Metaphor, computing systems, and active learning

John M. Carroll and Robert L. Mack

Computer Science Department, IBM Watson Research Center, Yorktown Heights, NY 10598

(Received 17 January 1984)

Recent discussion has resolved the question of how prior knowledge organizes new learning into the technical definition and study of "metaphor". Some theorists have adopted an "operational" approach, focusing on the manifest effects of suggesting metaphoric comparisons to learners. Some have resolved the question formally into a "structural" definition of metaphor. However, structural and operation approaches typically ignore the goal-directed learner-initiated learning process through which metaphors become relevant and effective in learning. Taking this process seriously affords an analysis of metaphor that explains why metaphors are intrinsically open-ended and how their open-endedness stimulates the construction of mental models.

1. Introduction

Over the last few years, cognitive scientists have increasingly acknowledged the critical role of prior knowledge in acquiring new knowledge. Of course, the bald fact that prior knowledge limits and structures the acquisition of new knowledge was never "missed", but it was often "abstracted away from" as being a complication that theories of human cognition could not well accommodate. This theoretical consideration focused empirical work on simple and artificial concepts and skills, precisely to avoid the complication of prior knowledge.

This situation has changed substantially in current Cognitive Science. The problem of prior knowledge has become a central concern. There is growing interest in the study of real world problems of human knowledge and learning. Indeed, there is a natural confluence between the interest in the role of prior knowledge in learning, and the interest in studying real domains of human knowledge and skill. All of this converges directly upon the topic of metaphor and the design of the human–computer interface: can interfaces be designed to take advantage of the metaphors new users generate spontaneously as they apply their prior knowledge to this novel learning situation?

Traditionally, metaphor is a topic for students of rhetoric. A metaphor—so conceived—is a linguistic figure with the form "X is Y". Recent writing has more seriously considered the metaphoricity of thought underlying these verbal figures, not only in Cognitive Science (Lakoff & Johnson, 1980; Ortony, 1979) but in rhetoric itself: "Thought is metaphoric, and proceeds by comparison, and the metaphors of language derive therefrom." (Richards, 1965).

Several investigators have used this broadened concept of metaphor in their descriptions of cognitive intercourse with complex domains. Carroll & Thomas (1982) claimed that the activity of learning to use a computer system is structured by metaphoric comparisons. For example, the metaphor "a text editor is a typewriter" could be spontaneously referred to during early learning about text processors. Carroll & Thomas
(1982) argued that system interface designs succeed or fail with respect to their learnability depending on what metaphors they suggest and how helpful these are to learners. They cited Zloof's (1977) Query-by-Example system as a case in point: this system uses the "screen is a table" metaphor and has been found to be easy to learn (Thomas & Gould, 1975).

In this paper, we survey some of the current thinking about metaphor as it relates to this problem. Theories are reviewed under two headings, operational theories and structural theories. For both classes there appear to be open questions. Finally, we raise our own proposal: metaphors are kernel comparison statements whose primary function in learning is to stimulate active learner-initiated thought processes. Metaphors are open-ended because open-ended comparisons stimulate these processes more than explicit and comprehensive comparisons. The latter we would call "models." This active learning view of metaphor resolves several of the open questions we raise in the early sections of the paper.

2. "Operational" theories of metaphor

Educators have often observed that providing students with comparisons can help them learn. Robert Riekert, who teaches an adult course in computer literacy, uses the TV set metaphor to address his students' anxiety:

I usually begin by saying, "suppose I took you outside the room, told you to go in and turn on the computer switch under the right side of the computer... to press a key marked ENTER twice, and the word READY would appear on the screen. And you did that, and nothing happened. Your reaction might be, 'What did I do wrong? I don't understand computers!' "But if I told you to go into the room and turn on a TV set and dial Channel 4 and nothing happened, what would you do? You would probably see if the TV set were plugged in and your reaction might be, 'What's wrong with the TV?' The purpose of this course is to make you feel as comfortable and confident in the operation of the computer as you are with your home TV."

Riekert's students know little about TVs, but they are not frightened of TVs. The metaphor is designed to help them extend their fearless TV attitudes to computers.

Metaphors are also used to convey more purely denotative information to students. Riekert refers to storage locations as "buckets", using a physical metaphor for electronic storage. Keller (1982) uses a pegboard metaphor to introduce variable typing in Pascal ("we cannot put a round peg in a square hole"). Mayer (1975, 1976) has experimentally studied the effectiveness of metaphors in teaching programming concepts in Basic. He showed that many programming constructs could be learned more easily when they were presented in the context of a concrete metaphor referring to memory locations and transactions analogous to how the constructs actually worked in a program.

Most recent metaphor research in the area of human–computer interaction has been directed at end-user application domains like text-editing. For example, Foss, Rosson & Smith (1982) studied the task of learning to use a text editor by providing their subjects with an "advance organizer," presented prior to the actual system manual. The organizer explained how the computer could serve as a tool, follow commands, and create, store and retrieve files—referring extensively to the metaphor of conventional file folders stored and retrieved from file cabinets (e.g. lines can be added or
deleted from a computer file just as pages can be inserted and removed from a conventional file). Rumelhart & Norman (1981) employed three "models" in teaching people about text processing functions. The "secretary model" was used to explain that commands can be interspersed with text input. The "card file model" was used to describe the deletion of a single numbered line from a file. The "tape recorder model" was used to convey the need for explicit terminators in files.

