Metaphor and Understanding: The Work of Lakoff and Johnson and Natural Language Processing

Karen M. Jakubowsky

College of Saint Benedict/Saint John's University

Follow this and additional works at: https://digitalcommons.csbsju.edu/honors_theses

Part of the Computer Sciences Commons

Recommended Citation
https://digitalcommons.csbsju.edu/honors_theses/720

Available by permission of the author. Reproduction or retransmission of this material in any form is prohibited without expressed written permission of the author.
METAPHOR AND UNDERSTANDING:
THE WORK OF LAKOFF AND JOHNSON AND NATURAL
LANGUAGE PROCESSING

A THESIS
The Honors Program
College of St. Benedict/St. John's University

In Partial Fulfillment
of the Requirements for "Departmental Distinction"
and the Degree Bachelor of Arts
in the Department of Computer Science

by
Karen Jakubowsky
April, 1999
PROJECT TITLE:

*Metaphor and Understanding:*
*The Work of Lakoff and Johnson and Natural Language Processing*

Approved by:

[Signature]
Project Advisor, Assistant Professor of Computer Science

[Signature]
Reader, Associate Professor of Computer Science

[Signature]
Reader, Associate Professor of Psychology

[Signature]
Chair, Department of Computer Science

Director, Honors Thesis Program

[Signature]
Director, Honors Program
INTRODUCTION

Imagine yourself back in elementary school. You are sitting in the front row as your favorite teacher, with perfect penmanship, is writing one word on the board: Metaphor. Do you remember what this first lesson was like? Most likely, you were introduced to metaphor as not much more than a literary device to “spice up” your writing. We learned that it is a comparison of one thing to another, to give a more descriptive account of something. “He was an iceberg” or “time is money” are obvious metaphors. They are regarded as simply vivid expressions to explain our ideas. The comparison to an “iceberg,” for instance, is not to be taken literally. It is used to show that someone is unfeeling or unemotional. In recent years, however, ideas about metaphor have changed from our elementary school days. George Lakoff, professor of linguistics at the University of California, Berkeley, and Mark Johnson, professor of philosophy at the University of Oregon, assert that metaphors are much more than literary devices. In their 1980 book, Metaphors We Live By, they claim that the way we think, act, and understand our world is rooted in metaphor (3). Their theory represents a shift in the focus of metaphor, taking it from merely poetic substitution to a bold statement regarding human understanding. The theory has opened a new door, causing an entire field of metaphor research to develop.

Clearly, this change in how we view metaphor is important to cognitive psychology and perhaps even to our old elementary school teacher, but its impact and importance does not end there. Its influence spans several different fields, most notably that of artificial intelligence (AI). The theory provides an important application to one of the main facets of AI: natural language processing. It could profoundly affect this area for two main reasons. First, it provides a theory that encompasses broad understanding and context, and second, it can be computationally modeled. At the same time, a growing field fosters numerous opinions, and there are many fundamental issues about metaphor that remain unresolved. There are disagreements, or in some cases a lack of discussion, on several key points. Therefore, we find two sides to the partnership of metaphor and artificial intelligence. Lakoff and Johnson’s theory of metaphor is important and has the potential to be a major force in natural language processing; however, its true impact is and will continue to be limited until a number of issues are resolved.
BACKGROUND

Before delving into the applicability of Lakoff and Johnson’s theory to artificial intelligence through natural language processing, we need to first understand the theory itself. Beginning with the ideas of Aristotle, we trace some of the early theories of metaphor, turning points in metaphor history paving the way for Lakoff and Johnson, and finally, the theory of Lakoff and Johnson and its key points.

Traditional Views

The roots of the study of metaphor can be traced back to the philosopher Aristotle. He describes metaphor as a substitution of one term for another, and provides us with one of the first models of metaphor (The Metaphor Home Page). His model is based on strict categorization of genus and species. A genus is a large, broad category, and its species are more specific instances of that genus. A metaphor is seen as a transfer of a name of a genus to one of its species, or a species to its genus, a species to another species, and so forth. Aristotle uses “Here stands my ship” as an instance of metaphor because it comes from substituting “to be at anchor” with the more generic term (its genus) “to stand still” (The Metaphor Home Page). Many traditional computational models of metaphor are based on Aristotle’s ideas. His notions on the purpose for metaphor parallel our own beginning introduction: to achieve a more “colourful expression” by replacing one idea with another (The Metaphor Home Page). Philosopher David E. Cooper echoes this thought in an Aristotelian Society Series book on metaphor when he says:

I may utter a metaphor to stimulate an image, or to provoke an interesting comparison, or to register a beautiful turn of phrase.... When a speaker draws on a systematic established metaphor, there is likely to be no reason why he did this in preference to employing an alternative, metaphor-free, vocabulary (140).

John Searle, a prominent philosopher, continues the theme that a metaphor is simply another way to say what you mean. To define metaphor, he divides meaning into two categories: the speaker utterance meaning, and the word or sentence meaning. The speaker utterance meaning is the speaker’s intended meaning, while the sentence meaning is what the words of the sentence actually mean (93). A metaphor, he states, is when the speaker utterance meaning is different from the word or sentence meaning, as illustrated by the diagram below:
Both of these diagrams are expressing the speaker utterance meaning “S is R.” In the first diagram, the sentence and speaker utterance meaning are the same (P is the same as R), so no metaphors are used. In the second, the speaker uses P (a metaphor) to express that “S is R.” For example, “He was an iceberg” is a way to express that someone is cold and unfeeling.

Many other philosophers, cognitive scientists, and linguists express similar, though varied views. No matter how varied, however, many metaphor theories are built upon this same basis of metaphor: substituting one thing for another. While praising metaphor and its uses, it was thought by many as merely a choice of language and that we could just as easily express our thoughts without its use.

