01 * score\_pattern_k) \tag{8.1}$$

Once the best facet instances are identified in this ranking step, the new instances are added to the original seed words, and the process repeats. The lexico-grammar patterns and property instances are clustered again and rescored after each iteration. The process stops after a specified number of iterations. For our experiments, we found five iterations to be sufficient. We leave an exploration of more heuristic stopping criteria to future work.

8.4 Evaluation

We evaluate our learned facet clusters, which define a particular metaphor frame template, with respect to how well they perform for an application, metaphor detection. In so doing, we assess the performance of the represented frame information and compare to our previous models in Chapter 6 and Chapter 7. The evaluation results are presented in Table 8.4. The results show that the new frame facet model performs significantly better than the previous models, which indicates that modeling metaphor in terms of frames is promising for distinguishing metaphorical and literal usage of words.

Section 8.4.1 explains our evaluation task, and which datasets we have used for the evaluation. Section 8.4.2 illustrates how we model the frame information as features for classification, and explains the classification settings used in our experiments. Finally, Section 8.4.3 provides the experiment results.

8.4.1 Evaluation Task

For our experiments, we use the metaphor detection task used in the previous chapters. The task is to decide whether a given target word is metaphorically or literally used. Because there is a set of pre-determined target words, this task is beneficial to see whether the applied model has disambiguating power.

We conducted our metaphor detection experiments on a subset of our breast cancer metaphor dataset. We chose to work on this dataset because this dataset contains conversational texts so that we can observe how people use metaphor in discourse. In addition, more importantly, this dataset has multiple target metaphors from a single frame, *journey*. From the cross-validation and development datasets used in the previous chapters, we select the journey-related words *road*, *train*, and *ride* to evaluate the journey frame template we built. We exclude other target words, *spice*, *boat*, *light*, and *candle* for our experiments because they do not belong to the journey frame. After filtering out these target words that are not relevant to the journey frame, the development dataset contains 488 instances, and the cross-validation dataset contains 1,119 instances.

To learn templates for the *journey* frame, we use unannotated data from the BookCorpus (Zhu et al., 2015). The corpus contains 11,038 books in 16 different genres. Particularly for our experiments, we use 74,004,228 sentences from the books, which are provided together with the original book files in the corpus. We use this data instead of more conversational data in order to minimize errors from detecting sentence boundaries and parsing, and to ensure broad topical coverage.

8.4.2 Features and Classification Settings

We extract a vector of binary features for each target word to indicate which of the learned facets of the journey frame appear in its immediate context. The presence of each cluster in the same sentence, preceding sentence, and following sentence relative to the target word; as well as the presence of each cluster in any of those three contexts, is indicated respectively by features in a vector of length four times the number of clusters.

We used the support vector machine (SVM) classifier provided in the `LightSIDE` toolkit Mayfield and Rosé (2010) with sequential minimal optimization (SMO) and a polynomial kernel of exponent 2. This enables the model to make use of contingencies between features. We expect that in order for a frame to be meaningfully identified, an appropriate topic shift coupled with identification of associated slot fillers in the nearby context is needed. The nonlinearity in this model enables this. For each experiment, we performed 10-fold cross-validation. We also trained the baselines with the same SVM settings.

8.4.3 Results

The results of our classification experiment are shown in Table 7.1. We tested our frame facet features alone (Facet), with context unigram features (Unigram + Facet), and with features from our previous model (Transition + Facet).

Model	κ	F1	P-L	R-L	P-M	R-M	A
Facet	.204	.602	.381	.369	.826	.833	.732
Unigram	.446	.720	.707	.434	.858	.950	.837
Unigram + Facet	.485	.742	.665	.520	.874	.927	.838
Transition	.618	.808	.789	.615	.899	.954	.880
Transition + Facet*	.655	.827	.814	.648	.907	.959	.891

Table 8.4: Performance on metaphor detection. (**Metrics**) κ: Cohen's kappa, F1: average F1 score on M/L, P-L: precision on literals, R-L: recall on literals, P-M: precision on metaphors, R-M: recall on metaphors, A: accuracy, *: statistically significant ($p < 0.05$) improvement over corresponding baseline by Student's t-test.

Adding our frame features to the baselines improved performance in predicting metaphor detection. We see that our features combined with the unigram features slightly improved over the Unigram baseline. However, when our features are combined with the features from our previous model using frame contrast and transition, we see large gains in performance, which suggests that there is an synergistic interaction between the frame facet features and the frame contrast and transition features.

8.5 Discussion

Our experiments show that frame facets that appear in surrounding sentences can be strong indicators of metaphor detection. This is promising, and suggests that observing frame facets can be crucial key to understanding how metaphor is used in discourse. However, the frame facets themselves are not as informative as when used with other features from the baseline. The improved performance when the frame facets are used with baseline features in the nonlinear model suggests that there are interactions among the features. In this section, we discuss the benefits of our model by examining prediction errors of our model and the previous model baseline.

The majority of the instances where the baseline model and our model do not agree is where our model improves on classifying literal instances as literal. In these cases, a topic shift is sufficient evidence of a metaphor, but the model without our template slots is not able to determine that. EX(46) and EX(47) show some specific examples where the baseline failed by incorrectly predicting metaphor. In both of these examples, a target word *road* is used literally, but the baseline classified it as metaphorical. Although their own topic transition features correctly captured that there is no topic transition in both cases, in combination with the frame contrast features, the baseline model did not make a correct prediction.

EX(46) ... Planning on having my right removed then reconstruction on both sides . I am an avid runner , road biker and downhill skier . Was looking at the tram flap. ...

EX(47) ... I did go to my son 's for Christmas , 500 miles away . My husband drove and we spent one night at our daughters to break up the time on the *road* .

When our frame features are added, however, the model correctly predicted that they are literal. This is probably because our frame features that picked up frame facet words surrounding the target word in combination with topic transition features strongly signaled literal usage of the target word. In EX(47), for example, our model picked up the distance word, *miles*, in the sentence prior to the sentence where the target word *road* resides.

From this, we can see that adding the frame facet information allows having more complete frame information for distinguishing metaphorical and literal usage of the topic frame. Our model seems to provide more fine-grained information about what pieces of the frame make it metaphorical or literal.

Conducting an error analysis on the instances where both baseline and our model failed reveals the limitations of using a topic frame based approach in general. EX(48) shows that *train* is used literally in the post. However, because there are different topical words around the target word and there is no other journey frame words, both the frame transition (sentence-level topic transition) model and our model classify the target word as metaphorical by picking up the topic transition.

EX(48) ... I woke at 2 a.m. because it was so quiet . I could n't hear the frogs or crickets and then I heard a *train* getting louder and louder and then it threw us around . When we got out the giant trees looked like x-mas trees from all the clutter in the tops of them . ...

8.6 Generalization to Other Domains

Thus far, we have focused on building a *journey* frame and using the frame facet information to detect journey-related metaphors in the cancer domain. This narrow focus was because we wanted to control for contextual expectation, target, and source (refer to Chapter 4) in order to not confound topical patterns when building a frame. Now that we have demonstrated the potential of the journey frame to help detect metaphor in the cancer domain, we take the same approach to building other frames and detecting metaphors in other domains. This will help verify if these approaches are still effective in other domains. The experiments in this section apply our three earlier approaches to a new domain.

8.6.1 Data

For a new domain and dataset for metaphor detection, our work uses the political discussion forums on Ravelry [1], a social networking site for fiber arts and one of the largest online communities. In the political domain, people use a wide variety of metaphors. Out of the many different metaphors in this domain, our work uses two popular frames: *illness* and *battle*. These two metaphor frames are chosen because they have related metaphors used with relatively high frequencies. Our specific target words for annotation are *headache* and *cure* for the illness frame and *battle*, *war*, and *warrior* for the battle frame. We use the same annotation instructions for

[1] https://www.ravelry.com/

Amazon Mechanical Turk as in Chapter 6 to annotate the data. The annotated data is 882 instances for the illness-related metaphors, and 1,196 instances for the battle-related metaphors.

8.6.2 Experiments

We apply all three approaches in this thesis to detect battle-related metaphors and illness-related metaphors in the political texts. For evaluation, we use the same metaphor detection task and experiment settings used in our previous chapters. The results are displayed in Table 8.5 and Table 8.6.

Model	κ	F1	P-L	R-L	P-M	R-M	A
U	.604	.802	.804	.738	.809	.860	.807
U+C	.619	.809	.804	.759	.821	.856	.814
U+C+T*	.647	.823	.813	.784	.837	.860	.827
U+C+T+F	.656	.828	.812	.799	.846	.856	.831

Table 8.5: Performance on metaphor detection for battle-related metaphors in the political domain. (**Metrics**) κ: Cohen's kappa, F1: average F1 score on M/L, P-L: precision on literals, R-L: recall on literals, P-M: precision on metaphors, R-M: recall on metaphors, A: accuracy, *: statistically significant ($p < 0.05$) improvement over the above one by Student's t-test, U: unigram model, C: frame context model, T: frame transition model, F: frame facet model.

Model	κ	F1	P-L	R-L	P-M	R-M	A
U	.155	.565	.868	.971	.450	.140	.849
U+C	.213	.598	.875	.965	.490	.194	.853
U+C+T	.247	.617	.879	.964	.518	.225	.856
U+C+T+F	.241	.614	.879	.961	.500	.225	.853

Table 8.6: Performance on metaphor detection for illness-related metaphors in the political domain. (**Metrics**) κ: Cohen's kappa, F1: average F1 score on M/L, P-L: precision on literals, R-L: recall on literals, P-M: precision on metaphors, R-M: recall on metaphors, A: accuracy, U: unigram model, C: frame context model, T: frame transition model, F: frame facet model.