It is clear these operational metaphors for text-editing exert a measurable effect on learning. Observational studies of actual learners have documented this extensively (Bott, 1979; Carroll & Mack, 1983, 1984; Mack, Lewis & Carroll, 1983). Typewriter knowledge is routinely invoked by learners first encountering text editors. Experimental studies are beginning to document these effects as well. Foss, Rosson & Smith (1982), in the experiment referred to above, found slightly better performance for subjects who were provided with the advance organizer (i.e., these subjects required slightly less time to complete performance tasks and used fewer commands).

However, as cognitive psychological analysis, the operational theories don't go far enough. For despite the learning efficacies they correctly anticipate, it remains obscure precisely how metaphor operates in the mind to achieve any of this. The operational theories offer examples of "good" metaphors, and sometimes of "bad" metaphors, but they offer no principles that could predict these properties extensibly. And in any case there is much more that one might like to know: which metaphors are "natural," "likely to arise," "possible," etc. The operational theories have little to say here. Such questions have led to the development of structural models of metaphor.

3. "Structural" theories of metaphor

In an effort to resolve issues regarding the underlying representational mechanism of metaphor, the notion has been given definition in relatively more structural terms (Bott, 1979; Gentner, 1983; Ortony, 1979). A typical definition might run like this:

Given two domains A and B, taking A as a metaphor for B is equivalent to providing a formal mapping from the primitives defining A into the primitives defining B.

Such a definition makes the metaphor question a question about representational formats, structural primitives, and the properties of formal mappings. From the standpoint of cognitive "process" these analyses reduce metaphor to primitive pattern matching operations defined over the elements and relations of a structural description.

A typical analysis is that of Gentner (1980, 1983), who has developed a "structure-mapping" analysis of metaphor. This view interprets metaphor as a mapping between two (graph theoretically expressed) domains, pairing the nodes of each. The relations of the two domains are constrained to be identical. One of Gentner's chief examples is Rutherford's famous analogy between the solar system and the structure of the hydrogen atom. The nodes of the former domain are the sun and the planets; those of the latter domain are the atomic nucleus and the electrons.

This structure-mapping is a successful metaphor because the relational predicates that apply to one set of nodes apply in like fashion to the other. Thus, the planets and the electrons revolve about the sun and the nucleus, respectively. The sun is more massive than any of the planets; and this relation too holds between the nucleus and
the electrons. In contrast to these common relational predicates, Gentner observes that attribute predicates generally fail to carry across the mapping. Thus, the sun is yellow and hot; while the nucleus may not be.

3.1. NEW METAPHOR CONCEPTS

This structural formulation of what metaphor is allows us to define many further concepts regarding metaphor relations. Gentner defines a variety of such relations: base specificity, clarity, richness, abstractness, systematicity, validity, exhaustiveness, transparency, and scope. “Base specificity” is defined as the extent to which the structure of the metaphor base, or source, is understood. “Clarity” refers to the precision of the node correspondences across the mapping. “Richness” is the density of predicates carried across the mapping. “Abstractness” refers to the level at which the relations carried across the mapping are defined. If they are the individual predicates of the base, the mapping is less abstract than if they are relations among predicates in the base.

Metaphors are “systematic” to the extent that the mapped relations are mutually constrained by membership in some structure of relations. In the Rutherford metaphor, the relations “attracts” and “orbits around” are systematized by their membership in the “inverse-square” relationship (Gentner, 1980; 1983). Metaphors are “valid” to the extent that the base relations carry their truth values across the mapping. “Base exhaustive” metaphors map each of their relations into the target (“target exhaustive” metaphors are defined analogously). Metaphors are “transparent” to the extent that it is obvious which relations in the base are able to be carried into the target. “Scope” refers to the extensibility of the mapping. For example, the Rutherford solar system analogy applies less usefully to heavier atoms.

This inventory of metaphor properties affords further explanatory power to the metaphor concept. A study of rated goodness for metaphors showed that both expressive (or literary) metaphors and explanatory (or scientific) metaphors were rated as better when they were higher in clarity (i.e., when the node correspondences were more self-evident). Explanatory metaphors, however, depended less on richness (i.e. on the density of predicates mapped). In this contrast, the concept of richness affords an analysis of the difference between expressive and explanatory metaphor. This concept and that of scope are used further in an attempt to analyse the difference between novices and experts with respect to the types of metaphorical models each engage.

... we speculate that a difference between experts and novices in a scientific domain is that the expert has an abstract global model with broad scope, while the novice has a pastiche of rich, only locally useful models. ... naïve models of science appear more like expressive analogies than do expert models (Gentner, 1980).

That is novices rely more on lower level relational predicates, applicable only in a limited domain—perhaps only in the current case. Expert models are more abstract and systematic.

3.2. THE PROBLEM OF ARBITRARY PRIMITIVES

Despite these promising achievements in providing a more articulated statement of the metaphor concept, structural theories suffer a variety of technical problems. First, the structural primitives posited by the theories can be shown to be arbitrary in ways that render some of their descriptive clarity specious. Second, the new concepts (albeit
precisely expressed by these theories) are sometimes problematic in that they map inconsistently onto the ordinary metaphor phenomena that motivated the analysis in the first place. Finally, the formal character of these theories has so far failed to actually resolve relevant empirical issues.

As noted, from Gentner’s structural rendering of the Rutherford solar system metaphor, one gathers a very neat result. However, scrutinizing her account reveals that much has been left out. One of the most salient relations holding between the sun and its planets is what we might call the “warms” relation: The sun warms the planets. Notably though, this relation does not hold between corresponding nodes of the atom model; the atomic nucleus of hydrogen does not warm the electrons (sub-atomic entities do not have temperature). Thus, the mapping as presented by Gentner appears to be far better than it in fact is.