Turning points

Later, some began to take a different approach to metaphor. Philosophers Ivor Richards and Max Black, for instance, assert an “Interaction View” on metaphor. In this view, metaphors not only describe one thing in terms of another, but the actual conceptual meanings of these two things are changed as well (The Metaphor Home Page). For instance, take the metaphor, “Man is Wolf.” When we use this metaphor, our concept of man becomes closer to a wolf, and our concept of wolf becomes closer to a man, according to Black and Richards (The Metaphor Home Page). This represents a change from the idea that a metaphor is simply a literary device, with no impact on our cognitive structure. One philosopher, Nietzsche, went so far as to claim that metaphors are the root of all knowledge. They give us a process that we use to interact in our world, he says. Little attention, however, was given to Nietzsche’s ideas, and they were dismissed as part of the romanticism of the period (Johnson 15).
Lakoff and Johnson

Lakoff and Johnson, with their book called *Metaphors We Live By*, revived Nietzsche’s ideas and solidified a major shift in the study of metaphor, claiming that metaphors shape the entire way we think and act. Our everyday functioning, they state, is based upon metaphor (3). One of the many examples they use throughout their book to illustrate this is the metaphor “argument is war.” They give some of the following sentences to illustrate that war is a metaphor for argument in our everyday speech:

He attacked every weak point in my argument.
His criticisms were right on target.
I’ve never won an argument with him.
I demolished his argument (4).

At first glance, these sentences don’t seem to have metaphors at all. Lakoff and Johnson, however, claim that a closer look at the sentences, which use words such as “attacked” and “demolished” to talk about arguments, reveal that they are based in the “argument is war” metaphor. In addition, their key thesis is not just that we talk about argument in terms of war, but that the way we think about arguments, and our actions in an argument are structured around this metaphor:

We can actually win or lose arguments. We see the person we are arguing with as an opponent. We attack his position and defend our own…. Many of the things we actually do in arguing are partially structured by the concept of war (4).

Similarly, the metaphor “time is money” structures our everyday dealings with time as illustrated by the sentences, “How do you spend your time these days?” and “I lost a lot of time when I got sick.” The metaphor “time is money” partially structures our thought about time into something that can be quantified like money—something we can spend, use, and lose (8).

The metaphorical examples above are what Lakoff and Johnson term structural metaphors, where one concept (e.g. argument) is “metaphorically structured in terms of another” (e.g. war) (14). However, Lakoff and Johnson expand their theory to include other types of metaphors, namely orientational and ontological metaphors. Orientational metaphors, instead of structuring one concept in terms of another, structure a “whole system of concepts with respect to one another,” usually in terms of space (14). An example is the “happy is up” and “sad is down” metaphors, illustrated by some of the following sentences:
That boosted my spirits.
I'm feeling up.
I fell into a depression.
He's really low these days (15).

Again, our fundamental concepts are structured through metaphor. Lakoff and Johnson claim that it is this “happy is up” metaphor that accounts for why we would never think of “I'm feeling up” as being happy while at the same time “my spirits rose” as being sad (18). In addition, ontological metaphors also structure how we think about and understand certain concepts. In an ontological metaphor, we refer to a concept in terms of a physical object or substance (25). For example, “the mind is an entity” is a common ontological metaphor, further elaborated in the “mind is a machine” and the “mind is a brittle object” metaphors:

My mind is just not operating today.
We're still trying to grind out the solutions to this equation.

He broke under cross-examination.
His mind snapped (27-28)

Both of these two metaphors, Lakoff and Johnson argue, structure the way we view and understand the mind. We treat it both as something very fragile, and as having an on-off state with a level of productivity (28).

Through this brief introduction, we find that Lakoff and Johnson hold a very deviant stance on metaphor in relation to traditional theories. Although their three types of metaphor (structural, orientational and ontological) are very different from each other, they all have one unifying element: a claim that metaphors are much more than literary devices, structuring the way we think and act and forming the foundation for our conceptual system.

LAKOFF AND JOHNSON AND ARTIFICIAL INTELLIGENCE

Lakoff and Johnson clearly offer a very different view on metaphor, taking it from a literary technique to the basis of our understanding. What impact could their ideas have on artificial intelligence? Perhaps most obviously, one of the main goals of artificial intelligence is for a computer to “understand” as a human would. If there is some truth to their theory, then
metaphor is clearly important to artificial intelligence. At a deeper level, we find two things that solidify the importance of their theory to AI. First, it represents a fundamental shift from current research into computers and understanding within the field of natural language processing. By taking into account broad understanding and context, it counteracts many of the current problems of the field. Second, it provides the means to implement this shift through the characteristics that allow it to be computationally modeled. These two things, coupled together, could provide a powerful tool for artificial intelligence.

Computers and Understanding

Terry Winograd of the Department of Computer Science at Stanford University, and Fernando Flores, President of Logonet, Inc. and Chairman of Action Technologies, claim that most attempts in AI at programming computers to have understanding have failed. These attempts are the focus of natural language processing, the goal of which is for computers to understand and respond to human language. In their book, Understanding Computers and Cognition, they assert that most natural language processing systems operate on a specific domain that does not take into account broad understanding or context. Using programs such as ELIZA, SHRDLU and others as examples, they show how these systems respond to a “fixed set of patterns” anticipated by the programmer (121). According to Winograd and Flores, this is not understanding. The theory of Lakoff and Johnson, in contrast, offers an alternative to these criticisms. Their theory of metaphor is a broad one that captures understanding in our entire conceptual system, an idea that has been supported by others in the field. In addition, it takes context into account when they emphasize that we each have our own metaphors we live by, within our culture and within ourselves as individuals. Overall, the field of natural language processing is moving toward an appreciation of general understanding and specific context, and Lakoff and Johnson’s “metaphors we live by” embody this important shift.

One of the examples Winograd and Flores use to demonstrate the problems of computers and understanding is the program named ELIZA. Developed by Joseph Weizenbaum at M.I.T. in 1966, ELIZA mimics a non-directive psychotherapist, whose goals are to reflect a patient’s feelings back onto the patient in order to help the patient understand and deal with those feelings
(Sternberg 485). A sample of ELIZA’s interaction with a patient is shown below (ELIZA’s responses are shown in capital letters):

Men are all alike.
IN WHAT WAY?
They’re always bugging us about something or other.
CAN YOU THINK OF A SPECIFIC EXAMPLE?
Well, my boyfriend made me come here.
YOUR BOYFRIEND MADE YOU COME HERE?
He says I am depressed much of the time.
I AM SORRY TO HEAR THAT YOU ARE DEPRESSED (486).

In the above dialogue, ELIZA may seem to be a sympathetic and understanding therapist, when in fact she simply uses keywords to pick out appropriate responses. For example, the keywords “always” or “everybody” prompt ELIZA to ask for a specific example, as was seen in the above answer to “They’re always bugging us....” (Winograd and Flores 120). In fact, Weizenbaum chose the role of a psychotherapist because of how easy it would be to simulate the responses (Sternberg 485). Seen this way, ELIZA does not really understand the patient, but rather, simply responds to expected input.