Both the frame transition model (Chapter 7) and the frame facet model ((Chapter 8) show improvements over the unigram baseline when detecting battle-related metaphors, as shown in Table 8.5. However, the effects of the frame contrast features (Chapter 6) and the frame facet features are not large. The analysis of these results reveals several issues to be addressed in our current models. The experiments with battle-related metaphors show the following limitations:

- Our approach uses FrameNet semantic categories to compute semantic category features (discussed further in Chapter 6). However, FrameNet does not have a semantic category for our target word *warrior*, which results in poor performance. Manually constructed language resources like FrameNet will continue to have this challenge, because they have a limited amount of information available.

- The learned frame facets for the battle frame underperformed for two reasons. There were two reasons. First, our approach of obtaining lexico-grammatical patterns between frame facets is not generalized. In our approach, we just use words as they appeared. However, the collected lexico-grammar patterns are too sparse to do proper clustering. Second, we use the BookCorpus for inferring frame facets. With this corpus, the lexico-grammatical patterns in the sentences that contain battle-related words were not obviously related to the relationships between frame facets. For the battle frame, it seems that we need to either generalize lexico-grammar patterns or find another corpus where battle-related facets are represented in simpler syntactic structures.

Our three models do not show improvements over the above model in Table 8.5 when detecting illness-related metaphors. In addition, the overall metaphor detection performance in this illness domain is lower than the journey domain and the battle domain. The experiments with illness-related metaphors show the following limitations:

- Unlike the battle frame, the illness frame facets are inferred decently. However, when the model searches for the frame facets (related words) in the neighboring sentences, it can only identify metaphors it has already found with other methods. In other words, the topically related words near a target word do not help metaphor detection for the illness-related metaphors.
- When topically related metaphors appear in the illness frame, they appear quite far from each other, not in neighboring sentences. Our approach only considers related words in neighboring sentences. This may be why our approach does not help detect illness metaphors in the political domain.

By applying our model to other domains, we are able to understand the weaknesses of our current models. Based on the weaknesses revealed, to generalize our current models, the immediate next steps will be (1) generalizing lexico-grammatical patterns by using a more abstract form, such as part-of-speech (POS) or semantic word categories; (2) generalizing lexico-grammatical patterns by using more context than just the relationships within a sentence; and (3) extending how to use frame facet information in metaphor detection, finding it in a wider lexical context (like the whole text) rather than only within neighboring sentences. All these expansions will help improve our models in future work.

8.7 Conclusion

In this chapter, we argued that a frame-based approach is useful for metaphor detection and may be useful in subsequent work for studying metaphor from a social perspective. To support this, we developed a semi-supervised computational approach for constructing a metaphor frame from unlabeled text. We demonstrated the effectiveness of this frame information in metaphor detection when used together with other proven features in a nonlinear machine learning model, which suggests interactions among the features. We discussed the ways in which the frame and topic information anchor the classifier to allow for more accurate metaphor detection. We also discussed the issues that were revealed when we applied the model to other domains.

Although our approach showed promising results suggesting that how frame facet information is used in text helps determine a frame's metaphorical usage, applying frame information to metaphor detection in this way is limited in scalability – we need to know which frame target words belong to in advance. Our contributions here demonstrated the potential of modeling metaphor through the lens of frame theory; we hope to address scalable ways of leveraging frame information, For example, future work could explore automatically detecting primary frames in a text.

Chapter 9

Wrapping up

In this chapter, I summarize the contributions of this thesis and suggest future research directions in computational modeling of metaphor.

9.1 Summary of Contributions

In this thesis, I model metaphor as it is used in every-day discourse, with the goal of laying a foundation for analysis of metaphor from a social perspective. Adopting the concept of *framing* in discourse, I suggest three approaches that leverage lexical contextual information for metaphor detection: (1) features of frame contrast, (2) features of frame transition, and (3) features of frame facets. Each approach implements a different piece of the idea of metaphor as frame switching. For the first time in the computational linguistics and NLP fields, these approaches detect metaphors that can be nearby related metaphors, which is a big initial step for discourse-level metaphor processing. These approaches stand in contrast to previous conceptualizations of metaphor as a violation of narrowly defined linguistic rules such as selectional restrictions. Instead, they adopt a softer, Gricean notion that a broadly construed expectation of coherence has been flouted. This thesis itself does not entirely "solve" the problem of metaphor detection at the discourse level yet. However, the new potential to adopt the frame view into the computational modeling of metaphor will make the work in this thesis foundational for metaphor detection in and extended naturalistic discourse. Detailed contributions of this work are as follows.

- I present analyses (Chapter 3) that show how much everyday conceptions of metaphoricity diverge from theoretical perspectives, and provide an insight on how models of metaphoricity may need to be adapted in order to adequately characterize metaphors in strategic use.
- I present a new annotation scheme for metaphor (Chapter 3) that maps variations in metaphoricity to a three-point scale of nonliteralness: *nonliteral*, *conventionalized*, and *literal*. Annotators can easily assign an expression to place on the scale by simply answering a set of decision questions.
- I present multiple metaphor corpora (Chapter 3, Chapter 6, Chapter 8) based on laypeople's recognition of metaphor using Amazon's Mechanical Turk crowdsourcing marketplace.

- I present a case study (Chapter 4) to uncover the relationship between people's use of metaphor and internal situational factors. Specifically, logistic regression model analyses show that people's metaphor usage is influenced by psychological distress conditions, including PTSD, depression, and anxiety. Unlike prior work, our annotation scheme allows us to separately assess the effects of three factors on language choices: contextual expectations, content of the message, and framing. Separating these factors helps us understand their impacts across different conditions at a deeper level.
- I present a case study (Chapter 5) to uncover the relationship between people's use of metaphor and external situational factors. This study provides evidence that people's propensity to employ metaphorical language increases around the time of stressful cancer events. Additionally, the study quantitatively verifies the effectiveness of considering the external situational features in metaphor detection.
- I present several models (Chapter 6, Chapter 7, Chapter 8) for detecting metaphors that can be nearby related metaphors, and are not restricted in their syntactic positions. This is the first time detecting metaphors that can be nearby related metaphors has been attempted in computational modeling of metaphor. To do so, the models find topical regularities appearing in the metaphor occurrences by leveraging lexical contextual information, based on the *framing* concept in discourse. The metaphor detection experiments using these models demonstrate the effectiveness of the frame information when used together in a nonlinear machine learning model, which suggests interactions among the features. These experiments show the potential of computational models that look at metaphor through the lens of framing.
- I present a model (Chapter 6) for capturing frame contrast around metaphorical frames. I propose several textual descriptors that can capture global contextual shifts within a discourse, such as semantic word category distribution, homogeneity in topic distributions, and lexical chains. Additionally, I show that global and local contextual information are complementary in detecting metaphors, and that leveraging these syntactic features is crucial in better describing lexico-semantic information in a local context.
- I present a model (Chapter 7) for capturing frame transition patterns around metaphorical frames. This model explores topic transitions between a metaphor and its context, rather than only detecting lexical discrepancies.
- I present a model (Chapter 8) for capturing frame facet patterns around metaphorical frames. I construct a metaphor frame from unlabeled text by using semi-supervised bootstrapping, which is used for finding frame facets in a lexical context when detecting a metaphor. This study is especially meaningful because it exploits semantically related words in a lexical context to disambiguate the metaphoricity of the target word, which is a big first step in extended metaphor detection.
- I present the issues exposed when applying the models to other domains (Chapter 8), beyond "journey" related metaphors in cancer discussions. I construct metaphor frames for "battle" and "health", and apply the frame information to metaphor detection in political discussions. The issues revealed in these experiments will help develop a more generalized

computational model for metaphor detection at the discourse level, which does not require controlling for topic domains.

At a high level, the first part of this thesis (Chapters 4, 5) illuminates the value of studying people's metaphor use in conversation through case studies about internal situational factors (psychological distress conditions) and external situational factors (stressful cancer events). By studying this, we can both learn more about the influences of situational factors and gain insight into how to computationally model metaphor in discourse. Motivated by the usefulness of studying metaphor in discourse shown in these case studies, the second part (Chapters 6, 7, 8) illustrates the value of looking at metaphor based on the frame concept, through simple implementation of frames by leveraging lexical contextual information. This new lens of operationalizing metaphor can open up new approaches to computational metaphor processing, and thus create opportunities for modeling the use of metaphor in interaction (e.g., allowing social intervention using metaphor) as well as improving the performance of downstream NLP applications (e.g., machine translation and dialog agent systems).

9.2 Future Directions

Although this thesis demonstrated the potential of situational factors and frame-based models for metaphor processing, this is just an initial step towards discourse-level metaphor processing. In this section, I discuss limitations of the models presented in this thesis, and suggest possible future directions for metaphor research in the fields of computational linguistics and natural language processing.

9.2.1 Follow-up Experiments

In this thesis, we learned frame facets automatically from unannotated text by using a bootstrapping approach. These learned facets were then used as features for metaphor detection. Since I focused this thesis on showing the potential of the frame notion for metaphor detection, when I inferred the frame facets, I simplified the process for a proof of concept. Therefore, there is room for improvement on this current model.

9.2.1.1 FrameNet

One way to make such an improvement is to utilize FrameNet (Baker et al., 1998). FrameNet is a lexical resource that contains frame elements, which are equivalent to the frame facets in this thesis. FrameNet is manually annotated, which makes it more accurate but offers less coverage. Our frame facets are semi-automatically learned, which means we have more coverage, but less accuracy. Therefore, it would be beneficial to test FrameNet in combination with our model. In light of this, I suggest the following experiments:

- FrameNet would be a great baseline to test our model of learned frame facets against in order to determine if our model's learned facets make sense. In addition, comparing the FrameNet features and our learned features in metaphor detection could give valuable lessons for extending our model.

