

METAPHOR INTERPRETATION AS SELECTIVE INFERENCING: COGNITIVE PROCESSES IN UNDERSTANDING METAPHOR* (PART 1)**

JERRY R. HOBBS
SRI International

ABSTRACT

The importance of spatial and other metaphors is demonstrated. An approach to interpreting metaphor in a computational framework is described, based on the idea of selective inferencing, in which a processor draws or refrains from drawing certain inferences in a controlled fashion. Two examples of metaphors are examined in detail in this light—a simple metaphor and a spatial metaphor schema. In Part 2 a novel metaphor will be examined and there will be a discussion, from this perspective, of some classic issues concerning metaphor, including the analogical processes that underlie metaphor, the stages in the life of a metaphor, and the definition of metaphor.

METAPHOR IS PERVASIVE

I. A. Richards, in speaking of metaphor, said, "Literal language is rare outside the central parts of the sciences" [1]. But it is rare even in the central parts of the sciences. Consider for example the following text from computer science. It comes from an algorithm description in the first volume of Knuth's *Art of Computer Programming* and is but one step removed from the domain's most formal mode of expression.

Given a pointer P0, this algorithm sets the MARK field to 1 in NODE(P0) and in every other node which can be reached from NODE(P0) by a chain of ALINK and BLINK pointers in nodes with ATOM = MARK

*This work was supported by the National Science Foundation under Grant No. MCS-78-07121, by the Advanced Research Projects Agency under Contract No. N00039-79-C-0118 with the Naval Electronics Systems Command, and by the National Library of Medicine under Grant No. 2R01 LM03611-03.

**Part 2 of this article will appear in *Empirical Studies of the Arts*, Vol. 1(2).

= 0. The algorithm uses three pointer variables, T, Q, and P, and modifies the links and control bits during its execution in such a way that all ATOM, ALINK, and BLINK fields are restored to their original settings after completion, although they may be changed temporarily [2, p. 417].

In this text, the algorithm, or the processor that executes it, is apparently a purposive agent that can perform such actions as receiving pointers; setting, changing, and restoring fields; reaching nodes; using variables for some purpose; modifying links and bits; and executing and completing its task.

Nodes are apparently locations that can be linked and strung into paths by pointers and visited by the processor-agent.

Nodes also seem to be containers that can contain fields.

Fields are also containers that can contain pointers, among other things. In addition, fields are entities that can be placed at, or set to, locations on the number scale or in the structure collection of nodes.

Pointers, by their very name, suggest objects that can point to a location for the sake of some agent's information.

In fact, there is very little in the paragraph that does not rest on *some* spatial or agent metaphor. Moreover, these are not simple isolated metaphors; they are examples of large-scale "metaphor schemata," or "root metaphors" [3], which we use to encode and organize our knowledge about the objects of computer science. They are so deeply ingrained that their metaphorical character generally escapes our notice.¹

The pervasiveness of metaphor was noted as early as the eighteenth century by Giambattista Vico [4] and somewhat later by Jeremy Bentham [5]. In our century, this observation has been the basis for a rejection of Aristotle's and Quintillian's views that metaphor is mere ornament, and an elevation of metaphor to an "omnipresent principle of language" [1] and "the law of its life." [6] Richards argued that metaphor involved complex interactions between two domains, which he called the "tenor," that which is being described, and the "vehicle," that which it is being described in terms of. The tenor is seen in a perspective provided by the vehicle, either bringing to the fore certain aspects of the tenor or allowing the tenor to be viewed in ways that would not have been possible without the metaphor.