Another natural language processing program called SHRDLU, developed by Winograd himself, also falls into his own criticisms of computers and understanding. SHRDLU was developed in the 1970’s and was considered a landmark in AI (Sternberg 485). In this program, a robot responds to an operator within a three-dimensional world of blocks:

Pick up a big red block.
OK.
Grasp the pyramid.
I DON’T UNDERSTAND WHICH PYRAMID YOU MEAN.
Find a block which is taller than the one you are holding and put it into the box.
OK.
What does the box contain?
THE BLUE PYRAMID AND THE BLUE BOX (Winograd and Flores 109).

The problem with SHRDLU, Winograd and Flores argue, is that the words of a sentence can be interpreted independently of the individuals in the situation (111). While the robot seems to be deciphering the context of the situation when he asks which pyramid the person is referring to, he is actually just comparing the pyramid mentioned to a “list” of previously-defined blocks (113). It did not “understand” anything beyond that list of input that, again, the programmer had anticipated. Winograd and Flores give a wealth of these kinds of examples to support their
claims, including more recent approaches of frames and scripts, which try to incorporate a broader knowledge. They all fail, they point out, because "...the structures are written with a few specific examples in mind and work well only for those examples and minor variations on them..." (118).

Many have recognized the fundamental problems of this approach to AI, and a new view has emerged with a shift in focus from understanding individual words or sentences into a broader level of general understanding—a shift very much aligned with the metaphor theory of Lakoff and Johnson. For instance, in their comprehensive summary on the field of artificial intelligence, Stuart Russell and Peter Norvig model discourse understanding as the following equation: KB' = DISCOURSE-UNDERSTANDING(TEXT, KB). In this equation, KB is the hearer's "knowledge base" before the text, and KB' is the knowledge base after the text (715-716). The difference between KB and KB' represents an understanding of the text. The metaphor theory of Lakoff and Johnson fits into this crucial shift in natural language processing. It looks beyond the words themselves in language. For example, the sentence "I've never won an argument with him" is not understood by just looking at the meanings of the individual words. Instead, it is understood through the "argument is war" metaphor that surrounds it, taking into account how the way we act in and talk about an argument is partially structured by the concept of war. It moves a step up from understanding what individual words may mean, to one of the true goals of natural language processing: human understanding. According to Lakoff and Johnson, the understanding modeled by the difference of KB' and KB is rooted in the metaphors that we live by.

Many psychologists and philosophers give support for metaphor as a tool for this broader understanding, including Donald Schön of the Department of Urban Studies at M.I.T.¹ In his essay on metaphor, Schön describes how the metaphors we use can shape the way we view social issues and transform our social policy. For example, he considers the issue of housing, and describes two metaphors, "slum as a disease" and "slum as a natural community" (144-145). In the first metaphor, the housing problem is seen as a healthy community that has become destroyed by disease. This metaphor was evident in the literature on housing in the 1950's, in phrases such as "if the community were to be healthy" or "...as though possessed of a congenital disease" [italics added] (145). The "slum as a disease" metaphor not only affected the way housing problems were understood in the 1950's, but also the way the policies were carried out.
True to the metaphor, the actions taken included getting rid of this "disease" by rebuilding homes, schools, and streets (146). The 1960's, in contrast, used a different metaphor to describe the problem, and consequently, different policies were carried out. Using the "slum as a natural community" metaphor, the focus of social policy shifted from cleansing to rebuilding within the community. With Schöns's ideas, we see metaphors supporting a broad understanding of our world.

Other researchers have used their study of a specific metaphor to support this view of broad understanding. One of the most notable of these is the conduit metaphor of Michael Reddy, from the Department of Linguistics at Columbia University. In this metaphor, communication is structured by the following ideas: "ideas (meanings) are objects," "linguistic expressions are containers" and "communication is sending" (Lakoff and Johnson 10). In this way, our words and sentences are seen as packages, with our ideas as the objects inside, sent along a conduit to another person. All the receiver needs to do is open the package and take out the ideas, and they will have the full meaning of the sender. In this metaphor, other aspects of communication, such as the context of the situation and the fact that people interpret things differently, are ignored (10). Examples of the conduit metaphor throughout our language are extensive and Reddy estimates that 70% of the sentences we use to describe communication use the conduit metaphor:

You know very well that I gave you that idea.
That thought is in practically every phrase!
Next time you write, send better ideas.
John says he can't find your idea anywhere in the passage (189-193).

Not only do we use this language to talk about communication, Reddy argues, but it also shapes the way we think about communication. To demonstrate this, Reddy issues the challenge of a simple exercise: try to avoid expressions using the conduit metaphor and replace references to it with a neutral sentence. He described the awkwardness and difficulty of the results with himself and his students. "Did you get anything out of that article?" was replaced with, "Were you able to construct anything of interest on the basis of the assigned text?" and telling someone to "communicate more effectively" was substituted in place of the more powerful, "you've got to learn to put your thoughts into words" (177-178). While Reddy provides a wealth of detailed examples showing our dependence on this conduit metaphor in our cognitive system,² perhaps the most appropriate illustration is natural language processing itself. As described earlier by
Winograd and Flores, the problem with most natural language processing systems is their failure to take into account context and the interpretation of the individual. ELIZA, for instance, is assumed to understand because she responded appropriately and the meaning was contained within the words. Disapproval of systems such as ELIZA and SHRDLU focus on how they lie within the conduit metaphor. We have traditionally viewed communication in terms of the conduit metaphor, and this has influenced how we integrate it with natural language processing. The conduit metaphor, then, illustrates both how metaphors support a broad understanding that influences how we interact in our world, and how using this metaphor has its limitations when applied to natural language processing.

More and more researchers are recognizing the constraints of the conduit metaphor. They try to remedy its faults by working at a broader level of natural language processing that includes context, something very much a part of the theory of Lakoff and Johnson. In Russell and Norvig’s description of discourse understanding, for example, they list several types of knowledge that have a role in this understanding, including knowledge about the situation, and knowledge about the beliefs of themselves and the speaker (716). Lakoff and Johnson’s emphasis on metaphors unique to various cultures and to the individual provides a rich basis for context. For instance, in our initial glimpse into Lakoff and Johnson’s theory, we looked at the “argument is war” metaphor common to our culture. What if, Lakoff and Johnson propose, another culture did not use this metaphor, but rather used the metaphor “argument is dance”?