This weakness becomes clearer when the relations she does emphasize are critically examined. For example, of the four predicates Gentner’s solar system metaphor analyses, two are “attracts” relations: the planets attract the sun and the sun attracts the planets; with the relations holding analogously in the atomic domain. For starters, why consider these two unidirectional relations instead of one bidirectional relation (i.e., the sun and the planets mutually attract)? As far as one can tell, there is no independently motivated principle to rule this possibility out for Gentner, except that it would decrease the apparent richness of the metaphor. But this begs the question.

Beyond this question though, it is unclear why “attracts” should be the predicate at all. Why not “gravitates toward” or “electromagnetically attracts”? Again there is no principle to rule these out; there is just the bald fact that doing so reduces the richness of a metaphor which seems intuitively rich. But given all these apparently arbitrary degrees of freedom one can ask just what the structure mapping analysis of metaphor is doing for us?

3.3. THE PROBLEM OF DESCRIBING REAL METAPHOR PHENOMENA

A second class of problems with these theories pertains to the concepts that are formally defined from structural representations. For example, consider Gentner’s definition of “transparency”, namely as the ease with which relations carried across the mapping are determinable. Clearly, a first order fact we want to get out is that “good” metaphors are transparent. However, in light of this, consider Gentner’s own remark that the familiar analogy of water waves to sound waves is rather non-transparent (Gentner, 1980). The trouble here is that the metaphor is well-established, attestedly useful—in short manifestly “good”. It is the theoretical concept of transparency that isn’t working well.

A similar problem crops up regarding Gentner’s concept of “base specificity”: “the degree to which the structure of the base is explicitly understood” (Gentner, 1980). Clearly, the success of a metaphor depends on having a familiar domain to analogize from and on recognizing enough in the new domain so that some correspondence can be established. But applying Gentner’s formal reconstruction of this ordinary language wisdom is problematic. She gives this as an example of a metaphor that fails on grounds of poor base specificity:

...sometimes in introductory chemistry texts, molecular bonding is explained by analogy with interpersonal attraction, e.g. ‘The lonely sodium ion searches for a compatible chloride ion.’ Interpersonal attraction is certainly familiar, but its rules are unfortunately unclear;
so this metaphor does not tell the student precisely what to map from the base. (Gentner, 1980).

The problem here is that the facts are wrong. The ion metaphor is well established, frequently used, popular, helpful for fledgling students of chemistry. A theory that tells us why it fails is making the wrong prediction.

Structural theorists have tended to focus only on similarity as a basis for comparison. That is, metaphors are "good" if and only if lots of nodes and relations map obviously between two domains. But however elaborately articulated, these are "counting" theories: matches are "good", mismatches are "bad". In traditional discussions more complexity is envisioned. For example, it is often noted that the disparity between a metaphor vehicle and its source can also be a potent factor contributing to the force of the metaphor (Deutsch, 1962; Richards, 1965). Richards' example is Hamlet's comparison of men to vermin through use of the word "crawling" in the remark, "What should such fellows as I do crawling between earth and heaven?" The comparison is salient as much in virtue of the differences between the compared entities as it is in virtue of their shared properties (e.g. locomotion).

An example from the history of physics shows that disparities in mappings can also operate effectively in scientific metaphors. Electrons are charged; and moving charges emit electromagnetic radiation which diminishes their kinetic energy. Thus revolving electrons would radiate away all their kinetic energy in a very short time, causing them to plunge into the nucleus. The fact that this does not happen posed a problem in classical physics: on this model all atomic matter should be unstable, with atoms collapsing in a release of ultraviolet radiation (see Misner, Thorne & Wheeler, 1973). Its resolution led to the overthrow of classical mechanics as an adequate model of atomic structure (on which the solar system model is based), and led to the prevailing quantum mechanical view of atomic phenomena.

A disparity in the mapped comparison, not a similarity, led to the insight. Note also that the metaphor in this case pertains to an attribute of electrons (their charge), as well as relations they enter into (e.g. revolving around the nucleus). Gentner draws a fundamental distinction between attributes and relations, claiming that only the latter are relevant to metaphor mappings. This is quite ironic, since couching the Rutherford solar system metaphor strictly on relational predicates obscures its role in defining this consequence of the metaphor—historically perhaps its most important derivative.

3.4. THE PROBLEM OF NON-EMPIRICAL FORMALISM

A third problem area for the structural theories, somewhat ironically, devolves from their formal character. For example, although these theories attempt to very precisely define primitive entities and relations, different theorists have isolated different primitive types. Gentner (1980), as we have seen, emphasizes node relations over node attributes, but Ortony (1979) does the reverse. This would be a useful contrast if it made any contact with empirical issues, but apparently it does not. In reviewing the various approaches, Gentner (1980) emphasizes their compatibility despite such differences. However, the availability of such a reconciliation only serves to underscore the fundamental arbitrariness of the various theories' formal expression.

Leaving aside the question of whether current structural theories can, in fact, provide adequate and univocal descriptions of metaphor relations, we must face the processing problem of how the node-node and relation-relation mappings are computed: struc-
tural theories like Gentner's (1980) merely assume (via cover terms like “transparency”) that these mappings obtain, but finesse the inevitable question of how. When these computational problems are faced, however, they might undermine the initial attractiveness of the structural theories. Winston (1980) implemented an analogy system that matches relations and attributes. He found that some classificatory pre-processing was required in order to avoid a computationally unmanageable process of brute force node matching. Winston's specific solution was to impose categorical pre-structuring on the program's input:

...the examples in this paper assume a beneficial teacher who gives only relevant facts and who does not deliberately try to confuse the system by shoveling detritus at it.

However, this is not how real learning occurs, and a theory of real learning will need to directly confront the processing problems of how corresponding nodes and relations are recognized as such.