As we saw in our example, the spatial metaphor especially is pervasive. Jespersen remarked on this [7]. For Whorf it was a key element in his view that language determines thought: the spatial metaphors provided by one's

¹ I have occasionally had a computer scientist argue that some of the metaphors, e.g. the "variable as container" metaphor, were not metaphors at all but true descriptions of physical reality. To see that this is not the case, note that when we place a value in a variable, its previous value is no longer there; we did not have to remove it. (I once had a beginning FORTRAN student who was puzzled by this very fact. He had not yet learned the limits of the metaphor.)

language determines how one will normally conceptualize abstract domains [8]. Urban saw in the use of originally spatial words for more abstract concepts an "upward movement" of language from the physical to the spiritual [9]. More recently, Clark examined the physical and psychological motivations behind our most common spatial metaphors for time [10]. I have previously attempted to exploit the pervasiveness of metaphor in a computational framework [11, 12]; the present article continues the attempt. In Jackendoff we find a similar effort in theoretical linguistics [13]. The most extensive recent treatment of metaphor in everyday language is found in Lakoff and Johnson [3]; they identify the root metaphors that underlie our thinking about a vast array of domains, and argue that we can understand the domains only by means of these metaphors. The fundamental insight that informs all this work is this: metaphor is pervasive in everyday discourse and is essential in our conceptualizations of abstract domains.

In this article I wish to explore how metaphors and metaphor schemata might be treated in a computational setting, from the perspective of artificial intelligence, in a way that accommodates the fundamental insight. In the next section, the framework of selective inferencing is presented, and certain interesting previous proposals concerning metaphor are examined within this framework. Then, three successively more difficult examples of metaphors are considered—first a simple metaphor, next a metaphor schema that has become a part of the language, and finally a novel metaphor. The aim is to discover some of what is needed to represent and reason about metaphorical usage. Finally, a number of issues of classical interest are examined in light of this approach. (The novel metaphor and the discussion of issues will appear in Part 2 of this article.)

SELECTIVE INFERENCING AND ITS RELATION TO PREVIOUS APPROACHES

One of the principal thrusts of natural language processing research in artificial intelligence in the last decade has been to develop systems that allow inferences to be drawn selectively [12, 14-17]. One reason that such systems are needed is that it is difficult, if not impossible, to axiomatize in a consistent manner any domain more complex than set theory. Workers as early as Collins and Quillian [18] noticed that it is a very powerful device to allow the following inconsistent set of axioms:

- (1) $\text{bird}(x) \longrightarrow \text{fly}(x)$
- $\text{ostrich}(x) \longrightarrow \text{bird}(x)$
- $\text{ostrich}(x) \longrightarrow \sim \text{fly}(x)$

This is a much more economical representation than replacing the first of these axioms with something like

- (2) $\text{bird}(x) \ \& \ \sim \text{ostrich}(x) \ \& \ \sim \text{penguin}(x) \ \& \ \sim \text{kiwi}(x) \ \& \ \sim \text{emu}(x)$
 $\ \& \ \dots \ \& \ \sim \text{injured}(\text{wing}(x)) \ \& \ \sim \text{dead}(x) \ \& \ \sim \text{newborn}(x)$
 $\ \& \ \dots \ \longrightarrow \text{fly}(x)$

The idea is that one can draw an inference as long as it does not result in an inconsistency, and that when an inconsistency does result, some means must be applied to decide among the inconsistent inferences. McDermott and Doyle [19] have developed a nonmonotonic logic in which the various exceptions of (2) are encoded with a special operator *M*, meaning "it is not inconsistent to assume that." Thus, (2) would be written

$$\text{bird}(x) \ \& \ M \text{ fly}(x) \ \longrightarrow \text{fly}(x).$$

That is, if *x* is a bird and it is not inconsistent that *x* flies, then *x* flies. Reiter investigates a similar approach [20].

For the purposes of this paper, however, it will be more convenient to keep the simple notation of (1) and complicate the calculus that manipulates it. For there are further reasons beyond the avoidance of inconsistency to be selective in the inferences one draws—there are too many true inferences that can be drawn in a specific situation and most of them are irrelevant. Consider the following text:

John couldn't find Mary's house. He drove up one street and down another.

Among the inferences normally relevant to understanding this text are the facts that

Houses are visible objects.
 Houses are located on streets.
 People live in houses.