Imagine a culture where argument is viewed as a dance, the participants seen as performers, and the goal is to perform in a balanced and aesthetically pleasing way. In such a culture, people would view arguments differently, experience them differently, carry them out differently, and talk about them differently (5).

Their theory of metaphor accounts for context by recognizing that all cultures live by different metaphors. Moreover, Lakoff and Johnson break down this context further with their emphasis on the individual. While our culture may unify many components of our conceptual system, we each have our individual metaphors that we live by, depending on personal experiences. For instance, some people in our culture live by the metaphor, “love is madness,” where love is seen as something beyond our control. This is shown in the sentences, “I’m crazy about her” and “She drives me out of my mind” (49). Others, though, may live through an opposite metaphor, such as “love is a collaborative work of art.” In this metaphor, love is something that is in within
our control and requires constant work (141). One person’s metaphor for a concept may be quite contrary to another’s. Our own metaphors create and define a reality for us (145).

Clearly, then, the importance of the metaphor theory of Lakoff and Johnson to artificial intelligence is too great to be ignored. In two ways, the theory offers insight into the problems pointed to by Winograd and Flores and the conduit metaphor view of natural language processing. First, it overcomes the restriction of merely looking at the words of a sentence to extract meaning. Instead, we find a theory that incorporates a higher-level understanding based in our conceptual system. Second, we find the application of context to this understanding, as is shown within their cultural and individual views on metaphor. As we stop viewing communication and understanding through the conduit metaphor, Lakoff and Johnson’s ideas will become even more significant to artificial intelligence and natural language processing.

**Computational Models**

Not only does Lakoff and Johnson’s theory represent an essential shift as to how we view natural language processing, but it also has the capability of being computationally modeled, moving its influence from the theoretical into actual uses in artificial intelligence. It is not enough to say that we understand through metaphor—we need concrete ways of modeling it. Lakoff and Johnson’s ideas on metaphor can be modeled in part because of three elements of the theory. First, there are mathematical mappings in structural metaphors between target and source domains, meaning the mapping from one concept to another to form a metaphor is not arbitrary. Second, our metaphors form a hierarchy in which certain metaphors are subsets of others, allowing us to abstract a small set of core metaphors for the model. Finally, diverse metaphors for the same concept are not contradictory, but structure the concept in a coherent and logical manner. All of these things provide a rich framework in which to develop computational models of metaphor that can be used for natural language processing, as is already being done in many cases.

First, Lakoff and Johnson’s theory defines a mathematical mapping between target and source domains in structural metaphors. The target domain is what is being understood (e.g. argument) in terms of the source domain (e.g. war). The mapping from a source to a target is highly structured, depending on both the characteristics of the target and the source domains (Lakoff 209). This means that metaphor mappings, and therefore metaphors themselves, do not occur by chance. There is a systematic structure to them. For example, the following is a
diagram of one of the three “modes” of metaphor of Bipin Indurkhya, a professor in the Department of Computer Science and Engineering at the State University of New York:

In this illustration, consider the target environment to be the way we use argument in our everyday life. The target environment is transformed into the target model, which is how we view argument conceptually. The source model, which could be war, is mapped onto the target environment through the target model, and is represented by the equation “C_{ST} \circ C_{T}.” In this way, we can see how the source domain is mapped to the target domain to form a metaphorical relationship, depending on the structure of each. One thing to notice in particular is how the metaphorical relation highlights some aspects of the target domain while downplaying others. For instance, we know that there are some things about argument that are not like war; otherwise argument would actually be war. We see how the metaphorical mapping highlights the way in which they are alike, which is shown on the target environment with darker lines, while the other lines on the target environment that do not necessarily fit the metaphor are not emphasized (Indurkhya “Modes”12). This, in essence, is one of the main functions of metaphor, according to Lakoff and Johnson. One concept partially, but never completely, structures another (13).

This structural mapping forms the basis of many metaphor models in existence today, including the dominant SME (Structure Mapping Engine) model that was created by Dedre Gentner, professor of psychology at Northwestern University (Veale 4.47). While different
models may not agree how a source domain is mapped onto a target domain (which will be examined later), the real importance here is that whatever the method, the mapping is not arbitrary. This generates the possibility for a computational model to represent metaphor, because the definition of “computation” is that it follows specific rules and restrictions. For metaphor, these rules are found within the mapping between target and source domains.

Another characteristic of Lakoff and Johnson’s theory is that the metaphors we live by form a hierarchy structure. This means that we have a small, superordinate set of core metaphors that we use, and the seemingly endless metaphors in existence are simply specializations of one of these core metaphors. For instance, take the metaphor “love is a journey.” There are specializations of this metaphor, such as “love is a car” (“long bumpy road,” “spinning our wheels”), “love is a train” (“off the tracks”) or “love is a boat” (“on the rocks,” “floundering”) (Lakoff 214). While they all are distinct, it is unlikely that one finds one of these specializations without some of the others in conjunction with it. This supports the idea that all of these different types of specializations should be grouped under the broader “love is a journey” metaphor (214). Certainly, it is much more fitting for a computational model to depict the metaphor “our relationship is on the rocks” as part of the “love is a journey” metaphor than to depict it as a “love is a boat” metaphor. The latter method would be far too complex to model all metaphors, making it difficult to capture metaphors in a single computational model. Instead, Lakoff and Johnson’s theory allows us to extract the core, general metaphors as the basis for metaphor interpretation.

Several models of metaphor, including those of Jamie Carbonell and James Martin, use this distinction between core and specialized metaphors. Jamie Carbonell, professor of computer science at Carnegie Mellon University, uses this very element as the entire reason for her model’s beginnings and development:

The computational significance of the existence of a small set of general metaphors underlies the reasons for my current investigation: The problem of understanding a large class of metaphors may be reduced from a reconstruction to a recognition task. That is, the identification of a metaphorical passage as an instance of one of the general metaphorical mappings is a much more tractable process than reconstructing the conceptual framework from the bottom up each time a new instance of a metaphor is encountered (416).