- Using FrameNet features together with our learned features could give some synergy between the two sets of features, since FrameNet's features have more accuracy but less coverage and our features have less accuracy but more coverage.
- The key to maximizing the usefulness of FrameNet would be expanding the frame elements by using corpus-based approaches together in order to overcome the limitation of its coverage.

9.2.1.2 Syntactic Patterns

One important assumption in our frame building is that syntactic structures reveal relationships between frame facets. Because I took a simple approach to creating syntactic patterns by computing lexico-grammatical patterns, other useful experiments would involve testing our assumptions as well as extending and improving our lexico-grammatical pattern approach.

- As a follow-up analysis, it would be helpful to probe more deeply into the ways in which the pairings of syntactic structures with specific metaphors in specific contexts can help or hurt the performance of our model.
- It will be valuable to generalize lexico-grammatical patterns and find out which level of generalization would be most helpful. We currently use raw lexical items, or words, together with dependency relations in our lexico-grammatical patterns. However, a more abstract form, such as part-of-speech (POS) or semantic word categories could be more beneficial when defining a pattern for a relationship between frame facets. In this way, this method could become more generalizable to other domains.

9.2.1.3 Situational Factors

We used two case studies to explore internal and external situational factors that cause distress and examined how they are related to people's metaphor usage. Future research could more thoroughly explore situational factors.

- It would be interesting to investigate whether metaphor use might indicate other psychological disorders, specifically lexical aphasia and schizophrenia.
- It would be beneficial to try differently sized context windows for the critical period of a cancer event, in order to see the effects of time with respect to situational factors.

9.2.2 Possible Extensions

In this section, I will suggest possible extensions that would build on the frame models I showed in this thesis as a foundation.

- **Extended Metaphor Detection**

  This thesis shows that modeling metaphor through the lens of frame theory could be the first step in detecting extended metaphor (defined as a series of related metaphors under the same frame) in discourse. Although the experiments in this thesis are performed only on a dataset where one target word in a text is annotated as metaphorical or literal, they touch on the very interesting problem of uncovering what makes a particular frame switch metaphorical or literal. We hope to use this frame information to help detect extended metaphor. Obtaining a metaphor corpus that contains a sufficient number of extended metaphors is a big challenge. However, once such a dataset becomes available, we believe that the findings and lessons from this thesis will be useful in that context.

- **Metaphor Interpretation and Generation**

  Once more research uncovers the mechanism of metaphorical mappings, the frame templates built in Chapter 8 could be influential in metaphor processing beyond just detection. If frames for both source and target domains of a metaphor become available, we could use them to infer which common properties between frames are used in the metaphorical mapping. Among the common properties, it will be interesting to explore which properties are emphasized by the metaphor. As the underlying mechanism of metaphorical mappings gets revealed, it will be informative not only for extended metaphor detection, but also for metaphor interpretation and metaphor generation. For example, common properties in both target and source frames could help with metaphor interpretation. Frame facets from the same domain could help in generating new extended metaphors that relate to a given metaphor. For example, some frame facets could be new related metaphors.

- **Metaphor Use in Interaction**

  All the above suggestions ultimately pave the way for computational modeling of metaphor use in interaction. The long-term goal of our research is to use computational modeling of social and discourse uses of metaphor to interpret metaphor at the level of social positioning and discourse functions. One immediate next step could be to investigate the different contexts that metaphors are used in.

In this thesis, I've looked at how people use metaphor, but haven't answered how metaphor functions from a social perspective. When we use metaphor, we are letting someone into our mind. We bring someone into our cognitive space with us to see things the way we see them. It is not just about the information, but the significance we place on it; it is not about *what* we see, but about *how* we see it. We make our way of describing things more vivid by giving the listeners a picture, which includes an additional layer of emotion and nuance. In other words, using metaphor creates a sense of inter-subjectivity (Gillespie and Cornish, 2010). This inter-subjectivity can give us the understood terms of persuasion, the understood terms of creating closeness, and thus enhances empathy.

All this can be more pronounced with extended metaphor since extended metaphor makes the metaphor frame more salient, maybe even as salient as the governing frame. It serves to project the metaphor frame more intensely in the reader's or listener's mind. Also, it opens up the possibility of inserting something (another frame) into that (second) metaphor frame, which would perhaps not been coherent if it were inserted directly into the main governing frame.

Therefore, we can characterize metaphor from a social perspective better with extended metaphors rather than single metaphors, because extended metaphor allows us to better study how people play with language and use it to interact with others. Once we model metaphor at the discourse level, especially extended metaphor in running discourse, it will open up more opportunities to uncover the mechanism behind how people achieve social goals by "tossing metaphors back and forth." For example, it will help us better understand appropriate metaphor timing and selection. This understanding will enable many NLP applications to interact with people better by using human languages more effectively.

Appendix A

Metaphor Annotation Scheme

An expression is nonliteral or figurative when its meaning departs from conventional and original senses of words. For example, when saying “Enjoy your journey, and good luck!” to a person who starts chemotherapy, “journey” means chemotherapy which is considered by the speaker to have some characteristics of a long voyage. Nonliteral language is frequently used everywhere from literature to ordinary language, in order to make text and speech more effective, attractive, and convincing to its readers or listeners. Thus, detecting and understanding nonliteral language is very important to understand human language.

There has been computational work on nonliteral language, especially focusing on metaphor detection (Mason, 2004; Shutova, 2013; Shutova et al., 2010; Turney et al., 2011) and idiom detection (Lin, 1999; Sporleder and Li, 2009). There is also work on general nonliteral language detection that does not distinguish between types of nonliteral language (Birke and Sarkar, 2006; Feldman and Peng, 2009, 2013; Li and Sporleder, 2010).

One of the main challenges in nonliteral language detection is the lack of large annotated datasets. Annotating nonliteral language is nontrivial because of non-consensus regarding annotation schemes and clear definitions. Pragglejaz-Group (2007) suggested a metaphor annotation scheme, which introduced a systematic approach with clear decision rules. In this scheme, words are considered to be metaphorically used if they are not used according to their most basic concrete meaning. In this way, however, most words, even words people often do not recognize to be metaphorically used due to their widespread usage for a long time, tend to be annotated nonliterally because of the hard standards. For instance, support in I support you will be considered to be metaphorical according to this scheme because support here is not used as its most basic concrete meaning bear all or part of the weight of. For some tasks, this is an ideal scheme to follow, but may not be suitable for other tasks where it is important to distinguish dead metaphors from living metaphors.

The goal of this coding manual is to provide an annotation scheme for metaphorical expressions in text. Unlike Pragglejaz-Group (2007), we aim to develop an annotation scheme for more obvious nonliteral language because we are more interested in nonliteral expressions that people purposely use. Our annotation scheme includes three types of nonliteral language: metaphor, simile, and idiom. We consider simile to be a special case of metaphor using explicit comparison with words such as like. We include metaphorical idioms because they are obvious to be nonliteral despite the fact that they have lost their source domains.

Specifically, this coding scheme is applied to annotate texts found in breast cancer discussion boards, which are more casual than news articles or research papers. However, this scheme should be applicable regardless of domain. The results of annotating can be used as gold standard labels for automatic nonliteral language detection tasks.

Target to be coded	Code to be assigned	Brief description
Nonliteral	N	Metaphor, simile, metaphorical idiom
Conventionalized	C	Potentially ambiguous between literal and figurative usage
Literal	L	None of the above

Table A.1: Target nonliteral language use (a word or phrase) corresponding to our scope of nonliteral language [A.1].

If there is a phrase or word [A.2] in the sentence you think might be nonliteral language, enter it into the span field, then follow the steps below. If you code more than one expression in the sentence, create a separate line for each expression after the first.

1. Do you not know what the word/phrase means? Or are you very unsure of whether its literal or not?
   - Make a note of it in the notes column, then continue
2. Is the expression using the primary or most concrete meanings of the words? [A.1]
   - Yes = L (literal)
3. Does the expression include a light verb that can be omitted without changing the meaning, as in I take a shower =¿ I shower? If so, the light verb expression as a whole is literal. [A.6]
   - Yes = L
4. Is the metaphor composed of a single compound word, like painkiller, used in its usual meaning? [A.7]
   - As a single word, analgesic medicine is its most concrete meaning, but it doesnt literally kill anything either.
   - Yes = L
5. Is the expression a conventional term of address, greeting, parting phrase or a discourse marker? [A.8]
   - Yes = L
6. Is the expression using terminology or jargon very common in this domain or medium? [A.9]
   - Examples in a web forum: thread, link, surf, post, board, etc.
   - Other examples. leg for furniture supports, body for essays and articles, eye for hurricane
   - Yes = L
7. Is the expression merely hyperbole/understatement, sarcasm or metonymy? [A.1-(3)]

    - Yes = L
    - Hyperbole/understatement are differences in intensity or magnitude, e.g. ancient for old
    - Sarcasm is a difference in polarity (good substitutes for bad)
    - Metonymy is a reference by association rather than a comparison, e.g. The White House denied the rumor, the White House stands in for the president because it is associated with him, it is not being not compared to him.