At another extreme, the structural enterprise tends to grade into purely formal concerns with notational conventions. VanLehn & Brown (1981), for example, developed a representation for procedures called “planning nets” such that the maximal partial isomorphism between two planning nets reconstructs the ordinary language sense of “similarity” between the procedures they represent. Their approach abstracts away from issues of actual learner performance to an even greater extent than Gentner's: the transparency, for example, of a mapping is simply not a factor in determining the comparison relation. On the way to developing the planning net formalism, they briefly consider alternatives, for example flow-charting, but they argue

Since the intersection graph (the overlap in structural representation between two intuitively similar procedures) is so small relative to the difference subgraphs, a reasonable closeness metric would have to report that the two procedures are not very close—a false prediction.

What is “a reasonable closeness metric” though? Is the proportion of overlapping nodes and relations the only appropriate consideration? Mere proportions cannot represent the cognitive processes by which analogical relationships are used and understood. But divorced from these processes, such analyses of similarity become strictly formal.

4. Toward an “active learning” theory of metaphor

To summarize the discussion so far, there is general consensus that some notion of metaphor can be brought to bear on the question of how prior knowledge organizes new learning. Operational theories have focused on the manifest effects of suggesting metaphoric comparisons to learners and have produced a variety of concrete demonstrations, but they have failed to provide insight into the mental mechanisms of metaphor. Structural theories attempt to do this, but thus far have failed to provide non-arbitrary accounts. What is lacking in both operational and structural approaches is an account of the mechanisms of metaphorical understanding. Such an account would tell us why one or more metaphors are useful and how they are generated and then used to support learning.

Speculation about the mechanisms of metaphorical understanding is not new. Rational analyses can be traced back to Aristotle. The gist of these proposals is that the understander somehow makes inferences about how properties of something known
can be applied [e.g. "projected" in Black's (1979), terms] to something that is not known. But only recently have cognitive psychologists begun to specify mechanisms in psychological terms and to obtain empirical evidence for them, especially in the domain of human–computer interaction (as general references to the psychological study of metaphor, see Holyoak, 1983; Ortony, 1979; and Tourangeau & Sternberg, 1982; for studies of metaphor in the domain of human–computer interaction, see Carroll & Thomas, 1982 and Bott, 1979).

Attempts to describe the mechanism underlying metaphor have often appealed to a mnemonic role facilitating the encoding and later retrieval of new knowledge (Bott, 1979), Carroll & Thomas, 1982; Paivio, 1979). Carroll & Thomas (1982), for example, suggested an account that appealed to consolidation and integration of new information. On their account material to be learned is apprehended and, by hypothesis, entered into working memory. Next, and as an automatic consequence, a framework of related general knowledge (Minsky, 1973) is retrieved from long-term memory and also entered into the working memory. Finally, with the apprehension of further new material, there is a need to consolidate and compress the contents of working memory into a more integrated format. One way that this can happen is for the new material to be assimilated to the retrieved frameworks.

The appropriateness of the retrieved knowledge framework for the new material being assimilated is crucial to this account. The retrieved framework cannot be completely appropriate, for, if it were, the "new" material would be recognized not assimilated. Hence, the framework must be partially appropriate and partially inapppropriate. When it is not, additional mechanisms of inference come into play to modify the old structure to accommodate novel features of the new object of knowledge (see Bott, 1978, for further discussion of such mechanisms).

Our current view extends these ideas by focusing on the inferential processes that construct a description of some new object of knowledge from pre-existing knowledge. These processes are inferential in a broader sense than simply assimilating one thing to another. Their precise nature depends not only how understanders understand the metaphor in itself, but on the understanders' goals, and can involve interactions between the metaphor and evolving understanding of the new domain of knowledge. (See also Holyoak, 1983, for similar proposals in the domain of problem solving). In our view, mismatches (and dissimilarities) play as important a role as similarities.

Our analysis of metaphor tries to explicate the fact that metaphoric comparisons are at least partially open-ended; that is, that the structural mappings implied by the comparison are incomplete. Focussing on this aspect of metaphor leads us to a pragmatic analysis of the learning situation. Metaphor, in this view, is a particularly vivid indication of a general picture of learning as an active process.

4.1. ACTIVE LEARNING

Our conclusion that learning is an active process is based on research into how computer naive people learn to use text-editors. We have observed that learners prefer to try things out rather than read or follow structured step-by-step practice. Rote descriptions and practice are resisted. And even when learners try to comply with instructions, these prove difficult to follow and assimilate. But perhaps more important, instructions appear inadequate for communicating the detailed procedural and declarative knowledge needed for understanding and solving problems in a complex task domain like text editing.
Instead, we have found that learners resort to more heuristic reasoning processes for figuring out how to do things or solving problems or understanding the operation of text-editors. These processes included abduction (generating hypotheses on basis of very limited information; see Lewis & Mack, 1982) and adduction (verifying hypotheses within these same limitations of information; see Carroll & Mack, 1984), rather than more principled reasoning processes (like deduction or induction). These processes produce conclusions that are less constrained than deductive and inductive processes, strictly defined, but which open up the possibility of discovering knowledge on the basis of impoverished information.

Metaphors can facilitate active learning in this situation by providing clues for abductive and adductive inferences through which learners construct procedural knowledge of the computer. The learner begins with a canonical but open-ended comparison statement, “X is Y”, where Y has many more known properties than X. From this, and in transaction with task-specific goals, concerns, and expectations, the learner can generate and test a variety of hypotheses about X, that is can learn actively about X.