The model of James Martin called MIDAS (Metaphor Interpretation, Denotation, and Acquisition System) continues the use of core and specialized metaphors in a hierarchy (Veale
4.12. “Thelma gave her flu to Louis,” for example, would be part of the more basic, core metaphor “Infection as Possession” (4.12). Martin uses large amounts of hierarchically organized metaphorical knowledge as the beginning base of natural language processing (Martin “Computer Understanding” 268). Clearly, it becomes much easier to create a computational model when information can be organized in such a hierarchical fashion. Otherwise, there is simply too much information to try to encapsulate within the model.

Finally, the last characteristic that solidifies Lakoff and Johnson’s connection to computational models and artificial intelligence is the coherence found across varied metaphors that structure the same concept. Much like the hierarchy of metaphors, this coherence also allows an organization to be placed upon the different metaphors, rather than treating every metaphor as a specific unrelated case. Lakoff and Johnson argue that all of the diverse metaphors for a particular concept or a set of concepts are very much structured into an organized whole. As stated earlier, the purpose of a single metaphor is to highlight some parts of a concept and hide others. Different metaphors for the same concept may highlight different things, but that does not mean they are contradictory (45). For example, the concept of love has a number of metaphors. “Love is a journey,” “love is magic,” “love is madness,” and “love is war” are some of them. No one metaphor could explain love, but all of the different ones we use highlight different aspects of love, and come together in a system to structure how we think about and act in love (49).

Lakoff himself argues that this coherence can be modeled in a connectionist network, a very familiar concept to artificial intelligence. He maintains that the various metaphors for one concept involve simultaneous multiple mappings that can be modeled as simultaneous connections between “nodes” (205). These nodes are simply graphical symbols to represent different parts of concepts, and all of the different connections form what is called a network. In this way, the different metaphors for a particular concept would not contradict one another, but rather highlight, or activate, different aspects (nodes) of the concept. Some have followed this example in their models of metaphor. One such model is ACME (Analog Constraint Mapping Engine) developed by Keith Holyoak, professor of psychology at UCLA, and Paul Thagard of the Department of Philosophy at the University of Waterloo. In this model, assorted activation levels between nodes are used to guide the system to make the most coherent mappings of metaphor (Wagman 108, Veale 4.48). Those who have followed suit, combining metaphor and
connectionism, include Garrison Cottrell, professor of computer science at the University of California San Diego, and Jaime Carbonell, who uses core and specialized metaphors in a recognition network (Cottrell 70; Carbonell 417). This last element of coherence of metaphor once again provides the means to computationally model metaphor.

The above three characteristics illustrate how Lakoff and Johnson’s ideas can be applied to artificial intelligence. The mathematical mappings between source and target domains, the hierarchy of metaphor, and the coherence of diverse metaphors for the same concept all provide a structure to their theory that take it beyond the theoretical realm and into practical application for natural language processing. Indeed, through various examples we have seen many different ways their ideas are being computationally modeled, perhaps paving the way for more concrete applications in AI and natural language systems.

Through the discussion of *Computers and Understanding* and *Computational Models*, we have seen that the importance of the theory of Lakoff and Johnson is not one that can easily be brushed aside. Their theory provides a comprehensive design of understanding that counteracts many of the criticisms of natural language processing, and it provides the means to implement this design into computational models. Lakoff and Johnson argue that metaphors influence and shape our conceptual system. Clearly, a more in-depth study of these metaphors could influence and shape the development of artificial intelligence as well.

**LIMITATIONS OF LAKOFF AND JOHNSON AND ARTIFICIAL INTELLIGENCE**

Since the time of Lakoff and Johnson’s book on metaphor, and some would argue because of it, there has been a dramatic increase in research involving metaphors and understanding (Martin “Computational Approaches” 91). The field is constantly growing, launching new discussions on metaphor at a rapid pace. In fact, metaphor has become so important that some of the prominent researchers in the metaphor field established a journal in 1986 dedicated solely to its debate (Ortony, xiii). Despite this enormous potential, the variety and growth in the metaphor field has its disadvantages. A number of key, unsettled issues have left the metaphor field very fragmented, weakening its utilization in natural language processing.
The issues that produce this fragmentation center upon three areas: interpreting Lakoff and Johnson’s theory in terms of computational models, fitting metaphors and their models into larger language processing as shown by the literal versus metaphorical debate, and overcoming the limitations of computers and understanding beyond the theoretical. In order to truly see the impact of Lakoff and Johnson’s theory of metaphor on artificial intelligence, some of these issues will need to be resolved.

**Interpreting Lakoff and Johnson in Computational Models**

When looking at the computational models described earlier, the application of Lakoff and Johnson’s theory seems promising for natural language processing. While not necessarily disputing this claim, a closer look at these models being developed reveals some essential, often contradictory differences among them. Two of these differences are how core metaphors and metaphor mappings are implemented. First, although many models use a small set of core metaphors as their basis, there is no consensus as to how small “small” is or how detailed the metaphors within the set should be. Similarly, almost all models follow a structure for mapping a source to a target domain, but the details of this structure vary greatly from model to model. We see how the same two components of the theory that provided the means for computational models also foster a number of problems as we compare the models to each other. Indurkhya captures one of the weaknesses of Lakoff and Johnson’s theory when he says, “Though several theories have been proposed in the past few years—mostly by philosophers and linguists—many of them are not spelled out with sufficient detail and rigor to be able to be used as a basis for designing computer models” (“Semantic Transference” 449). As a result, there is a general lack of consensus on how to apply Lakoff and Johnson’s theory to computational models.

First, there is disagreement on the core metaphors used as the basis for computational models. As previously mentioned, Lakoff and Johnson claim that a small set of generalized metaphors structured in a hierarchy provides the framework for metaphor interpretation and creation. This is one of the aspects of their theory that fosters computational models, as it allows us to reasonably extract a modest set of metaphors as a manageable foundation. A number of models use this trait of the theory, but they disagree as to how small this core set of metaphors should be and how detailed. Some core sets of metaphors are extremely small, while others
believe a larger, more detailed set is needed to accurately account for metaphors in our conceptual system.