8. Is the expression a fixed idiom like kick the bucket that could have a very different concrete meaning? [A.5]
    - Yes = N
9. Is there another common way to say it that would convey all the same nuances (emotional, etc.)? Or, is this expression one of the only conventional ways of conveying that meaning? [A.10]
    - If yes to the latter = C (Conventionalized)
10. Is the expression a simile, using like or as to make a comparison between unlike things? [A.4]
    - Yes = N (Nonliteral)
11. Is the expression unconventional/creative and also using non-concrete meanings? [A.3]
    - Yes = N
12. If you cannot otherwise make a decision between literal and nonliteral, just mark it as C.

A.1 Decision for nonliteral language

Decision for nonliteral Language: In our definition for the coding, an expression is nonliterally used when satisfying the following three conditions: (1) the expression, whether a single word or a group of words, needs to have an original established meaning, (2) the expression needs to be used in context to mean something significantly different from that original meaning, and (3) the difference in meaning should not merely be a difference in intensity or polarity. Explanations for each condition are following.

(1) **Have an original meaning**

The expression or the words within the expression need to have original established meanings. For example, in the phrase "kick the bucket" the words "kick" and "bucket" have clear and commonly known original meanings. The phrase "kick the bucket" could easily be used according to those meanings; in other words, to mean, "strike the bucket with ones foot".

(2) **Alter the original and established meanings of the words**

The usage needs to change the original meaning of the expression in some way. For the same example, "he kicked the bucket," the nonliteral meaning of "he died" is far from the literal meaning of "he struck the bucket with his foot." In many idioms such as this,

this meaning will be non-compositional, meaning that it cannot be determined from the meanings of the individual words.

Another example would be "keep your eyes open" which has an obvious original meaning. An ophthalmologist might use it in that way, to keep your eyelids open while examining your eyes. But it is also used to mean, "to watch carefully for someone or something, often while you are doing something else." This meaning does not derive from the original meanings of "keep," "eyes" and "open" but it is more easily interpreted than idioms like "kick the bucket."

(3) **Should not merely alter the intensity and polarity of the meaning**

The usage needs to alter the original meaning of the expression but should not simply be a change in the intensity or the polarity of the meaning. Language uses like hyperbole and understatement may simply change the intensity of the meaning without otherwise altering it. If one says "Im starving," the original meaning would be that the person is "dying from a lack of food." But if used to mean, "Im very hungry" then it is nonliteral and an example of hyperbole. The meanings of "starving" and "hungry" are very similar, but with "starving" having a much more intense meaning. One could also be hyperbolic with simple use of intensifiers like "very." Conversely, understatement works in the opposite direction. For example the sentence "I know a little about running a company" when said by a highly successful businessman would not mean that they actually know only a little bit about business. In that case, the true meaning is simply stronger than the literal meaning but not otherwise different.

Language uses like sarcasm instead change the polarity of the meaning. Sarcasm is when "you say the opposite of what you really mean in order to be rude to someone" [note for Collins dictionary]. As such, the polarity of the expression is changed. For example, a person can say "you are such a wise guy" when the person he or she was talking to did something stupid. In this context, what the expression actually means is "you are not wise," negating the literal meaning of the original sentence. Sarcasm is often difficult to detect in writing, which also would make it harder to code.

In this framework, we do not deal with nonliteral language related to the intensity of the language such as hyperbole and understatement, which exaggerate or weaken the original meanings in the same direction of the meaning, or sarcasm, which negates the original meaning. Note that although we do not include hyperbole, understatement, and sarcasm in general, those techniques can be used along with other forms of nonliteral language. For example, if a 20 year old referred to a 30 year old as "a dinosaur" the word "dinosaur" would be used nonliterally to mean "very old" as a metaphor but in that case it would also be hyperbole. In other words, "dinosaur" is both metaphor and hyperbole. We would count that usage because it is not merely hyperbole.

Our scope to code nonliteral language includes expressions such as metaphors and idioms which alter the meanings of the words. We also include similes which work similarly to metaphors but are explicitly marked.

A.2 Boundary to be coded

The unit to be coded is limited to the words or phrase whose meaning is altered. It can include any articles or modifiers in the phrase.

(1) A metaphor generally consists of target and source. For metaphor and simile, only the target is to be coded. In the following examples “a fox” is coded.

> Example1. My sister is a fox.
>
> Example2. My sister is like a fox.

(2) For an idiom, the whole idiom is coded.

> Example3. I never know how to break the ice with someone Ive just met at a party.

(3) If the expression contains modifiers such as additional adjectives or adverbs, the whole expression including the additional words are to be coded.

> Example4. My sister is a sneaky fox.
>
> Example5. I have come from the scared “Valley of Darkness” and you know it wasn’t nearly as dark as I expected.

A.3 Metaphor

A metaphor describes a subject, known as the source, by comparing it to another object, known as the target. A typical form of a metaphor is “A is B.” It is also possible for the source to be referred to as the target without being explicitly mentioned in the sentence. In other words, the source (A) can be omitted. Metaphors can use any kind of propositional words, but function words, expletives or other words that are empty of propositional content cannot be in themselves metaphors. A preposition can be a function word in a phrasal verb or it can be lexical and thus part of a metaphor. But since they are often used as function words, we exclude them from our coding. Metaphors can be nominal (e.g., “Juliet is THE SUN”), predicative (e.g., “The politician SAILED into the crowd”), or sentential (e.g., “THE LABORERS RETURNED TO THE FIELDS” to describe an office after a coffee break) [reference].

> Example6. I’m still working and feeling great and ready to just move forward with this journey.

In this example, “this journey” doesnt refer to a physical journey from one place to another given the context of chemotherapy. “This journey” refers to chemotherapy which is considered by the speaker to have some characteristics of a long voyage. Thus, “this journey” in this context is metaphorically used. This is a case where the source is not part of the sentence as it would be in “A is B.”

> *Example7. Sorry you had to join this special sisterhood.

In this example, "join this special sisterhood" is literally used because the breast cancer forum can be considered a special sisterhood, which is an association or unification of women in a common cause. While they are not literally sisters, the term "sisterhood" is almost always used to refer to this kind of association rather than familial relationships.

Not only nouns, but verbs and other content words can be used metaphorically. In this case, usually the metaphorically used words alter the original meaning by breaking selectional restrictions or extending the original meaning.

Example8. I am <u>fighting</u> with pain.

The literal meaning of "fighting" is "take part in a violent struggle involving the exchange of physical blows or the use of weapons." In this example, however, "fighting" is used with "pain," which is not a physical object but a sensation. This usage breaks selectional restrictions on an object of "fight" and extends the meaning of "fight." We can say "fighting" is metaphorically used in this case. Note that "fighting" in this case could be considered as a dead metaphor as well. Please refer to A.10 for more details about dead metaphor.

*Example9. I <u>support</u> your decision

"support" originally means, "Bear all or part of the weight of; hold up," but in the example, it is used emotionally. Likewise, if the original meaning is physical, the extended meaning is merely an emotional version of the original meaning, and it is a common usage of the expression, we consider it as dead metaphor. Please refer to A.10 for more details about dead metaphor.

A.4 Simile

A simile is very similar to a metaphor. We apply the same rules to similes as we do for metaphors. The primary difference is a simile uses words such as "like" or "as" when comparing two things whereas a metaphor does not use those words.

Example10. My sister is like <u>a fox</u>.

A.5 Idiom

An idiom is a set expression of two or more words that means something other than the literal meanings of its individual words.

Example11. kick the bucket, hang ones head, tripped the light fantastic

You can refer to idiom dictionaries on the web in order to determine an idiom. One dictionary available is here: http://idioms.thefreedictionary.com

A.6 Phrasal Verb

Note that we deal with only metaphorical idioms, which have both a literal meaning and a non-literal meaning. Phrasal verbs are not included even though their meanings may not be compositional. For example, in sentence, "I put up with a lot of nonsense," "put up with" is an idiom because it gives different meaning, “endure,” when the three words, “put,” “up,” and “with” are combined. However, since there isnt a compositional (or literal) interpretation of “put up with,” we exclude “put up with” in our coding.

Example12. *The beam holds up the roof.

In this example, “hold up” is literally used as “be the physical support of; carry the weight of.” So, in our scope, “hold up” in this context is not considered to be nonliteral.

A.7 Combined Word

When two or more words are joined together to make them one word, we call it a combined word. Although the origin of the word is metaphorical, we do not consider it to be metaphorical.

Example13. Try to make sure you have some painkillers on hand.

In this example, a “painkiller” does not literally kill pain, so the compound words meaning is non-compositional and could be considered a metaphor. However, we will not consider such compound words to be metaphorical since the word “painkiller” can only refer to analgesic drugs and does not have some other literal meaning.

A.8 Expletive & Discourse Marker

We do not consider words that are empty of propositional content to be metaphorical.

*Example14. Ok ladies, first Happy Thanksgiving to all - even though it may be harder this year to find the gratitude.

In this example, “Ok” is a discourse marker, which is a word or phrase that is relatively syntax-independent and does not change the meaning of the sentence, and has a somewhat empty meaning. Examples of discourse markers are “oh,” “well,” “now,” and “you know.” We dont consider discourse markers a nonliteral language use.

*Example15. Either way, I’m glad you ladies are here so we can share our experiences, and encourage each other.

Blessings

Paula

In this example, “Blessings” is a greeting. We dont consider greetings and parting phrases as nonliteral language use.

A.9 Jargon

When a word is a part of the terminology for the context domain, we do not consider it to be metaphorical, even if it is a metaphor in general domains.

*Example16. If I can find the thread, I will post the link.

In this example, “the thread,” “post,” and “the link” are used differently from their original meanings. However, they are standard internet terminologies. So, we do not consider them as metaphors.

> *Example17. I wish all of you luck and the 2011 December sisters are a great resource.