Consider again operational metaphors to card files, file cabinets, typewriters, tape recorders, etc. Not all properties of a typewriter can be carried over to a developing concept of a computer text processor. Some can (the layout and character-transmission function of the keys); some cannot (character keys cannot straightforwardly be overstruck using a text-editor); and some can be mapped from the typewriter base, but somewhat problematically (e.g. with respect to the storage of information, the tape recorder provides an alternate—and in some ways more accurate—metaphor).

The comparison of a text-editor with a typewriter carries all of these implications. The obvious similarities in function and form afford the metaphor in the first place: text editor learners almost never puzzle over what will happen when they strike a character key. In the context of such canonical and salient correspondences, the dissimilarities between the text editor and a typewriter become open questions—impelling further thought and leading then to further learning.

For example, keying two characters at the same location on a conventional typed page results in an overstrike. However, text-editors don't produce overstrikes (in this way). They either insert (i.e. place the new character adjacent to the old one, and adjust the text line accordingly) or replace (i.e., place the new character where the old one was—deleting the old one). Conventional typewriters, of course, don't have insert or replace capability: this is a clear dissimilarity in the metaphor. But this lack of transparency, clarity, richness, etc. is not functionally a limitation on the metaphor. Salient dissimilarities—in the context of salient similarities—stimulate thought and enhance the efficacy of the metaphor as a learning vehicle.

Consider an example from a computer system that is based on the metaphor of a desk top. In this system objects and their manipulations are represented concretely (at least on the surface): for example, to create a new document file, a user is prompted to initiate an action roughly described as "tearing off paper", in the context of an icon representing a pad of paper. One user we observed took the prompt quite literally. He tried to devise some action of "tearing" or sweeping the cursor across the icon representing the paper. In fact, the metaphor is misleading because actions applied to objects like files (or applications) must be selected in a more conventional fashion, from menus which describe the actions.

Was the metaphor a failure? In fact, the experience was informative because once the user understood how to select actions he changed his rule about how to do things.
At that point he understood that the metaphor was not to be taken too literally, but more important, when one wanted to do something, the action probably had to be selected from a menu.

This experience with a metaphor (supplied in part by the manufacturer, but elaborated by the user) was successful. Other examples demonstrate that the outcome of reasoning with a metaphor can be more complex. Here is a more complicated example described by Mack (1983). A secretary was trying to understand a design feature of a text-editor in which the end of a line is represented by a special character that can be inserted or deleted like an ordinary character, but which has surprising and unintuitive consequences. For example, entering a line end break (by hitting return) splits a line and puts everything to the right of the cursor on a new, inserted line. Deleting the line break for a given line causes the next line to be concatenated with it.

These effects are a complicated consequence of how text data is represented in this text-editor: Everything that is typed (including line break symbols) forms a "data stream". Line breaks cause the stream to be formatted into new lines at line breaks. Moreover, blank areas lying beyond line ends are not represented at all in the data stream.

One perplexing consequence occurs when one tries to append text to the end of a paragraph, and in doing so, types over the line break symbol: in this case the next blank line is concatenated to the original paragraph, as will subsequent lines and any text on them as one continues to type. This has the odd effect of "drawing up" and replacing material on subsequent lines. Needless to say, when a learner was asked to explain these effects, she was perplexed by them, and was not able to induce the underlying program model. Yet, she did formulate a "metaphor of the string" which seemed to explain part of what was happening.

According to this metaphor, what one typed occupied a fixed number of slots, like beads on a string. Trying to type over existing characters would replace what was there, and at the end of lines, this somehow caused characters afterwards to be "pulled up" like a string. Unfortunately, this metaphor did not explain why the blank area after the line break could not be filled up before encroaching on existing material. Nor did it explain why typing at or beyond line ends should begin to encroach on subsequent material. The metaphor was incomplete, but did lead to a positive outcome: the learner now had a rationale for the correct procedure which is to insert new material (make the "string" longer) instead of simply trying to type additional text starting from the end.

4.2. METAPHORS ARE OPEN-ENDED

We have argued that metaphors function to stimulate processes of inference that try to apply not only similarities between known and unknown object of knowledge, but also make use of dissimilarities and omissions in metaphor, relative to the new object of knowledge. Put another way, metaphors are open-ended: that is, not literal descriptions, but inherently incomplete, even indeterminate, not for lack of wit or by the fiat of intention, but by definition.

While this view may not be incompatible with structural theories in principle, it does nag at the attempt to develop explicit mapping representations for metaphors. It seems to us inherent in the nature of metaphor that its relation to a metaphorized
object or domain be not just incomplete, but indeterminate. Certainly this trait of indeterminacy is true of expressive and literacy metaphor (Frye, 1966; Richards, 1925), but we would argue that it is equally true of metaphors in more mundane areas, such as learning to use a computing system. It is this property of metaphor that affords cognitively constructive processes which can lead to new knowledge. From the perspective of active learning, the open-endedness of these kernel comparisons is intrinsic to the mechanism that allows them to work. In this case, learners are introduced to properties of text editor formatting and data-stream representation as problems they pose to themselves.

Consider, for example, what a learner gleaned from the comparison of a text file and a card file. In the context of trying to learn to edit computer files, there is a highly salient correspondence: when one card (line) is removed (deleted) its place in the deck (file) is filled by the next card (line). But having gleaned this correspondence raises all sorts of further questions which impel further thought and lead to further learning. To unpack the implications of this basic insight requires discovering many properties of computer text files which are either left out of the card file comparison, or would not be specifically illuminated by it. For example, the particular subparts of each complex object (card file versus computer text file) do not match in an obvious way. A single line on an index card does not literally match the video object corresponding to a line of a file, but corresponds only in an open-ended and rather abstract way that must be separately spelled out.