There are many more facts however that are not ordinarily relevant, and should not be "activated" by a natural language processing system. For example:

Houses have roofs.
 A house has a living room, a kitchen, several bedrooms, and one or more bathrooms.
 Houses contain furniture.
 Houses are made of such materials as wood, brick, stucco.
 Termites sometimes attack wooden parts of houses.
 Houses have exterior faucets to which hoses can be attached.
 Houses tend to rise in value.

All of these things may be true of Mary's house, but the text does not require them in any way.

A great deal of work in natural language processing can be viewed as addressing the problem of using the discourse itself to determine which inferences are relevant. For example, Grosz examines the clues in task-oriented dialogs that signal a shift to another part of the task and hence another part of the knowledge base [15]. Mann, Moore and Levin use what is explicit in an utterance to choose a "dialog game," and then assume that what is encoded in that dialog game is relevant [21]. Work on frame or script recognition [e.g., 16] is in a similar vein.

In my own work [e.g., 12] I have investigated the idea that the inferences that it is relevant to draw are the inferences required to solve various discourse problems, like recognizing the coherence structure of the text, forcing congruence between predicates and their arguments, and anaphora and ambiguity resolution.

To take a simple (literal) example, consider following definite noun phrase resolution problem. Suppose we hear

- (3) John picked up a book.

It is sometimes true that a book has an index, and sometimes it is relevant. But there is no reason, given (3) alone, that we would necessarily want to draw the inference that John's book has an index. However, if the next sentence in the text is

He turned to the index,

then we can be sure that the inference is both true and relevant. Resolution of the definite noun phrase "the index" requires us to draw the inference that the book mentioned in (3) has an index. Note that we still need the normative knowledge that a book often has an index, even though the text mentions an index explicitly. If John had turned to the door, we would not have assumed the book had a door.

For the rest of the paper we shall be assuming a natural language processing system, such as that described in Hobbs [12], and exemplified by the DIANA system implemented at SRI [22]. This system accepts a text translated by a syntactic front-end into predicate calculus formulae and draws those inferences necessary to solve the discourse problems posed by the text. The inferencing process is selective and driven by a collection of *discourse operations* that try to do such things as resolve pronoun and definite noun phrase references, find the specific interpretations of general predicates in context ("predicate interpretation"), reconstruct the implicit relation between the nouns in compound nominals, and recognize coherence relations between successive portions of

the text. The operations select inferences from a large collection of axioms representing knowledge of the world and the language. Associated with the potential inferences are measures of salience that change as the context changes. These help determine which inferences are drawn by the operations and hence how the text is interpreted. The control structure is such that the system does not try to solve the discourse problems independently, but rather seeks the most economical interpretation of the sentence as a whole. A more thorough description of the principles underlying this system can be found in Hobbs [12, 23-25] and its output is exhibited in Hobbs [22]. The two examples in the next section have been interpreted by the DIANA system, although not with a realistically large knowledge base. (It is important to note that the same problem still faces us in these examples as faces us in all of discourse interpretation—how to navigate in a flexible, efficient manner through the sea of possible inferences.)

There are at least two questions one might ask about a metaphor: What is its meaning? And what processes are involved in its comprehension? The first is a philosophical question and will not concern us here. It is the second which is addressed in this paper, and it is a psychological question. The psychological relevance of the answer depends on the computer metaphor for human cognition—the mind is a computer—but also on the level at which the answer is described. In artificial intelligence, there are three levels at which one can discuss one's work. One can simply discuss it in English prose, but the danger here is that the "procedures" described will not in fact be computable, and the point of appealing to the computer metaphor will be lost. Or one can discuss the actual code of some implemented system. However, in spite of the power of the computer metaphor, the brain and the present-day computer are sufficiently different that a discussion at this level of detail is of no psychological interest, whatever its technological merits. In this paper I aim for an intermediate position; I try to proceed at a formal enough level that the processes described can be imagined as computable, yet at an abstract enough level that the account is not *prima facie* implausible as a psychological account.