The literal meaning of “sister” is “A female having the same parents as another or one parent in common with another.” In this example, however, “sister” refers to “A woman fellow member, as of a sorority.” This usage could be seen as metaphor in general domains, but because in this domain (breastcancer.org) this is a extremely common case and the word is a kind of terminology to refer to other people in a group, we do not consider it as metaphor.

A.10 Conventionalized

When the literalness of a word or phrase is unclear, it is assigned to the Conventionalized category. For instance, this can occur when the expression is a dead metaphor. A dead metaphor is a metaphor that has lost the original imagery of its meaning due to extensive, repetitive popular usage. The standards for “dead” are diverse in the language community, but we will consider dead only when we are not sure whether the expression is nonliteral. Although idioms are also conventionalized nonliteral expressions that have lost their original imagery of the meaning, we classify it into the Nonliteral category because it is not confusing that idioms are nonliteral.

Some metaphors are completely dead, in that the phrase only has one meaning, but it originally was part of a metaphor. For example, the word “tulip” in Turkish originally meant a type of turban and was borrowed to refer to the flower due to a physical resemblance. However, this original meaning is completely unknown to most speakers and so the literal meaning is considered the one that refers to the flower. If the original literal meaning is lost or is never used in contemporary English, we would not code it as a metaphor, not even as a dead metaphor.

Other metaphors are very commonly used but the original meaning still persists and the usage can be recognized as a metaphor upon reflection. For example, “support” originally means, “Bear all or part of the weight of; hold up,” but in the sentence, “I support your decision,” it is used emotionally. Some people can say “support” is used metaphorically, but other people may not because this usage of the word is very conventional and does not bring to mind any imagery associated with the original metaphor. In this case, we can say “support” is a dead metaphor in the context.

Appendix B

Post Examples

Post 1 (1) Hello Ladies! (2) I was supposed to start chemo in January, but got an infection in my mastectomy site and had to have surgery to have my TE removed, and am on IV antibiotics. (3) I cant start tx until that is done. (4) So I will be joining you on your journey this month. (5) I AM SICK OF the ANXIETY and WAITING. (6) Ug,. Sorry. (7) BFF you are so right, Hurry up and wait.

(8) I was diagnosed first with dcis in one breast and a spot of it in the left breast as well. (9) I decided on a BMX and that turned out to be rather smart as they found the invasive 1.6 cm tumor in the left side that never showed on the mammogram or ultrasound. (10) I couldn't have an MRI as I have a pacemaker. (11) I have one boobette half filled, one big crater on the left where the TE was removed, feel like sh*t, have a nurse come to the house to give me the antibiotic by IV, and am still waiting for the chemo. (12) I had to have a picc line put in for that, and will still need the port. (13) I would prefer the port to the pic line anyday- nothing dangles from the port. (14) It is real similar (I have heard from my doc and nurses) to my pacemaker placement- that wasnt bad so if you are worried about that, it is one of the easier things. (15) The picc line was porbably almost the same thing, but it is real annoying to have the darn things dangle, even if I put on the cover they give you it slips down and the dangle anyway! (16) LOL.

(17) Thanks ****, for the hopeful words. (18) I know I need to hear there is an end. (19) I will be on TCH as well, sounds pretty much like the tx they want me to have. (20) TC x 6 nad H for the year.

(21) So Ladies, please add another member to this club. (22) Looks like we well all be leaning on each other. (23) But I promise to pick you up if you fall if you can catch me once in a while!

Post 2 (1) Well ****...So far I have been just trying to focus on this one step at a time. (2) I split this journey into 4 stages and I only deal with one. (3) The first was the diagnosis and then the surgery. (4) I had a double lumpectomy as the cancer was in both breasts. (5) Found thru mammo in the right and only because the surgeon insisted on an MRI did we find the cancer in the left. (6) I opted for the lumpectomys instead of the mastectomy and the surgery went really well. (7) So I figure .....ok got thru phase 1. (8) Chemo is going to be phase 2 and I am going to put all my focus and positive thoughts into this phase. (9) If I try and think to far ahead I start to panic. (10) Anyway, we will hold each others hand thru the next few months and we will get thru!!!

(11) So what is your story so far? (12) I see the IDC diagnosis but what have you had done so far and what is your chemo plan?? (13) Sorry if this is too nosy, let me know if I get too personal.

Post 3 (1) February 2011 girl chiming in to lend some support and hope to you February 2012 girls.

(2) I did 6 rounds of TCH (Taxotere, Carboplatin, Herceptin) starting February 24, 2011. (3) It sucked and I'm glad it's over but it was very doable. (4) I have four kids at home (14, 12, and twin 7's) and we maintained a pretty normal life. (5) I felt crappy about 9 days out of each cycle but I just paced myself and prioritized things. (6) I never had mouth sores, was never nauseated. (7) But I did have major diarrhea. (8) Wow. (9) But I was already predisposed to that sort of problem and my onc said the Taxotere just accentuated it. (10) I finish up my year of treatment on Thursday...#17 Herceptin. :) (11) I also did lumpectomy and 33 rounds fo rads. (I hated rads far worse than chemo. Ugh.)

(12) You guys will get through this. (13) The road seems long now but it really moves along fast. (14) Learn all you can and use it to come back and encourage the February 2012 ladies. (15) Chemo sucks but it wasn't nearly as bad as I had feared. :)

Post 4 (1) Saying a prayer for you ****, that your radiation treatments will go well as you move through your breast cancer journey.

Post 5 (1) hi ladies
(2) thought I would come by and say hi...
(3) I am getting ready to celebrate my 1 yr survivorship... and I have decided to put together a video clip to express my gratitude and thanks to God for getting me through this very tough year.
(4) Here is the link... (5) I hope you enjoy it. (6) I have called this video, a new creation. (7) It is my spiritual journey over the past year.

(8) Feel free to pass on to anyone who you think may enjoy this thanksgiving. (9) He really is an amazing God.

(10) God bless each and every one of you ladies.
(11) ****
(12) xox

Post 6 (1) ****, ****, ****, **** and ****, thank you all for your prayers and your loving hearts. (2) I called my doctor's office first thing, spoke with a nurse and they had
ONLY

wanted to tell me that my thyroid is
“normal”.

(3) Hallelujiah!! (4) My weekend was anything but normal.

(5) I took everyone’s advice and stopped worrying, because worry solves nothing and robs us of precious time. (6) And NS had her cat looking out for all the “worries” under my bed and in my closet

(7) so I knew I was safe on that front.

(8) But it was your prayers for me, to help me over these few days that carried me and helped to put the quiet in my heart and allowed me to just leave those worries with God so that I could enjoy the most of each day. (9) I spent yesterday planting two rose bushes, made a lovely meal for my family, spent some quiet time with my soul and my spirit, and had a wonderful conversation with my best friend. (10) Wednesday will be fine as well and I am optimistic that the muga (and all your prayers for this!) will have been fine and I will continue with my treatment.

(11) I would like though to ask you for one more prayer. (12) I know a woman through my son’s hockey. (13) She and I were never close. (14) She has always had an unpleasant personality she was rough, she gossiped, she was very much into money, power and having the best, being the most noticed, speaking only with those she thought could do something for her, disdainful of anyone else. (15) I think we all know people like this unhappy within themselves and looking to make others feel less or feel bad in order to make themselves feel better. (16) There is an emptiness and a void. (17) She and I did not always see eye-to-eye which made it hard (18) I managed the hockey team for the 3 years our sons played together. (19) Two years ago she was dignosed with cancer of the lungs. (20) She was treated and seemed to be doing well. (21) I had had minimal contact with her, but did speak often with her husband. (22) Apparently her personality did not “change” during this cancer journey which made me sad for her. (23) Shortly after I was dx’d with bc, her cancer (sarcoma) returned. (24) Her husband spoke with me often, about chemo, about treatment, about cancer, about life and asked advice on many occasions which I gladly gave. (25) I made peace with the past, I made peace with the person she had been, the hurt she had given me and gave that to the Lord to help me move forward. (26) I reached out to her, though I did not expect that she would reach back. (27) Cancer has a way of making us realize that we are all sisters and brothers, no matter what our personalities, or our differences. (28) But sadly, some people do not change. (29) She is, however, a mother who loves her husband and her sons and within that she does have redeaming qualities, as we all do.

(30) I was talking to my best friend last night (who was our sons’ hockey coach) and I asked after her (he still golfs occasionally with her husband).. (31) I knew from the husband that things were not good and she was going to NY for some treatments that are not available in Canada, but according to my friend, she is now at home, having hospice and has about one month left. (32) She has made her peace with this. (33) My heart goes out to her, to her family and to her children who are 14 and 17. (34) I’m asking if you would join me and include her in your prayers for peace, comfort and a very safe journey.

Post 7 (1) ****,

(2) I know how frustrated you must feel... wanting to be pro active in your cancer experience.

(3) My onc was wearing the“kid gloves” when we were forced to have a relationship 12 yrs ago.

(4) Everything was normal, by the book, during the first five years. (5) Chemo, rads, tamox, recon and hysterectomy.

(6) Appts were graduated each yr, 2 mths, 3 mths, 4 mths, 6 mths... then yearly in the fifth.

(7) Asked her then about screenings, to which she replied.... annual in the office, physical exam and bloodwork.

(8) She told me that I would be the first to know if something were wrong, and I’d know it before any screening would pick it up.

(9) In my 9th yr, she took off those kid gloves and laid it on me. (10) The lump[s] I felt on top of my implant was a recurrence.

(11) Along with the removal of my recon breast, more rads and meds came some shocking statements.