Many other elements of the computer file must be discovered on their own terms, not through tracing out mappings with the metaphor. To cite one example: computer file lines are numbered and renumbered as lines are inserted or deleted, quite unlike cards. The renumbering and adjusting of computer file lines (scrolling up with deleted lines or scrolling down with inserted lines) must be understood in its own right. Our familiarity with computer files may make it easy to presuppose that these mappings are obvious (card versus video line, card position versus line number, etc.) but this may be to beg important questions.

To summarize, the relationships comprising a metaphor comparison are not "read off", but constructed by actively working through the implications of the metaphor. There is of course a pedagogical application of metaphors in which they serve to codify and communicate new knowledge in a comprehensible way for new learners. Here metaphors may be carefully crafted and presented in a way that helps the learner make just the connections the instructor wants the learner to make. But this role for metaphor should not mislead us into thinking that all metaphors function in this seemingly straightforward way: that is, where the connections between the metaphor and new object of knowledge are clear cut and seemingly "read off".

The typewriter metaphor, in particular, provides an orienting framework but leaves many operational details for the learner to discover. The metaphor orientates the learner (the computer novice in this example) to hypothesize and verify similarities in structure and function, and to have a framework for recognizing and analyzing discrepancies. In this way, the metaphor serves to initiate more systematic processes which construct more complete procedural knowledge representations specifically appropriate to the text editor. But the metaphor itself "seeds" these constructive processes and may continue to serve as an orienting framework through the learner's interaction with the specific properties of the computer. Accordingly, we cannot enumerate the efficacies
of a metaphor just by counting mapped relations. Rather, we must take into account the learning process within which these relations have real meaning.

4.3. WHERE METAPHORS COME FROM

Active learning implies that effective metaphors will be self-generated by learners, and that working out the implications of these metaphors will be a principal process of learning. Self-generated metaphors, because they are developed spontaneously, and without apparent effort, could be regarded simply as "givens" of the learning process, as intrinsic to it or as intrinsic motivators of it, and therefore not as sources of additional processing demands. Other-generated metaphors, because they must themselves be explicitly comprehended by learners, raise other issues: whatever learning burden is associated with the metaphor itself must trade off favourably with the savings in learning burden derived from employing the metaphor subsequently.

There is indirect evidence for this in that it is often found that learners cannot make use of metaphors that they are taught. There are cases where metaphors are available but not recognized as such, or not applied. Gick & Holyoak (1980, 1983) went to great lengths to induce students to apply a story scenario which they understood to solve a problem whose form was analogous to the scenario in the story. Students were able to exploit the analogical connection when informed that there was one, but were much less likely to see or exploit the connection spontaneously. It remains unclear what motivates a learner to use a metaphor, or even consider the possibility that one might be relevant to understanding.

However, it is clear that there are risks inherent in self-generated metaphors which may not be appropriate for understanding. Gentner & Gentner (1980) provide an example in case of students trying to understand electrical circuits. Students who generated metaphors referring to flowing water were better able to understand properties of batteries and their serial or parallel combination than students who thought in terms of masses of moving objects (e.g. crowds of people). Conversely, students who generated metaphors involving the idea of moving objects were better able to understand aspects of circuits that the flowing water metaphor could not adequately describe. No one metaphor was sufficient for understanding all properties of the circuit (or changes made in it): failing to choose the right one made it hard to understand one or more features.

Bott (1979) provides an example in the domain of text-editing. Consider the contrast between the "pigeonhole" metaphor with the "card file" metaphor in understanding the structure of a text file in a line editing environment. In the former, printing consists of fetching the contents of a pigeonhole (a line of text) and displaying it. In the latter, printing consists of fetching a card from a card file and displaying what is written on it (also a line of the text file). To this extent, the metaphors are identical.

A contrast emerges in the case of deletion. Deleting a line of text, in the pigeonhole metaphor, amounts to fetching the contents of a pigeonhole and discarding it—*but the pigeonhole itself remains* (albeit empty). In the card file metaphor, deletion consists in simply removing a card from the file. On the former metaphor, a learner would expect that the operation of deleting a line would have the effect of leaving a blank line in the file. On the latter, this expectation would not occur. Bott found that the pigeonhole metaphor spontaneously occurs to learners and is not given up easily, however it makes the wrong predictions for deletion. Bott presumes that in the typical case successful metaphors will be explicitly taught by a teacher.
Sometimes conflicting metaphors are generated. The typing page display of a text editor is like a paper typing page, but data stored in a text-editor is also very much like tape recorded information. Both metaphors illuminate the new concept of computer storage, but the two are also somewhat inconsistent. Text-editor files can undergo insert and replace, which is more like splicing and recording over tape than like cutting and pasting a paper page. Neither metaphor covers well the possibility of reformatting a text-editor document.

Carroll & Lasher (1981) described a learner who maintained two distinct self-generated metaphors for the internal operation of a hand calculator. In one of these, the calculator kept track of data entry by internally “typing out” inputs. In the other, the calculator maintained an on-going computation which incorporated each successive input in real time. The subject used the typing metaphor to account for error correction as being a matter of “whiting out”. The on-going computation metaphor explained why when data entry is complete, all calculations are immediately available. The metaphors are at least somewhat inconsistent though: if information can be “whited out”, it must be isolably stored, not incorporated into an on-going process. Nevertheless, the subject generated and used both of these metaphors to understand the calculator.

5. Toward a pragmatics of metaphor

We have argued that understanding the role of metaphor in learning requires more than a structural description of two objects and how the relations in one map onto the relations in the other. It requires understanding how mental mechanisms of active learning, in conjunction with metaphorical descriptions, provide the means for understanding some new knowledge domain.

