Mental Models and Metaphor

Edwin Plantinga
Dept. of Computer Science
University of Toronto
Toronto, Ontario CANADA M5S 1A4
and
Redeemer College
Ancaster, Ontario CANADA L9G 3N6

1. Introduction

This paper investigates the significance of the mental models (MM) hypothesis for computational linguistics in general and for metaphor comprehension in particular. The MM hypothesis is the claim "that people understand the world by forming mental models."[1] The general form of this hypothesis is not new: Immanuel Kant and neo-Kantians such as Hans Vaihinger and Ernst Cassirer have argued that there is no direct access to the *things-in-themselves*. Concepts and conceptualizations mediate between the person and the world.

Although the general contours of the MM hypothesis have been around for some time, the emphasis on models and domains which one finds in the literature is a more recent phenomenon. Let us consider a definition of an MM:

> A mental model is a cognitive construct that describes a person's understanding of a particular content domain in the world. This contrasts sharply with much other work in cognitive psychology, which attempts to be domain-independent.[2]

Donald Norman, for example, investigated calculator usage and found that the models constructed by individuals varied considerably from user to user.[3] If we take the time to find out, we see that individuals do differ in the conceptualizations which they form.

---

[1] John Sowa, *Conceptual Structures: Information Processing in Mind and Machine*, (Reading, Ma.: Addison-Wesley, 1984), p. 4.

[2] John M. Carroll, Review of *Mental Models*, Dedre Gentner and Albert Stevens, eds., *Contemporary Psychology*, 30(9), September 1985, p. 694.

[3] Donald Norman, "Some Observations on Mental Models", in Dedre Gentner and Albert L. Stevens, eds., *Mental Models*, (Hillsdale, N.J.: Lawrence Erlbaum Associates, 1983), pp. 7-14.

2. In Search of *Homo Loquens*

An emphasis on individual differences does not mesh very well with the current linguistic paradigm. The individual has been banished from contemporary linguistics. Linguistics studies language but not *homo loquens*. There are a number of reasons for this.

First, linguistics wants the prestige and status that we bestow on disciplines which are sciences. To achieve this, linguists tend to the abstract and to the universal while ignoring much of the idiosyncratic nature of language use.

Second, the Saussurean distinction between *langue* and *parole* became a cornerstone of Chomskyian linguistics. Competence, the abstract linguistic system, became the major interest of linguists; performance, the actual output of language users, was only of passing interest.

Third, much of our thinking about language is shaped by a very powerful metaphor which Michael Reddy has named the 'conduit metaphor.'[4] According to Reddy, our model of human communication is based on the following:

1. Ideas (or meanings) are objects.

2. Linguistic expressions are containers.

3. Communication is sending.

A speaker puts ideas (objects) into words (containers) and then sends them (along a conduit) to a hearer who takes the ideas/objects out of the word/containers. What an expression means depends on what meaning the speaker inserted into the container. Since the meaning is *in* the expression, the recipient need only retrieve the meaning. In this model, the individual hearer contributes nothing — he merely receives.

But the hearer does not receive meanings — he receives words. To the hearer falls the task of generating meaning in response to these words. In short, meaning is response.[5] What is

---

[4] Michael Reddy, "The Conduit Metaphor — A Case of Frame Conflict in our Language About Language" in Andrew Ortony (ed.), *Metaphor and Thought,* (Cambridge: Cambridge University Press, 1979).

[5] I have argued this position in greater length in "Who Decides What Metaphors Mean?", *Proceedings of the Conference on Computing and the Humanities — Today's Research, Tomorrow's Teaching,* Toronto, April 1986, pp. 194-204.

manufactured depends on the architecture of the meaning generator. Abandoning the *conduit metaphor* forces us to bring the individual into linguistics so that the discipline focuses on both language and the individual language processor. Mental models give us a way of bringing the architecture of the individual language processor into linguistics.

3. Modeling Mental Models

A common strategy for software development is to precede the implementation phase with a problem definition phase. Normally, the implementation does not commence until the problem definition is complete. But this strategy will not work in constructing models of MMs. Philip Johnson-Laird argues that mental models cannot be defined currently:

> At present, no complete account can be given — one may as well ask for an inventory of the entire products of the human imagination — and indeed such an account would be premature, since mental models are supposed to be in people's heads, and their exact constitution is an empirical question.[6]

An alternative strategy is to use an iterative software development methodology. We learn by building so that the problem definition is refined during the development process. The computer-based modeling of mental models should shed light on their nature.

Assume the existence of some domain d.[7] An agent, agent-1, constructs a MM of that domain which we call $MM_{agent-1}(d)$. It is tempting to claim that another agent, agent-2, forms a second MM of 'that same domain.' But that assumes that agent-1 and agent-2 participated in 'exactly the same discourse.' The domain of agent-1 may be similar to the domain of agent-2, but they are not the same.

MMs are not restricted to 'domains in the world.' First, an agent can construct a MM of some imaginary domain. Second, an agent can construct a MM of some other agent's MM. Let $MM_i(MM_j(d))$ represent agent$_i$'s MM of agent$_j$'s MM of some domain.