(12) My onc laid it on the line.

(13) Catching a recurrence “early” means sh*t. (14) Once you recur, you no longer can ever be called cured.

(15) “Buying time” was a difficult pair of words to hear.

(16) Yes, new stuff comes along every day to help keep this demon at bay... and that is the best they can do.

(17) I recurred again last year, and I was the one who detected the lumps in my neck.

(18) So ****, be vigilant in your own exams. (19) If you find anything out of the ordinary or of concern... call your doc and have him/her take a look. (20) If warranted, the onc will set up scans or tests at that point.

(21) My onc listens to ME now. (22) I’m not the little innocent gal she took care of 12 yrs ago. (23) She does treat me like an educated consumer and respects my choices and decisions [even thou she voices her opinion].

(24) Good luck to you on your journey.

Post 8 (1) ****:
I just read part of your post above..
and am very glad you have this and it
takes you through.. your journey..
(2) I am a spiritual/religious person as well
(3) We all have our own beliefs and ways..
(4) Hoping that this week will be better
for you.. (5) Hugs and prayers....
and many angels..
(6) ps least you are not
freezing.. like me..
(7) ****

Post 9 (1) I know, the age thing struck me too when I read about your bc journey - we have been going down the same road at the same time, only in another part of the country! (2) It does help to know you are not alone! (3) How amazing with the size of your tumor, that you did not have positive nodes. (4) That is a miracle in itself. (5) I do beleive faith and prayer are our most powerful weapons against this disease. (6) It is what gets me thru each day.

(7) I had a double masectomy because of a strong family history of bc. (8) My onco. recommended it. (9) I think we all handle the news in different ways - for some reason, I had to read everything I could find and go over the pathology reports with a fine-tooth comb - trying to understand them was a major challenge! (10) I do think triple-negative tumors grow very quickly - just wish they knew what fuels them.

(11) Take care,
(12) ****

Post 10 (1) falling off the wagon is no big thing, in my opinion, the psychological good feelings of enjoyment weigh in big for feeling good.

(2) love hummus too...found this great indian bread called naan (?) that has 8 grams of fiber...it is kinda like pita but yummier in my opinion...

(3) love wasa and love popcorn but it gets caught in my teeth to easy.

(4) made pork fried rice without the pork for dinner last night. (5) brown rice, loads of colorful peppers, tofu, spinach, celery, tomatoes tossed in olive oil...at the end put some sweet and sour sauce on it..

(6) really good, even kid friendly...

(7) tina, got my hair ironed after haircut last week... it was so neat! (8) i thought of you...am now looking for the tool to buy and do at home.

(9) got the best haircut ever last week...decided i don't need a shrink, just more trips to the hairdresser...

(10) anyway, i fall off the wagon regularly in small portions, fiber rules, and wonder if we should take the calcium supplement and soy intake discussion to the debate thread...

(11) that was a joke.

(12) love you all just as much as my fiber and george clooney!

(13) no joke...
(14) ****

Post 11 (1) ahhhh a really good haircut is priceless! (2) That and a cleaning lady! (3) Only problem with staying home DH expects me to do it! (4) Yikes!!!
(5) **** falling off is part of this journey, it is stupid to deny yourself everything. (6) A cheat can spur you on! (7) Fiber is the key no doubt, in my life no fiber my body ain't happy!

Post 12 (1) Thanks Chicks!
(2) I straighten my golden locks 2-3 times a week depending on how much product is in it at any given time. (3) I am done chemo just over 1 year and my hair is still as curly as ever.
(4) I love fibre..... (5) I would seriously marry fibre if I could
(6) I am on the wagon so far today
...ongoing battle.
(7) I'm going to make the WW low fat brownie mix/all bran/muffins today.
(8) I had gorgeous orange roughy fish for supper last night.
(9) Have a good day, ****

Post 13 (1) **** - hope you stay on the wagon, or at least get back on after you fall!
(2) Tried the pumpkin in the brownie mix. (3) It was better than the black beans. (4) The result is a bit gelatiounus (sp?), don't know that I'll try that again. (5) My half jar of prunes were moldy so I didn't blender them into the pumpkin as I'd been considering. (6) All things considered, applesauce may be the best substitute. (7) I think I've sacrificed enough brownie mixes.
(8) Checked out the microwave popcorn. (9) My goodness, things have changed since I last went looking for it. (10) I like those mini-bags.

Post 14 (1) A SPECIAL POEM FOR A SPECIAL FRIEND AND SISTER ****
(2) This is the journey and this is the road,
(3) Helping each other and sharing the load,
(4) Walking together and seeking the way,
(5) Sharing the wonder and seizing the day.
(6) Making discoveries,lessons to learn,
(7) New exploration as seasons return,
(8) Finding a way to forget and forgive,
(9) Counting our blessings and learning to live.
(10) This is the journey and now is the time ,
(11) New paths to follow and mountains to climb,
(12) Travelling onward,enjoying each mile,
(13) Rejoice in the moment,the journey worthwhile...****
(14) Almighty God,giver of health and healing: Grant to this thy
servant,****,such a sense of thy presence that she may have perfect trust in thee. In all her suffering may she cast her care upon thee,so that, enfolded in thy love and power,she may receive from thee health and salvation according to thy gracious will ; through Jesus Christ

our Lord. Amen
(15) I wish you peace my friend may the Grace of God give you comfort .
(16) Jesus said , “NEVER WILL I LEAVE YOU ; NEVER WILL I FORSAKE YOU .” (HEBREWS 13:5)
(17) Love and warm thoughts.****

Post 15 (1) Hi everyone... (2) I just have a moment to pop in because I am feeling so dang good that I’m trying to get stuff done around the house today...yippie! (3) I tried my chemochicks.com turban again, this time I got it to work. (4) Wore it with the matching T shirt and a large necklace I had already. (5) 13 y.o. stepson said “oh, that looks tight! That means you look good!” (6) Husband also was a big fan of the turban. (7) I guess I can be a fashion diva with the turban after all!

(8) Thought someone might like a write up of the words to the survivor movie. (9) If you haven’t seen the website, go check it out! thesurvivormovie.com
(10) I have cancer, but cancer does not have me.
(11) Cancer is not who I am.
(12) It is only a bend in the road that is my life?s journey.
(13) An unexpected detour on my path.
(14) It is a lesson in the cosmic schoolroom that is human existence.
(15) So I will pause to rest? and heal? and study the lesson?
before I move on to my life beyond cancer.
(16) I will not give into fear, and I will not be discouraged by setbacks.
(17) Setbacks are only opportunities to review the lesson.
(18) I will not be ashamed of my scars.
(19) Scars are the brushstrokes in the masterpiece that is my life.
(20) I will be thankful for the many blessings cancer
has brought into my life?
(21) People I never would have known.
(22) Love that I had never been still or quiet enough to witness.
(23) Humility I needed,
(24) Strength I thought I had lost,
(25) Courage I never knew I had.
(26) I will remember that I can still have fun and that it?s okay
- even healthy! - to be silly.
(27) I will remember that to find the joy in rainbows,
(28) I must endure the rain.
(29) And I will always remember that?
(30) While I have cancer,
(31) Cancer does not have me.
(32) Hope everyone is well. I FEEL SO GOOD TODAY!

Bibliography

Dan Assaf, Yair Neuman, Yohai Cohen, Shlomo Argamon, Newton Howard, Mark Last, Ophir Frieder, and Moshe Koppel. Why dark thoughts aren't really dark: A novel algorithm for metaphor identification. In *Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2013 IEEE Symposium on*, pages 60–65. IEEE, 2013. 2.3.1.1

Collin F Baker, Charles J Fillmore, and John B Lowe. The berkeley framenet project. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, pages 86–90. Association for Computational Linguistics, 1998. 6.2.1, 9.2.1.1

Beata Beigman Klebanov and Eyal Beigman. A game-theoretic model of metaphorical bargaining. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 698–709. Association for Computational Linguistics, 2010. 2.3.2

Beata Beigman Klebanov and Michael Flor. Argumentation-relevant metaphors in test-taker essays. *Meta4NLP 2013*, pages 11–20, 2013. 2.2

Beata Beigman Klebanov, Eyal Beigman, and Daniel Diermeier. Discourse topics and metaphors. In *Proceedings of the Workshop on Computational Approaches to Linguistic Creativity*, pages 1–8. Association for Computational Linguistics, 2009. 6.1

Beata Beigman Klebanov, Chee Wee Leong, Michael Heilman, and Michael Flor. Different texts, same metaphors: Unigrams and beyond. *ACL 2014*, page 11, 2014. (document), 2.3.2, 5.5.4, 5.6, 6.1, 6.3, 6.4.3, 6.2, 6.5.1

Beata Beigman Klebanov, Chee Wee Leong, and Michael Flor. Supervised word-level metaphor detection: Experiments with concreteness and reweighting of examples. *NAACL HLT 2015*, page 11, 2015. 2.3.1.1

Julia Birke and Anoop Sarkar. A clustering approach for nearly unsupervised recognition of nonliteral language. In *EACL*, 2006. 2.3.1.1, A

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, March 2003. ISSN 1532-4435. URL `http://dl.acm.org/citation.cfm?id=944919.944937`. 6.2.1

George Aaron Broadwell, Umit Boz, Ignacio Cases, Tomek Strzalkowski, Laurie Feldman, Sarah Taylor, Samira Shaikh, Ting Liu, Kit Cho, and Nick Webb. Using imageability and topic chaining to locate metaphors in linguistic corpora. In *Social Computing, Behavioral-Cultural Modeling and Prediction*, pages 102–110. Springer, 2013. 2.3.1.1, 2.3.1.3, 6.1

Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior research methods*, 46(3):904–911, 2014. 6.2.2

Luana Bulat, Stephen Clark, and Ekaterina Shutova. Modelling metaphor with attribute-based semantics. *EACL 2017*, page 523, 2017. (document), 2.3.1.1, 2.2

Lynne Cameron. *Metaphor in educational discourse*. A&C Black, 2003. 2.1

Ronald Carter. *Language and creativity: The art of common talk*. Psychology Press, 2004. 2.1

Nathanael Chambers and Dan Jurafsky. Unsupervised learning of narrative schemas and their participants. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, pages 602–610. Association for Computational Linguistics, 2009. 8.3.2

Jonathan Charteris-Black. *Corpus approaches to critical metaphor analysis*. Palgrave-MacMillan, 2004. 2.1

J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46, 1960. 3.6.1

Jacob Cohen. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, 70(4):213, 1968. 3.6.1, 4.3.2

Scott A Crossley and Danielle S McNamara. Cohesion, coherence, and expert evaluations of writing proficiency. In *Proceedings of the 32nd annual conference of the Cognitive Science Society*, pages 984–989, 2010. 7.2.1

Dipanjan Das, Desai Chen, André FT Martins, Nathan Schneider, and Noah A Smith. Frame-semantic parsing. *Computational Linguistics*, 40(1):9–56, 2014. 6.2.1

Hal Daume III. Frustratingly easy domain adaptation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 256–263, Prague, Czech Republic, June 2007. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/P07-1033`. 7.2.2

Alice Deignan. The evaluative properties of metaphors. *Researching and applying metaphor in the real world*, 357, 2010. 2.1

Lynn Fainsilber and Andrew Ortony. Metaphorical uses of language in the expression of emotions. *Metaphor and Symbol*, 2(4):239–250, 1987. 4.3.1

Dan Fass. met*: A method for discriminating metonymy and metaphor by computer. *Computational Linguistics*, 17(1):49–90, 1991. 2.3.1.1

Gilles Fauconnier and Mark Turner. Conceptual integration networks. *Cognitive science*, 22(2): 133–187, 1998. 1.2

Anna Feldman and Jing Peng. An approach to automatic figurative language detection: A pilot study. In *Proceedings of the Corpus-Based Approaches for Figurative Language Colloquium, Liverpool, UK*, volume 22, pages 25–42, 2009. A

Anna Feldman and Jing Peng. Automatic detection of idiomatic clauses. In *Computational Linguistics and Intelligent Text Processing*, pages 435–446. Springer, 2013. A

Charles J Fillmore. Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the origin and development of language and speech*, volume 280, pages 20–32, 1976. 1.2

Joseph L Fleiss. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378–382, 1971. 3.6.1, 5.5.2, 6.3

Lisa Gandy, Nadji Allan, Mark Atallah, Ophir Frieder, Newton Howard, Sergey Kanareykin, Moshe Koppel, Mark Last, Yair Neuman, and Shlomo Argamon. Automatic identification of conceptual metaphors with limited knowledge. In *AAAI*, 2013. 2.3.1.1

Matt Gedigian, John Bryant, Srini Narayanan, and Branimir Ciric. Catching metaphors. In *Proceedings of the Third Workshop on Scalable Natural Language Understanding*, pages 41–48. Association for Computational Linguistics, 2006. 2.3.1.1

James Paul Gee. *An introduction to discourse analysis: Theory and method*. Routledge, fourth edition, 2014. 1, 1.2

Alex Gillespie and Flora Cornish. Intersubjectivity: Towards a dialogical analysis. *Journal for the theory of social behaviour*, 40(1):19–46, 2010. 9.2.2

Andrew Goatly. *Language of Metaphors: Literal Metaphorical*. Routledge, 1997. 2.3.1.3

Jonathan Gratch, Ron Artstein, Gale Lucas, Giota Stratou, Stefan Scherer, Angela Nazarian, Rachel Wood, Jill Boberg, David DeVault, Stacy Marsella, et al. The distress analysis interview corpus of human and computer interviews. In *Proceedings of Language Resources and Evaluation Conference (LREC)*, 2014. (document), 4.2, 4.1

HP Grice. Logic and conversation'in p. cole and j. morgan (eds.) syntax and semantics volume 3: Speech acts, 1975. 1

Zellig S Harris. Distributional structure. *Word*, 10(2-3):146–162, 1954. 8.3.3

Ilana Heintz, Ryan Gabbard, Mahesh Srinivasan, David Barner, Donald S Black, Marjorie Freedman, and Ralph Weischedel. Automatic extraction of linguistic metaphor with lda topic modeling. In *Proceedings of the First Workshop on Metaphor in NLP*, pages 58–66, 2013. 2.3.1.2

Dirk Hovy, Shashank Srivastava, Sujay Kumar Jauhar, Mrinmaya Sachan, Kartik Goyal, Huiying Li, Whitney Sanders, and Eduard Hovy. Identifying metaphorical word use with tree kernels. *Meta4NLP 2013*, page 52, 2013. 2.3.1.1

Ting-Hao Kenneth Huang. Social metaphor detection via topical analysis. In *Sixth International Joint Conference on Natural Language Processing*, page 14, 2014. 2.3.1.1

Hyeju Jang and Jack Mostow. Inferring selectional preferences from part-of-speech n-grams. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 377–386. Association for Computational Linguistics, 2012. 1.3

Hyeju Jang, Mario Piergallini, Miaomiao Wen, and Carolyn Penstein Rosé. Conversational metaphors in use: Exploring the contrast between technical and everyday notions of metaphor. *ACL 2014*, page 1, 2014. 5.1, 6.3

Mario Jarmasz and Stan Szpakowicz. Not as easy as it seems: Automating the construction of lexical chains using rogets thesaurus. In *Advances in Artificial Intelligence*, pages 544–549. Springer, 2003. 6.2.1

Yohan Jo and Alice Oh. Aspect and Sentiment Unification Model for Online Review Analysis. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 815–824, 2011. 7.2.1

Amy Johnson. The link between focusing on yourself and depression, 2003. URL http://dramyjohnson.com/2013/06/the-link-between-focusing-on-yourself-and-depression/. 4.1

Anna Albertha Kaal. *Metaphor in conversation*. Vrije Universiteit, 2012. 2.1

Zornitsa Kozareva. Multilingual affect polarity and valence prediction in metaphor-rich texts. In *ACL (1)*, pages 682–691. Citeseer, 2013. 2.3.2

Saisuresh Krishnakumaran and Xiaojin Zhu. Hunting elusive metaphors using lexical resources. In *Proceedings of the Workshop on Computational approaches to Figurative Language*, pages 13–20. Association for Computational Linguistics, 2007. 2.3.1.1, 2.3.1.1

George Lakoff and M. Johnson. Metaphors we live by. *Chicago/London*, 1980. 1, 1, 2.1, 2.3.1

Heidi Levitt, Yifaht Korman, and Lynne Angus. A metaphor analysis in treatments of depression: Metaphor as a marker of change. *Counselling Psychology Quarterly*, 13(1):23–35, 2000. 2.1, 4.1

Hongsong Li, Kenny Q Zhu, and Haixun Wang. Data-driven metaphor recognition and explanation. *Transactions of the Association for Computational Linguistics*, 1:379–390, 2013. 2.3.1.1

Linlin Li and Caroline Sporleder. Using gaussian mixture models to detect figurative language in context. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '10, pages 297–300, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics. ISBN 1-932432-65-5. URL http://dl.acm.org/citation.cfm?id=1857999.1858038. A

Paula Liendo. Business language: A loaded weapon? war metaphors in business. *Invenio: Revista de investigación académica*, (6):43–50, 2001. 2.1

Chih-Jen Lin. Projected gradient methods for nonnegative matrix factorization. *Neural computation*, 19(10):2756–2779, 2007. 8.3.3

Dekang Lin. Automatic identification of non-compositional phrases. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 317–324. Association for Computational Linguistics, 1999. A

Graham Low, Alice Deignan, Lynne Cameron, and Zazie Todd. *Researching and applying metaphor in the real world*, volume 26. John Benjamins Publishing, 2010. 2.1

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, 2014. URL http://www.aclweb.org/anthology/P/P14/P14-5010. 6.2.2, 8.3.2

Zachary J Mason. Cormet: a computational, corpus-based conventional metaphor extraction system. *Computational Linguistics*, 30(1):23–44, 2004. (document), 2.3.1.1, 2.2, A

Elijah Mayfield and Carolyn Rosé. An interactive tool for supporting error analysis for text mining. In *Proceedings of the NAACL HLT 2010 Demonstration Session*, pages 25–28. Association for Computational Linguistics, 2010. 5.5.4, 6.4.4, 7.3.1, 8.4.2

Linda M McMullen. Methods for studying the use of novel figurative language in psychotherapy. *Psychotherapy: Theory, Research, Practice, Training*, 22(3):610, 1985. 4.1

Marvin Minsky. A framework for representing knowledge. 1975. 1.2

Jeffery Scott Mio. Metaphor and politics. *Metaphor and Symbol*, 12(2):113–133, 1997. 2.1, EX(14)

Saif M Mohammad, Ekaterina Shutova, and Peter D Turney. Metaphor as a medium for emotion: An empirical study. *c2016 The* SEM 2016 Organizing Committee. All papers c2016 their respective authors. This proceedings volume and all papers therein are licensed under a Creative Commons Attribution 4.0 International License. License details: http://creativecommons.org/licenses/by/4.0*, page 23, 2016. 2.3.2