Who the metaphor will be used by and the ends for which it was designed are also crucial components of metaphor in this view. Structural descriptions of corresponding domains in a comparison relation provide only an abstract set of possible mappings. The actual relevance of any of these mappings to a real and usable metaphor depends fundamentally on the needs and goals of the learner. Put another way, we need to understand the pragmatics of the learning situation.

The learning efficacy of a metaphor is not only a matter of counting atomic correspondences, but an intricate balancing of correspondence, non-correspondence, and indeterminate correspondence. The relation between the learner and this intricate balancing is one of “transactions”, not a simple mapping in toto. Relations between the metaphor source and the target are brought into correspondence through the course of a process of thought. For example, when the keyboard of a text-editor is recognized as similar to that of a more familiar typewriter, this piece of relation can be mapped. Depending upon the learner’s goals and expectations, this partial mapping itself might encourage further mapping—or further metaphors. But it would never be the case that a complete node-correspondence mapping would be carried out non-transactionally and without intermittent reference to overriding goals and current actions.

5.1. STRUCTURAL AND OPERATIONAL METAPHORS

This view of metaphor is quite compatible with what we earlier called the operational theories. However, the process of self-initiated learning we have described departs from these views both in being more active [recall Carroll & Thomas’s (1982) notions of consolidation and integration] and in being more concretely grounded in real and
observable learning behaviour (Carroll & Mack, 1983, 1984). Where the operational theories focus chiefly on demonstrations that metaphor enhances learning, our analysis seeks to clarify how that advantage becomes manifest in active learning.

This view is also compatible with the structural theories. However, it provides an approach to the problem of arbitrary assumptions, which as we have argued threatens to deprive the structural theories of any serious content. Recall, for example, that some of the formal properties Gentner defined over her graph structures, seemed paradoxically related to the actual efficacy of metaphors for comprehension. Whereas Gentner showed formally why the metaphor of interpersonal attraction ought to be a bad account of ion bonding (because of its low base specificity), the fact is that it is a good metaphor. What seems to be relevant is the goal of trying to understand activity directed toward bonding. Having this goal in mind makes salient the analogy of human pair bonding and ion bonding. Base specificity per se is only a secondary consideration.

Another problematic case we considered earlier was that of relations. Gentner’s analysis arbitrarily suppressed, for example, the “warms” relation in her reconstruction of the Rutherford solar system analogy. On the active learning view, it becomes clear—and non-arbitrary—that the “warms” relation is irrelevant to the comparison being made in this metaphor. The goal of a learner employing this metaphor is to better understand the spatial properties of the hydrogen atom, and the manner in which they are dictated and maintained by force relationships between the various constituents. From the viewpoint of this goal, “colour”, “temperature”, and a host of other physical dimensions are irrelevant. But the analyst discovers this by including the learner’s goals within the analysis, not by arbitrarily preselecting only the relevant dimensions in the analysis. Active learning affords an explanation of what Gentner was forced to merely stipulate.

Holyoak (1983) has made similar criticisms regarding purely structural definitions of what properties of an analogy or metaphor will be relevant. He cites an example in which an attribute of an object was critical to understanding something new about another object. More generally, Holyoak argues that the relevance of attributes or relationships is determined by the casual role that they play in some situation. Attributes may be just as relevant as relations.

This pragmatic conception of metaphor makes it unsurprising that learners sometimes compose two, or even more, base domains (as did the learner studied by Carroll and Lasher, 1981). From the unembellished structural viewpoint, composite metaphors are cases in which the mapping from the target to the base domains is formally ill-defined. But as we have noted, these are not cases where comparisons become hopelessly unresolvable. They are cases in which a richer basis for open-ended comparison is accessed.

More generally, we should not expect that the description of a metaphor to be unique. There may be multiple descriptions serving different purposes for different goals. Holyoak (1983), for example, argues that analogical connections may be made in terms of concrete connections between target to base domains, or they may be made at a more abstract level in which target and base are understood as instances of more abstract generalization. Both levels of description presumably have advantages and disadvantages, and serve different goals. For example, abstracting a general schema may help the learner to disentangle their understanding from the myriad irrelevant details of the target and base, details which might otherwise obscure similarities between
the two domains. On the other hand, a general schema may be too abstract, so that its relevance to a specific instantiation obscure.

The problem of metaphor is intimately bound up with the transaction between agents and their current goals. This transaction is often advanced by reconceptualizing the current situation in terms of usefully related other situations. Among all the potential relationships between a target and a source there will be some that serve to illuminate properties of the target that are relevant to current goals, and some that do not. Purely operational theories overlook this and deal with the outcomes per se. Structural theories can only deal with this by positing further arbitrary relations in order to generate the a posteriori appropriate mapping relations (e.g. systematicity principle). But this unending resort to additional degrees of freedom undermines the explanatory efficacy of these theories.

The resolution of this, we suggest, lies in elaborating current analyses of metaphor to include explicit reference to the role of metaphor in active learning and to the personal goals that drive this process. There can be no purely structural index of the usefulness of a metaphor for understanding some novel entity. For this will always be a joint function of potential relationships between the target and metaphor domains, and the goals and needs of the learner.

5.2. MODELS AND METAPHORS

The active learning picture of metaphor permits us to sharpen the pre-theoretical distinction between models and metaphors. The distinction we envision resides chiefly in the open-endedness, incompleteness, and inconsistent validity of metaphoric comparisons versus the explicitness, comprehensiveness and validity of models. This contrast is not a dichotomy. Models typically rely on a kernel metaphor, or metaphors, and hence stimulate many of the same comprehension processes with respect to active learning (see also Miller, 1979, for related views on the generality of processes for comprehending literal and metaphorical text).