---

[6] Philip Johnson-Laird, *Mental Models* (Cambridge, Ma.: Harvard University Press, 1983), p. 398.

[7] As Stephen Regoczei and I have argued, the domain of discourse is created by the discourse. This idea is consistent with the Whorfian hypothesis and much of post-structuralist thinking. See Stephen Regoczei and Edwin Plantinga, "Ontology and Inventory: A Foundation for a Knowledge Acquisition Methodology", *Proceedings of the Workshop on Knowledge Acquisition*, Banff, Alberta, November 1986, to appear.

In order to model a MM on a computer, we must select some individual, perform knowledge acquisition operations with the individual, and then build a model of the informant's MM. But what we are constructing is not a model of the informant's MM (i.e., $MM_{informant}(d)$) but a model of the analyst's MM of the informant's MM (i.e., $MM_{analyst}(MM_{informant}(d))$). If the development involves a number of individuals, then the model constructed will not correspond to any particular agent's model.

John Sowa has defined a notation called conceptual graphs (CGs), which is ideal for modeling MMs. CGs are suitable for both knowledge representation and also for the knowledge acquisition phase which must precede the representation phase.[8]

> Sowa suggests that concepts are the atomic components of mental models:

> Concepts are inventions of the human mind used to construct a model of the world. They package reality into discrete units for further processing, they support powerful mechanisms for doing logic, and they are indispensable for precise, extended chains of reasoning.[9]

MMs have a structure which can be modeled using CGs. Each conceptual graph consists of nodes which either represent concepts or conceptual relations. In their linear notation, conceptual graphs are directly machine representable. Operations on MMs can be modeled by operations on conceptual graphs. Since Sowa has defined the algorithms necessary to implement a conceptual processor,[10] CGs form a basis for modeling both MMs and operations on MMs.

4. Natural Language Processing and Mental Models

Although our vocabularies overlap considerably, the concepts which each of us hold have our own personal stamp upon them. George Steiner has stated this most elegantly:

---

[8] The merits of Sowa's approach are outlined in more detail in Regoczei and Plantinga *op cit.*

[9] John Sowa, *Conceptual Structures - Information Processing in Mind and Machine,* (Reading, Ma.: Addison-Wesley, 1984), p. 344.

[10] At least one conceptual processor has been implemented. See Jean Fargues, Marie-Claude Landau, Anne Dugourd, Laurent Catach "Conceptual Graphs for Semantics and Knowledge Processing", *IBM Journal of Research and Development,* 30(1), January 1986, pp. 70-79.

Each living person draws, deliberately or in immediate habit, on two sources of linguistic supply: the current vulgate corresponding to his level of literacy, and a private thesaurus. The latter is inextricably a part of his subconscious, of his memories so far as they may be verbalized, and of his singular, irreducibly specific ensemble of his somatic and psychological identity. Part of the answer to the notorious logical conundrum as to whether or not there can be a private language is that aspects of every language-act are unique and individual. They form what linguists call an idiolect. Each communicatory gesture has a private residue. The 'personal lexicon' in every one of us inevitably qualifies the definitions, connotations, semantic moves current in public discourse.[11]

Is this 'personal lexicon' a blessing or a curse? It is this 'personal lexicon' which makes language understanding idiosyncratic. While there is some overlap in the concepts each of us possess, there is also considerable non-overlap; while there is room for understanding, there is also considerable room for non-understanding or misunderstanding.

If this 'personal lexicon' is a deficiency, why should we build this into computers? Why should computers misunderstand? So far, attempts have concentrated on making computers understand. Understanding in this case means translating linguistic input into the meaning representation. For example, if the representational system is CGs, then the translation maps words into concepts. But which concepts should the machine have?

The temptation is to say, "Only those which are true." But this poses two problems. First, as Lakoff and Johnson have pointed out, our conceptual systems are metaphorical. To lock the door on concepts which do not 'correspond to reality' will exclude machines from modelling a large part of our mental life. Second, who decides what is true? This is a pragmatic issue which must be faced in the knowledge acquisition phase. Should the analyst argue with the informant? Should the analyst claim that the informant's concepts are wrong?

During the knowledge acquisition phase which precedes construction of a natural language processing (NLP) system, the analyst should attempt to acquire the concepts of the informant without judging the concepts to be acceptable or unacceptable. In practice, this is difficult to achieve. Once acquired and represented in a machine usable form, the words which act as input to the system are mapped to concepts.

---

[11] George Steiner, *After Babel: Aspects of Translation,* (New York: Oxford University Press, 1975), p. 46.

Sowa has suggested a mechanism for connecting words and concepts: a lexicon which lists the concepts into which a word can be mapped. If a word has multiple senses, multiple concepts are stored in the lexicon. In Sowa's lexicon, for example, the word 'occupy' is associated with three different concepts: [OCCUPY-ACT], [OCCUPY-STATE], and [OCCUPY-ATTENTION]. The following sentences illustrate the three concepts:

The enemy occupied the island with marines.