On the side in favor of a miniature, very undetailed set of core metaphors are Tony Veale and Mark Keane, both from the School of Computer Applications at Trinity College Dublin in Ireland. Veale and Keane’s model begins with only a very small set of abstract, unstructured metaphors (Martin “Computational Approaches” 92). Their model is based upon a term they coined, called “conceptual scaffolding,” where first the broader meaning of the metaphor is found, with the specific meaning elaborated over time (The Metaphor Home Page). They divide metaphor interpretation into two stages. In the first stage, the initial scaffolding is constructed to give the “broad semantic themes” of the metaphor. In the second, the scaffolding is elaborated by various factors, such as world knowledge, domain knowledge and context (Veale and Keane 495). With this approach, only a very small set of core metaphors is needed to begin metaphor interpretation. Carbonell’s model, too, agrees with this philosophy. Similar to Veale and Keane, we have already seen that the key to Carbonell’s model is a very small set of general metaphors.

Others focus on a larger, more specific set to form the base of the metaphor hierarchy. Martin’s MIDAS, for example, begins with a large number of highly structured metaphors. From empirical research based on analyses of large bodies of text, Martin asserts that an extremely small set of core metaphors, such as those of Veale and Keane, simply does not account for the wide range of metaphors encountered in practice (“Computational Approaches” 93). Therefore, Martin claims that a more extensive, highly structured set of core metaphors is necessary to represent the endless metaphors found in our language. We see, then, how this differs fundamentally from the ideas of Veale, Keane, and Carbonell.

There are also distinct divisions on another fundamental issue within computational models: the mappings between target and source domains. One of the advantages of Lakoff and Johnson’s theory, as stated previously, is that there is a definite mathematical structure as to how a source is mapped onto a target domain in a metaphor. However, there are a wide variety of ways this metaphorical mapping has been implemented within different models. Some focus on exact structure mappings between the target and the source domains, some downplay either the target or source domain’s influence within the mapping, while still others ignore both of these and focus on the content of metaphor. Looking at these three cases in turn, we see that one important question has not yet been answered when attempting to apply Lakoff and Johnson’s
theory to computational models: how are the mappings between source and target domains determined?

First, there is one extreme with the school of Dedre Genter and Arthur Markman, proponents of the Structure Mapping Engine. Originally developed by Genter and later joined by Markman of the Department of Psychology at Columbia University, the SME remains one of the dominant computational models for metaphor today (The Metaphor Homepage). This model is based upon the “systematicity principle,” which states that a metaphor is created if and only if there is a one-to-one mapping between target and source domains (Veale 4.47). A high-level view on how this works is as follows: stage one tries to find local mappings in structure between the target and the source domains, stage two coalesces the consistent mappings that were found, and stage three merges the consistent mappings into one consistent image (Genter and Markman 49). Genter’s theory concentrates on using existing structural mappings to create similarities between a target and a source domain. Therefore, the structure of both the target and source domain must match perfectly in order for a metaphor mapping to be created.

Some move away from this strict structural approach to metaphor. Indurkhya’s three modes of metaphor, for instance, include metaphor models that create similarities between target and source domains where structural similarities did not exist before (“Modes” 4). This model concentrates on finding new metaphors to completely change our understanding and view of the world. Indurkhya says the key to Lakoff and Johnson’s theory is that a target domain can be restructured in a different way through the use of metaphor (4). We saw this especially through Schön’s arguments, where our metaphors of housing changed over time, modifying the way we dealt with slums and other social problems. Indurkhya seeks to model this phenomenon through his projective metaphor. This model works by “completely disregarding the structuring of the target environment under the target model, and projecting the source model on it anew, as if the target environment were being encountered for the first time” (20). This, Indurkhya states, accounts for the generative metaphors supported by Schön, explaining how we can restructure a concept by using a different metaphor (21). Genter would argue that these shifts in metaphor could be found by using different one-to-one structural mappings, while Indurkhya maintains that this can be accomplished only by relaxing such a structural restriction (Genter and Markman 49).
Genter and Indurkhya hold opposing viewpoints as to how much the actual structure of the target and source domains should be used in metaphorical mappings. Still others, such as Carbonell, disregard structure completely. Carbonell, as we have seen, fully develops Lakoff and Johnson’s hierarchy of metaphor. However, Carbonell’s mappings in metaphor completely ignore the structure of the target and source domains and focus on content. Carbonell suggests a prioritized “list” should be used for comparison when trying to recognize a metaphor (Martin “Computational Approaches” 95). Some have criticized this approach for focusing only on how to understand metaphors semantically, with no indication on how to create new metaphors (Indurkhya “Semantic Transference” 448). The model, they say, contradicts Lakoff and Johnson because it ignores the emphasis on structure as well as one of the main ideas of their theory, which is that metaphors are constantly shifting and being created to define our reality.

Through the examples of core metaphors and mathematical mappings, we see different models highlighting different aspects of the theory in numerous ways. There is a lack of consensus as to how Lakoff and Johnson’s theory should be applied, leaving the field too fragmented to build a unifying model of metaphor. It is difficult to compare computational models that are based upon fundamentally different ideas, and therefore the use of Lakoff and Johnson’s theory of metaphor in artificial intelligence suffers.

**Metaphor and Larger Natural Language Processing**

Not only is there disagreement on how to apply Lakoff and Johnson’s theory to computational models, but there is another important dispute found within the broader level of metaphor study. It deals with the question of how metaphor fits into larger natural language processing systems. More specifically, should a line be drawn between metaphorical and literal language, and if so, where? Take, for instance, the ELIZA program described earlier. In what terms should this “conversation” be understood? Are even these simple statements rooted in metaphor, or is there some division between what is interpreted literally and what is interpreted metaphorically? The proposed answers to these questions remain widespread. Some claim that natural language systems based upon Lakoff and Johnson’s ideas entail, by definition, no distinction between metaphorical and literal language. Several others, however, either ignore this idea or reject its premise. Certainly, we need to come to some agreement on such an important question in order to determine how Lakoff and Johnson’s theory of metaphor fits into natural language processing.
From one perspective, Andrew Ortony argues that an acceptance of Lakoff and Johnson’s ideas would entail a rejection of literal language. Ortony is the editor of *Metaphor and Thought*, a book containing essays from the prominent figures in metaphor research as of the early 1990’s. In the introduction to the collection of essays, Ortony writes the following of Lakoff and Johnson and literal language:

Such [literal] approaches seem ill equipped to deal with what many view as a central aspect of human thought and cognition, namely the ability to understand something from a different point of view.... Therefore, if metaphors are important because of their ability to provide alternative or new ways of viewing the world, then so-called literal language may be too restrictive because of its inability to provide those perspectives (13).