On the one hand, a model is a description of a target domain or object that seeks to faithfully represent the actual elements, relations and mechanisms that are constitutive of objects in that domain. Although models are necessarily selective and abstract, and accordingly to some extent incomplete, their incompleteness is not the same as that of metaphors. For while models are designed to represent some target domain, metaphors are chosen or designed to invite comparisons and implications which are not literally true. Metaphors are not “right” or “wrong” descriptions, as models are: rather they are “stimulating” (or unstimulating) invitations to see a target domain in a new light. Black’s (1979) “interactionist” view of metaphor understanding develops this point of view in the case of interpreting metaphorical expressions (see also Tourangeau & Sternberg, 1982).

On the other hand, a practical complication enters into this contrast when we consider a model as codified in some physical representation. In such a case, understanding the model presupposes interpreting the representational format of its presentation. The latter can be in part a problem of metaphor. The problem can be relatively trivial, as when the relationship between representational format and the model itself is explicit and even conventional; examples would be following flow-charts or recognizing conventions of perspective in drawings. But the familiarity of these representational formats should not obscure the cognitive mechanisms presupposed in interpreting their
implications for respective target domains, as cross-cultural research has shown in the case of pictorial representation (see Cole & Scribner, 1974).

Apart from presupposing that learners understand particular representational formats, learning models may simply not be possible. Recall the example of the "metaphor of the string". The secretary was perplexed by the surface effects of a rather complex editor feature: the data stream. What would it mean to provide a literal model of the data stream to this secretary? Taken literally, it would mean introducing programming concepts that are technical and difficult to understand in their own right.

This view of the distinction between model and metaphor has immediate implications for what Halasz & Moran (1982) have called "conceptual models", and their role in learning. Halasz and Moran contrast conceptual models with metaphoric, or analogical, models. By the latter, they intend suggestive but typically incomplete descriptions referring to near-neighbor domains, or to compositions of these. In contrast, their view of conceptual models is intended to cover highly accurate and arbitrarily complete descriptions, usually in some abstract format, like a flow-chart or a graph. They cite du Boulay, O'Shea & Monk (1981), Moran (1981), and Young (1981) as examples.

Halasz & Moran (1982) try to make the argument that conceptual models have a distinct advantage over analogical models for learning, namely that they are truly appropriate vis-a-vis the target domain, indeed to arbitrary limits, by design. Analogical models, by their very definition, cannot be truly appropriate. They are at best only roughly appropriate as descriptions. Unfortunately, this argument suffers two serious difficulties: First, as we have observed above, so-called conceptual models are not at all dichotomous with metaphors. The Platonic form of a conceptual model may indeed be so abstract as to be metaphor-free. But when the conceptual model is codified on paper as a graph or a chart, as it would have to be in order to function as an instructional tool, its interpretation requires prior knowledge about such representational formats and their characteristic interpretation. To the extent that this process is not automatic and determinate, it engages metaphor.

The second difficulty with the Halasz & Moran (1982) argument is that highly accurate and arbitrarily complete descriptions may simply not be optimal learning aids (Jones, 1982). The view that they are presupposes a passive learning theory, which indeed is presupposed in many state-of-the-art approaches to instruction, but which we have found to be inconsistent with the propensities and capacities of actual learners trying to master text-editing systems (Mack, Lewis & Carroll, 1983). Indeed, in our view, a metaphor that truly provided a completely transparent and comprehensive mapping of the target domain would be a bad metaphor for learners—there would be nothing left for the learner to do.

Finally, it is important to stress that the learning outcome we—along with Halasz & Moran (1982)—imagine, is that a relatively complete, explicit and valid model of the target domain is attained. They suggest that this can be achieved by explicitly presenting a representation of the model to the (passive) learner. In contrast, we suggest that metaphor can stimulate the learner to actively self-generate a model.

Figure 1 schematizes the role of metaphor as mediator in the construction of a knowledge representation for some target domain. The outcome of the comprehension process is not an understanding of some metaphorical description of the target domain per se, but rather a representation of the domain itself—that is to say, a mental model. The role of metaphor may vary, but crucially, it serves to draw the comprehender's
attention to certain critical elements or relations of the target domain—elements or relations which will be critical components of the final representation but ones which cannot perhaps be readily understood on their own terms.

In The poetics, Aristotle wrote, “The greatest thing by far is to have a command of metaphor”. But he was quite frankly dubious about human prospects for ever really getting an analytic command of metaphor: “This alone cannot be imparted to another: it is the mark of genius, for to make good metaphors implies an eye for resemblances”. If current discussions of complex learning provide any gauge, Aristotle scarcely overestimated the importance of the concept of metaphor. What is troubling is the possibility that he has also not overestimated our prospects for developing an analytic understanding of metaphor.

We cannot know our ultimate prospects; the cognitive study of metaphor is really only beginning. However, we can be certain that the special nature of metaphor will proscribe theories of metaphor. For the nature of a domain always dictates the type of theory we may hope to develop to describe that domain. In this paper, we have tried to stress that a metaphor is more than mere resemblance. As Aristotle said, it implies “an eye for resemblance”; this “eye” (for us, the mind of the learner) takes an open-ended, incomplete—even indeterminate—kernel comparison, and makes it a focus for the self-initiated construction of new knowledge. A metaphor is a crystal seed disturbing a saturated suspension to precipitate the formation of a growing crystal.

We are grateful to Linda Misek-Falkoff, Clayton Lewis, Mary Beth Rosson, and John Thomas for discussions and clarifications of metaphor, and for comments on this paper. This paper was presented at the IEEE International Conference on Cybernetics and Society, as part of the Symposium on Metaphor and Computers, on October 28, 1982, in Seattle, Washington.

References