Debbie occupied the office for the afternoon.

Baird occupied the baby with computer games.

Using this-word-to concept mapping, the conceptual processor constructs a conceptual structure (graph) which represents the meaning of the linguistic input. The nature of this graph depends upon the contents of the mental model and upon the word-to-concept mapping.

5. Metaphor Processing Without Mental Models

Metaphor and analogy have always been very closely associated in AI research. Consider a sentence such as (1).

(1) Peter's argument is full of holes.

If this sentence means anything, it does not mean what it says. The conventional way of producing a 'metaphorical' meaning is to assume that there is an underlying analogy which must be computed. What it means to compute an analogy depends on which knowledge representation scheme you are using but generally means something like analogical reasoning, inferencing, or transfering information from one domain to another. Since computing analogies is computationally expensive, metaphorical interpretations should not be generated for gibberish. Hence the emphasis in the work of computational linguists such as Jerry Hobbs and Jaime Carbonell has been twofold:[12]

---

[12] See Jerry Hobbs, "Metaphor, Schemata, and Selective Inferencing," Technical Report 204, SRI International, December 1979 and Jaime Carbonell, "Metaphor: An Inescapable Phenomenon in Natural Language Comprehension", Technical Report, Computer Science Department, Carnegie-Mellon University, May 1981.

1. Find criteria whereby ill-formed input is rejected and metaphors are accepted.[13]

2. Define the rules which govern what additional inferences may be drawn.

Metaphors are expensive to process and hence it is crucial that NLP systems are able to label input as metaphoric or non-metaphoric. Now, some metaphors signal their presence by violating semantic constraints. A sentence such as

(2) John hit the nail with a hammer.

fails to violate semantic constraints whereas a sentence such as (1) does since arguments do not 'literally' have holes. But a sentence such as (3) does not violate semantic constraints.

(3) Zeke's father is an accountant.

By most definitions of 'literal', (3) has a literal reading. But a metaphorical reading can also be generated, a reading in which attributes such as meticulous, finicky, boring, dull, and mousey are predicated of Zeke's father.[14] On the basis of the sentence alone, it is not possible to tell which reading of (3) is preferred. While the violation of semantic constraints may be used to detect some metaphors, it will not reveal them all. When multiple readings or interpretations are available, we say that a sentence is ambiguous and that disambiguation requires 'context.'

6. Mental Models and Metaphor

Mental models provide some conceptual clarity to some aspects of metaphor processing. I will examine three such aspects.

First, it is incorrect to appeal to 'context' as an aid in disambiguation. A user has no access to 'context' although he (potentially) has access to his mental model of the context.

---

[13] George Lakoff and Mark Johnson's *Metaphors We Live By* has been helpful on this score and its popularity among computational linguists is undoubtedly due to Lakoff and Johnson's suggestion that metaphors are systematic and not *ad hoc*.

[14] Such a metaphorical reading should be easy to generate for fans of *Monty Python's Flying Circus* who are familiar with their caricatures of accountants and bank clerks.

Second, it has become common to distinguish between 'dead' metaphors and 'live' metaphors. This distinction is made purely on the basis of the linguistic expression. A 'dead' metaphor, so the explanation goes, has acquired a fixed meaning through repeated use. Retrieving the meaning is simple: it only requires a table lookup. But since there seem to be no interesting research issues here, 'dead' metaphors have received little attention from computational linguists.

But a 'dead' metaphor is not dead for everyone. Children, for example, are frequently puzzled by a 'dead' metaphor such as 'out to lunch.'

(4) Charles is permanently out to lunch.

What is 'dead' and what is 'live' does not depend on the linguistic expression, but upon the mental model of the language processor. Since MMs are evolving models, we can use them to model this kind of change.

Third, it appears that some 'metaphors' can be processed without relying on analogical reasoning. Since each agent participates in multiple discourses, he possesses multiple mental models. An agent might even have a number of inconsistent models of the 'same domain.' Depending upon which model is running, there may or may not be a mapping from word to concept. Hence what was not a metaphor at time t may be a metaphor at time $t + n$ simply because another model is running.[15]

MMs allow us to make distinctions which cannot be made reliably otherwise. What is and is not a metaphor and what is a 'live' and what is a 'dead' metaphor cannot be decided just by looking at the linguistic expression. Nor can it be decided by looking at the expression and the agent. These determinations can only be made with respect to a particular mental model at a particular point in time.

---

[15] It may be helpful to think of Lakoff and Johnson's conceptual metaphors as inconsistent MMs of this type. Each one of their conceptual metaphors would have a different ontology. What is permissible in one ontology may be forbidden in another. The alternative to multiple ontologies is what we have now: one 'pure' ontology and lots of computation.

7. Conclusion

Mental models have been used as explanatory models for investigating the conceptualizations which individuals form of fairly structured domains. Little research has been done in using MMs in linguistics. Since CGs provide a basis for modeling MMs, it is now feasible to use MMs in computational linguistics. A linguistics based on mental models is in its infancy and many open questions remain. But MMs appear to offer a promising approach.