Ortony argues that if the central thesis of Lakoff and Johnson, that metaphors structure the way we understand and view our world, is correct, then we only need to concentrate on this metaphorical language. This is because understanding things from a new perspective is seen as the main component of human cognition. Ortony casts somewhat of a division between metaphorical and literal language, but argues that metaphorical language is the more important part of language to consider. Literal, non-metaphorical language exists, but it is of little use in understanding our world through various points of view (13).

Others take Ortony’s basis and expand it even further, claiming that *no* distinction between literal and metaphorical language can be made. Max Black, one of the pioneers of the shift in metaphor theory, claims that any attempted division between literal and metaphorical language will fail. He returns to the problems with traditional views of metaphor to illustrate his point. Recall that, in the past, metaphors were viewed solely as a literary device, another way to say “what we really mean.” However, the whole basic concept of Lakoff and Johnson is that metaphors are not simply literary devices. Rather, metaphor structures the way we understand things, and as we understand through metaphor and use it in our everyday language, we are saying “what we really mean” (Black 22). Whenever we refer to literal language at all, Black states, we are referring to the concept of what the truth of our statements are, and Lakoff and Johnson’s theory allows us to express this truth through metaphor (22). Cathy Wheeler supports this view in an essay on metaphor as she proposes that there is not one absolute “literal” truth, but multiple truths and realities defined through metaphor (236). Whenever we try to make any kind of division between metaphorical and literal language, we accept the “misleading view of
metaphor as some kind of deviation or aberration from proper usage” (Black 22). Therefore, Black argues that any distinction between metaphorical and literal language fundamentally rejects the entire basis of Lakoff and Johnson’s theory.

Ortony, Black and Wheeler’s arguments are based deeply within the theoretical views of metaphor. In practice, however, within the world of computational models, much of the research differs from these views. This is demonstrated by metaphor comprehension studies, which compare the interpretation of literal and non-literal statements to determine if the same or different processes are used to comprehend each type. The entire basis of this research assumes that there is a distinct division between metaphorical and non-metaphorical language. Richard Gerrig of Yale University, for instance, argues that computational models of metaphor need to take into account a total time constraint. This constraint asserts that metaphorical comprehension does not take any longer than when we are understanding something literally (236-237). The support for this argument is found within reaction-time studies. In these studies, participants are presented with some stimulus, such as the reading of a sentence. Then they are required to produce some sort of response, including pressing a button or giving a verbal response. The reaction time (RT), or how long it takes the participant to respond, should give an indication as to the depth of the mental processes involved in understanding the sentence (Hoffman and Kemper 152-153). The problem with most computational models of metaphor, Gerrig argues, is that they first assume a literal interpretation. Only when this interpretation fails do they shift into a metaphorical interpretation, making metaphorical understanding much slower than literal. Instead, both types of interpretations should take equivalent amounts of time (236). Here, we find a new paradigm within the debate on metaphorical and literal language. According to Gerrig, it is not the distinction between literal and metaphorical language that is the problem; rather, it is when one assumes that the processes for each are different.

On another extreme of the literal and metaphorical debate, Bipin Indurkhya claims not only that all language is not metaphorical, but that Lakoff and Johnson themselves support this idea. First of all, Indurkhya states that the thesis that all language and thought is metaphorical is based upon the assumption that cognition works by “creating conceptual structures in the world and not by adapting to pre-existing ones” (“Knowledge” 66). This would mean that our entire conceptual system is created by metaphor, with no beginning, non-metaphorical concepts. However, Indurkhya points out that Lakoff and Johnson describe some non-metaphorical
concepts within their theory of metaphor. A prime example of this is their explanation of orientational metaphors (67). A key concept of their theory is that orientational metaphors are grounded in our own physical experiences. For example, the metaphors “happy is up” and “sad is down” originate within the physical experience that when we are sad our body physically sags, while it is upright when we are happy (Lakoff and Johnson 15). Ontological metaphors, similarly, arise from our physical experience (29). According to Indurkhya, this qualifies as a pre-existing, non-metaphorical conceptual structure. In addition, Indurkhya also looks at Lakoff and Johnson’s mathematical mappings between target and source domains. We already know that Lakoff and Johnson argue that a mapping between such domains depends upon the structure of both the target and the source. Indurkhya maintains this would mean that the target and the source both have a pre-metaphorical structure before the mapping takes place. This disputes the claim that all understanding is only through metaphor (“Knowledge” 70). Differing very much from the views of Black and Ortony, who refuse or downplay literal language, Indurkhya claims the theory of Lakoff and Johnson favors the existence of literal language and understanding.

Finally, we have the view of Eileen Cornell Way, from the departments of Philosophy and Computer and Systems Sciences at the State University of New York. Way assumes that Lakoff and Johnson’s theory does support the idea that all language is metaphorical, but disagrees with such an assertion. She points out the way we use the word “literal” in our everyday language: “We often qualify what we say with ‘I literally meant that...’ in contrast, say, to exaggerating, being ironic, being facetious...”(18). The problem, she states, is that the view of literal has been reduced to an “absolute truth,” as in the case of Wheeler and Black. Literal, she points out, is used to make a linguistic distinction as to your personal intention of a sentence (18). She gives the example of the following sentence: “My common sense literally flew out the window.” Here, she states, the “literal” interpretation does not assume that the sentence should be interpreted within the exact meanings of the words. Rather, she remarks, a literal interpretation of this metaphor is used to show that the person really did lose all common sense; they were not just being ironic or glib (19). Way credits literal language as a way to clarify and qualify our intended meanings in language, putting her in direct opposition to Ortony, Black and others on the issue of metaphorical and literal language.

From the varying viewpoints on literal and metaphorical language, we can see that the issue remains undecided and very complex. The viewpoints range from no distinction between
metaphorical and literal language, to very clear cutoffs between the two. Moreover, Lakoff and Johnson’s theory has been used as an argument in diametrically opposing sides. Therefore, Lakoff and Johnson’s fit within natural language processing is uncertain. Can metaphorical models be incorporated into existing natural language models or must they replace them? Clearly, only when researchers reach some common conclusions about literal and metaphorical language will we know the role metaphors and understanding will play within natural language processing. With such a wide array of contradictory views, Lakoff and Johnson’s current applicability to artificial intelligence remains somewhat limited.