It is often advanced as an argument against a particular formal approach that it does not take context into consideration. As Black has emphasized, metaphors occur in some context and must be interpreted in that context [26]. It does not make sense to ask about the interpretation of a metaphor outside of context. That is not an argument against the approach used here. On the contrary, the framework outlined above is specifically designed to formalize a notion of context, and to provide a way of interpreting expressions in context.

Finally, it should be pointed out that metaphors operate primarily at the conceptual level, and we shall be dealing at all times at the conceptual level, not at the surface linguistic level. At the conceptual level, we talk about "predicates," not "words." Although we will generally have, for every word, a predicate of the same name, the predicate should not be thought of as exhausting what is conveyed and suggested by the word. Rather, we should think of

the word as corresponding to the possible sets of inferences that might be drawn because the word has been used in a particular context. That is, words do not merely translate into a single expression in a formal notation; they trigger an inference process that could result in any one of a large set of possible expansions in this notation. Hence, we have not stripped words of their mysterious quality, but rather translated the mystery into the mystery of choosing the right set of inferences.

Much previous work on metaphor can be seen to fall within this framework. In *The Art of Rhetoric* (III.II.12), Aristotle said that "clever enigmas furnish good metaphors; for metaphor is a kind of enigma." [27] In a sense, then, the idea of metaphor interpretation as problem solving—like most other ideas—is originally due to Aristotle.

More recently, in computational linguistics, the earliest detailed proposal for handling metaphor was that of Russell [28]. Her proposal concerns abstract uses of verbs of motion and involves lifting selectional constraints on the arguments of the verb while keeping fixed the topological properties of the motion, such as source, path, and goal. Thus, to handle "the ship plowed through the sea," one lifts the restriction on "plow" that the medium be earth and keeps the property that the motion is in a substantially straight line through some medium. Russell exemplifies an approach that finds its most complete development in the work of Levin [29], but it is also seen in linguistics in the work of Matthews [30] and Kahn [31]. Metaphor is treated as a species of semantic deviance; selectional constraints are lifted until the expression can plow through the interpreter without difficulty. One can view a selectional constraint as a particular kind of inference. Thus,

(4) plow-through(x,y) \longrightarrow earth(y)

That is, if x plows through y, then y is earth. In terms of selective inferencing, lifting this constraint is equivalent to not using (4) to draw an inference about the substance that is being plowed.

But the problem of interpreting "the ship plowed through the sea" is not just to avoid rejecting the sentence because the sea is not earth, but to notice the similarity of the wedge-shaped plow and the wedge-shaped bow of a ship and the wake that each leaves, and perhaps more importantly, to take note of the ship's steady, inexorable progress. In short, metaphor interpretation is less a matter of avoiding certain inferences than it is a matter of selecting certain others. Any approach to metaphor that does only the first of these is not a way of interpreting metaphors, only of ignoring them. Under this view, the fundamental insight about metaphor is simply bizarre and inexplicable.²

²For further arguments against this approach to metaphor, see Nunberg [32].

Several more recent approaches can be seen as aiming toward the selection of an appropriate set of inferences. For Miller [33], the basic pattern of metaphor is given by the formula

$$(5) \quad G(x) \longrightarrow (\exists F) (\exists y) (\text{SIM}(F(x), G(y)))$$

In words, this means the following. A predicate, G , is applied metaphorically to an entity x . To interpret the metaphor, one must discover a property F which literally describes x , an entity y which G literally describes, and the similarity between $F(x)$ and $G(y)$. By similarity, Miller means that there are "features" which $F(x)$ and $G(y)$ share.³ The notion of "feature" is probably equivalent to or subsumed by the artificial intelligence notion of "inference." Thus for Miller interpreting a metaphor $G(x)$ is a matter of selecting the inferences which one can draw from G which can also be drawn from the known (literal) properties of x .