Michael Mohler, David Bracewell, David Hinote, and Marc Tomlinson. Semantic signatures for example-based linguistic metaphor detection. In *Proceedings of the First Workshop on Metaphor in NLP*, pages 27–35, 2013. 2.3.1.2

Jane Morris and Graeme Hirst. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Comput. Linguist.*, 17(1):21–48, March 1991. ISSN 0891-2017. URL `http://dl.acm.org/citation.cfm?id=971738.971740`. 6.2.1

Yair Neuman, Dan Assaf, Yohai Cohen, Mark Last, Shlomo Argamon, Newton Howard, and Ophir Frieder. Metaphor identification in large texts corpora. *PloS one*, 8(4):e62343, 2013. 2.3.1.1

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. 8.3.3

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, 2014. URL `http://www.aclweb.org/anthology/D14-1162`. 8.3.4

Xuan-Hieu Phan and Cam-Tu Nguyen. Gibbslda++: Ac/c++ implementation of latent dirichlet allocation (lda), 2007. 6.2.1

Pragglejaz-Group. Mip: A method for identifying metaphorically used words in discourse. *Metaphor and symbol*, 22(1):1–39, 2007. (document), 2.2, 2.1, 3.6.2, A

Ashequl Qadir and Ellen Riloff. Bootstrapped learning of emotion hashtags# hashtags4you. In *Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis*, pages 2–11, 2013. 8.3

Ashequl Qadir, Ellen Riloff, and Marilyn A Walker. Automatically inferring implicit properties in similes. In *Proceedings of NAACL-HLT*, pages 1223–1232, 2016. 2.3.3

Sophia Rabe-Hesketh and Anders Skrondal. *Multilevel and longitudinal modeling using Stata*. STATA press, 2008. 5.4.2

Sophia Rabe-Hesketh, Anders Skrondal, and Andrew Pickles. Gllamm manual. 2004. 5.4.2

Gary M Reisfield and George R Wilson. Use of metaphor in the discourse on cancer. *Journal of Clinical Oncology*, 22(19):4024–4027, 2004. 1.2, 3.3

Steffen Remus and Chris Biemann. Three knowledge-free methods for automatic lexical chain extraction. In *HLT-NAACL*, pages 989–999, 2013. 6.2.1

Philip Resnik. Selectional preference and sense disambiguation. In *Proceedings of the ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How*, pages 52–57. Washington, DC, 1997. 2.3.1.1

Ellen Riloff, Rosie Jones, et al. Learning dictionaries for information extraction by multi-level bootstrapping. In *AAAI/IAAI*, pages 474–479, 1999. 8.3

Ellen Riloff, Janyce Wiebe, and Theresa Wilson. Learning subjective nouns using extraction pattern bootstrapping. In *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4*, pages 25–32. Association for Computational Linguistics, 2003. 8.3

SL. David Ritchie. *Metaphor (Key Topics in Semantics and Pragmatics)*. Cambridge university press, 2013. 1.2, 2.1, 3.3, 3.3, 5.1

Peter Mark Roget. *Roget's Thesaurus of English Words and Phrases...* TY Crowell Company, 1911. 6.2.1

Roger C Schank and Robert P Abelson. *Scripts, plans, and knowledge*. Yale University, 1975. 1.2

Marc Schulder and Eduard Hovy. Metaphor detection through term relevance. *ACL 2014*, page 18, 2014. 2.3.1.3

Ella Freeman Sharpe. Psycho-physical problems revealed in language: An examination of metaphor. *The International Journal of Psycho-Analysis*, 21:201, 1940. 4.1

Ekaterina Shutova. *Computational approaches to figurative language*. PhD thesis, University of Cambridge, 2011. 2.3.1.3

Ekaterina Shutova. Metaphor identification as interpretation. *Atlanta, Georgia, USA*, page 276, 2013. A

Ekaterina Shutova and Lin Sun. Unsupervised metaphor identification using hierarchical graph factorization clustering. In *HLT-NAACL*, pages 978–988, 2013. 2.3.1.1

Ekaterina Shutova and Simone Teufel. Metaphor corpus annotated for source-target domain mappings. In *LREC*, 2010. 2.2, 2.2

Ekaterina Shutova, Lin Sun, and Anna Korhonen. Metaphor identification using verb and noun clustering. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 1002–1010. Association for Computational Linguistics, 2010. 2.3.1.1, 2.3.1.1, A

Ekaterina Shutova, BarryJ. Devereux, and Anna Korhonen. Conceptual metaphor theory meets the data: a corpus-based human annotation study. *Language Resources and Evaluation*, 47

(4):1261–1284, 2013. ISSN 1574-020X. doi: 10.1007/s10579-013-9238-z. URL `http://dx.doi.org/10.1007/s10579-013-9238-z`. 2.2

Ekaterina Shutova, Douwe Kiela, and Jean Maillard. Black holes and white rabbits: Metaphor identification with visual features. In *Proc. of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 160–170, 2016. 2.3.1.1

Dan Sperber, Deirdre Wilson, ł, and . *Relevance: Communication and cognition*, volume 1. Citeseer, 1986. 1

Caroline Sporleder and Linlin Li. Unsupervised recognition of literal and non-literal use of idiomatic expressions. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 754–762. Association for Computational Linguistics, 2009. 2.3.1.3, 6.1, 6.2.1, A

Jane M Spurr and Lusia Stopa. Self-focused attention in social phobia and social anxiety. *Clinical psychology review*, 22(7):947–975, 2002. 4.1

Mike Startup, Lebogang Makgekgenene, and Rosemary Webster. The role of self-blame for trauma as assessed by the posttraumatic cognitions inventory (ptci): a self-protective cognition? *Behaviour research and therapy*, 45(2):395–403, 2007. 4.1

Gerard J Steen, Aletta G Dorst, J Berenike Herrmann, Anna Kaal, Tina Krennmayr, and Trijntje Pasma. *A method for linguistic metaphor identification: From MIP to MIPVU*, volume 14. John Benjamins Publishing, 2010. 2.2, 2.2, 6.3

Deborah A Stone. *Policy paradox and political reason*. Addison-Wesley Longman, 1988. 2.3.2

Tomek Strzalkowski, Samira Shaikh, Kit Cho, George Aaron Broadwell, Laurie Feldman, Sarah Taylor, Boris Yamrom, Ting Liu, Ignacio Cases, Yuliya Peshkova, et al. Computing affect in metaphors. *ACL 2014*, page 42, 2014. 2.3.1.3, 2.3.2, 2.3.2

Ernesto Suárez Toste. Metaphor inside the wine cellar: On the ubiquity of personification schemas in winespeak. *Metaphorik. de*, 12(1):53–64, 2007. 2.1

Deborah Tannen. *Framing in discourse*. Oxford University Press, 1993. 1.2

Deborah Tannen and Cynthia Wallat. Interactive frames and knowledge schemas in interaction: Examples from a medical examination/interview. *Social Psychology Quarterly*, pages 205–216, 1987. 1.2

Yla R Tausczik and James W Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54, 2010. 2.3.2

John D Teasdale. Negative thinking in depression: Cause, effect, or reciprocal relationship? *Advances in Behaviour Research and Therapy*, 5(1):3–25, 1983. 4.1

Yulia Tsvetkov, Elena Mukomel, and Anatole Gershman. Cross-lingual metaphor detection using common semantic features. pages 45–51, 2013. 2.3.1.1, 2.3.1.1

Yulia Tsvetkov, Leonid Boytsov, Anatole Gershman, Eric Nyberg, and Chris Dyer. Metaphor detection with cross-lingual model transfer. pages 248–258, 2014. (document), 2.3.1.1, 2.3.1.1, 6.1, 6.4.3, 6.2, 6.5.1

Peter D Turney, Yair Neuman, Dan Assaf, and Yohai Cohen. Literal and metaphorical sense identification through concrete and abstract context. In *Proceedings of the 2011 Conference on the Empirical Methods in Natural Language Processing*, pages 680–690, 2011. 2.3.1.1, 6.2.2, A

AM Wallington, JA Barnden, P Buchlovsky, L Fellows, and SR Glasbey. Metaphor annotation: A systematic study. *COGNITIVE SCIENCE RESEARCH PAPERS-UNIVERSITY OF BIRMINGHAM CSRP*, 2003. 2.2, 3.4.2

David Watson, Lee A Clark, and Greg Carey. Positive and negative affectivity and their relation to anxiety and depressive disorders. *Journal of abnormal psychology*, 97(3):346, 1988. 4.1

Carl F Weems, Armando A Pina, Natalie M Costa, Sarah E Watts, Leslie K Taylor, and Melinda F Cannon. Predisaster trait anxiety and negative affect predict posttraumatic stress in youths after hurricane katrina. *Journal of consulting and clinical psychology*, 75(1):154, 2007. 4.1

Adrian Wells and Costas Papageorgiou. Social phobia: Effects of external attention on anxiety, negative beliefs, and perspective taking. *Behavior therapy*, 29(3):357–370, 1998. 4.1

Miaomiao Wen and Carolyn Penstein Rosé. Understanding participant behavior trajectories in online health support groups using automatic extraction methods. In *Proceedings of the 17th ACM international conference on Supporting group work*, pages 179–188. ACM, 2012. 5.2, 5.4, 5.4.1

Miaomiao Wen, Zeyu Zheng, Hyeju Jang, Guang Xiang, and Carolyn Penstein Rosé. Extracting events with informal temporal references in personal histories in online communities. In *ACL (2)*, pages 836–842, 2013. 5.3, 5.4

Yukun Zhu, Ryan Kiros, Richard Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In *arXiv preprint arXiv:1506.06724*, 2015. 8.4.1