Computers and Understanding Revisited

The above discussions focus on some of the unresolved issues limiting Lakoff and Johnson’s applicability to artificial intelligence. Along these same lines, we also need to revisit the issues of natural language processing established by Winograd and Flores. By briefly reexamining the ideas of computers and understanding, we find that although Lakoff and Johnson provide a theoretical alternative to its faults, it is not so clear-cut in practice. Instead, through a repeat of small domains and a lack of discussion on what it means for a computer to understand through metaphor, some problems established by Winograd and Flores remain unsettled.

First of all, one of the major criticisms Winograd and Flores have of natural language processing is its emphasis on small domains. Programs such as ELIZA, they argue, did not truly understand, because it simply took expected input and produced the appropriate output based upon these expectations. In theory, Lakoff and Johnson provide an alternative to this because their theory of metaphor is one of a broad human understanding and context. However, in practice, this aspect has not been fully utilized, partly because of the irresolution of some of the issues mentioned earlier. Most models of metaphor, rather than representing a broad, general theory, focus upon a specific domain. Indurkhya writes the following in description of this:

...their [computational model’s] main concern is to produce an acceptable output within a reasonable time in the domain that they are designed for....for all these reasons, one cannot always abstract a general theory...embodied in the model ("Semantic Transference" 448).

This, in effect, echoes the condemnation of Winograd and Flores, showing how Lakoff and Johnson’s theory can come full circle as it is applied to natural language processing. While
providing the basis for answering Winograd and Flores’ criticisms, it also, depending on how it is applied, has the potential to fall victim to them.

In addition, we see how there is a general lack of discussion on what it means for a computer to understand through metaphor, again allowing the faults found by Winograd and Flores to remain. For instance, the various metaphor models discuss metaphor interpretation and creation for human understanding. How would this, however, transfer to computer understanding? Would it again be a computer understanding by responding to expected input, only this time with responses using the correct metaphors within the language? Or does the criteria for judging computers and understanding need to change? In either case, would such a model truly answer the criticisms of Winograd and Flores? Clearly, it does no good to offer a broader theory of understanding if we apply it within the traditional narrow definitions. Instead of a discussion on these issues of computers and understanding, we find a host of unanswered questions. Therefore, we observe that the applicability of metaphor theory to artificial intelligence depends not only on resolving issues within the metaphor field itself, but resolving on a broader level what it means for a computer to understand.

Through small domains and the absence of discussion on what it means for a computer to understand through metaphor, we see how the application of Lakoff and Johnson to artificial intelligence resurfaces the problems pointed to by Winograd and Flores. In order to see any real advancement in Lakoff and Johnson’s application, we need to dig deeper into natural language processing to answer these fundamental questions, and to apply the advantages of Lakoff and Johnson’s metaphors not only in theory, but in practice as well.

Through the discussions of Interpreting Lakoff and Johnson in Computational Models, Metaphor and Larger Natural Language Processing, and Computers and Understanding Revisited, we see some of the problems associated with applying Lakoff and Johnson’s theory to artificial intelligence and natural language processing. There are fundamentally distinct ways that researchers are applying the theory in computational models, major disagreements on the relationship between metaphorical and literal language, and a repeat of the problems found earlier with computers and understanding. In order to bring the broad theory of metaphor, so
critical to answering the criticisms of Winograd and Flores, into artificial intelligence, some of these issues need to be resolved.

CONCLUSION

It is only fitting that a paper on a theory of metaphor should end with its own metaphorical illustration. In their book, Lakoff and Johnson offer a metaphor to structure the way we view our problems. This metaphor, called the “problems are chemicals” metaphor, structures solutions to our problems in the form of a “...large volume of liquid, bubbling and smoking, containing all of [our] problems, dissolved or in the form of precipitates, with catalysts constantly dissolving some problems (for the time being) and precipitating out others” (143). In this metaphor, there are no definite answers to a problem that solve it “once and for all.” We find only partial solutions, and the problem never entirely disappears. Furthermore, the dissolving of one segment of the problem may cause others to precipitate out (143-144). Lakoff and Johnson argue that the chemical metaphor is so much more enlightening than traditional metaphors used in viewing problems, as it causes us to focus less on finding the “right” answer, and more on finding “what catalysts will dissolve your most pressing problems for the longest time, without precipitating out worse ones” (144).

The metaphor theory of Lakoff and Johnson as it applies to the problems in natural language processing can be seen most insightfully within the context of this chemical metaphor. It has been demonstrated that Lakoff and Johnson’s theory of metaphor “dissolves” some of the criticisms of Winograd and Flores by its theory of broad human understanding and account for context. Furthermore, other components of the theory contribute partial solutions within the bubbling chemical, such as the aspects that promote the construction of computational models. However, the theory of Lakoff and Johnson does not provide the ultimate solution to the faults of natural language processing. We noted that the plethora of computational models have precipitated out new problems involving how to interpret Lakoff and Johnson’s theory, and how the growth of metaphor research has spurred a debate on literal and metaphorical language. In addition, without a careful analysis of some important issues such as computers and understanding, some problems can reappear. Therefore, in order to advance Lakoff and Johnson’s application to AI, we need to concentrate on the aspects of the theory that dissolve
some of the criticisms of natural language processing, while also resolving some of the key problems that have precipitated out or returned because of it. Seen through the chemical metaphor, we are not disappointed by the lack of a “perfect” solution within the theory of Lakoff and Johnson. Instead, we see great promise and potential in artificial intelligence for the partnership of Lakoff and Johnson’s theory of metaphor and natural language processing.
ENDNOTES

1. Another work that supports the use of metaphor in understanding is Hugh G. Petrie and Rebecca S. Oshlang’s essay, “Metaphor and Learning.” In Metaphor and Thought, Andrew Ortony, Ed. (New York: Cambridge University, 1979), 580-609.


4. The diagram of Indurkhya’s shown earlier in the section on Computational Models was his syntactic model of metaphor, having a similar structure to Genter’s SME.

5. For additional discussions on this issue see the following: Jerrold Saddock, “Figurative speech and linguistics.” In Metaphor and Thought, Andrew Ortony, Ed. 2nd Ed. (New York: Cambridge University, 1993), 42-57; and David E. Rumelhart, “Some problems with the notion of literal meanings.” In Metaphor and Thought, Andrew Ortony, Ed. 2nd Ed. (New York: Cambridge University, 1993), 71-82.
BIBLIOGRAPHY