There have been a number of recent proposals that may be viewed as specifications, prior to interpretation, of which inferences are the best to select. One proposal is that of Ortony, who also uses the notion of "feature." Ortony has suggested a breakdown of the knowledge about the vehicle and the tenor into classification facts, other high-salience facts, and low-salience facts [35]. Classification facts are not transferred from the vehicle to the tenor. Thus, from "John is an elephant" we do not infer that John is a (nonhuman) animal. What gets transferred from the vehicle to the tenor are other high-salience facts whose correlates in the tenor are of low salience. It is a high-salience fact that elephants are large, whereas John's size is generally of low salience. The effect of the metaphor is to bring to the fore this low-salience fact about John. In terms of selective inferencing, one draws the high-salience inferences associated with the vehicle that are not contradicted or confirmed by high-salience inferences about the tenor.

Carbonell [36], working in an artificial intelligence framework, proposes pre-packaging the inferences associated with Lakoff and Johnson's root metaphors, recognizing on the basis of explicit content which "package" or root metaphor is being tapped into, and then drawing all the inferences in the package that are not explicitly contradicted by the text.

In view of the close relationship that is generally asserted between metaphor and analogy, the work in artificial intelligence that should be most relevant to a study of metaphor is research on analogical reasoning. There are a number of examples. Evans wrote a program for solving geometric analogy problems [37]. Kling built a system for proving theorems in ring theory by examining proofs of analogous theorems in group theory [38] (a class of analogies that forms the

³This is a weaker notion of similarity than Tversky's [34], which also takes into account features that are not shared.

basis of Galois theory) [compare 39]. Dershowitz and Manna [40] and Moll and Ulrich [41] attempt the automatic synthesis of programs by analogy with known programs. Most of this work has been conducted at too specific a level to be of use in our work on metaphor. Where the specific domain has been abstracted away from, e.g., in Kling [38] and J. McDermott [42], the framework has been too general to offer any new insights.

An exception to this is the work of Winston [43]. He presents an algorithm in which properties are transferred from the vehicle to the tenor if they are extremes on some scale, are known to be important, or serve to distinguish the vehicle from other members of its class. Thus, properties of elephants that are not shared by other animals would be transferred. Again, one can view the transfer of a property from the vehicle to the tenor as an inference one selects, and what Winston has suggested are criteria for selecting these inferences.

Gentner presents evidence that relations are more likely than attributes to be transferred from the vehicle to the tenor [44]. That is, inferences are more likely to be selected if they involve a two-place predicate rather than a one-place predicate in the consequent. Thus, from the simile "the atom is like a solar system" one is more likely to infer that electrons go around the nucleus (a two-place predication) than that the nucleus is yellow (or roughly spherical).⁴

Toward the end of the paper cited above, Carbonell suggests a more refined classification of possible inferences [36]. Inferences about goals and plans of agents and causal facts are most likely to be transferred from the vehicle to the tenor. Somewhat less likely are functional attributes, temporal orderings, and structural relations, and least likely, almost never relevant, are physical descriptive properties and object identity. It is not surprising that this should be the case, since the function of metaphor is usually to make sense of some abstract domain.

All of this research seeks to specify certain classes of inferences that are typically transferred—on the basis of salience, arity of the predicates, convention, semantic content of the inferences, and so on. But these approaches suffer from the fact that they do not explain how context influences the interpretation of metaphors. None takes into account the text in which the metaphor is embedded.⁵

The approach taken in this paper is to subsume the metaphor interpretation problem under the more general problem of making sense of a discourse as a whole. The discourse operations a natural language processor must possess anyway—operations like coherence relation recognition, predicate interpretation,

⁴There has been other work on metaphor by psychologists. A good review can be found in Ortony, Reynolds and Arter [45].

⁵If we imagine salience as something which varies with context, then Ortony's proposal can be viewed as depending on context, but it is a rather blunt sort of dependence. Carbonell's choice of the pre-packaged root metaphor is dependent on explicit context, so this step in his algorithm at least is context-dependent.

and compound nominal interpretation—will often serve to pick out the relevant inferences in cases of metaphor. Often the correct interpretation of the metaphor will simply “fall out” as a by-product of other interpretation processes.

TWO EXAMPLES

A Simple Metaphor

Let us now consider how a simple metaphor would be interpreted using context-dependent selective inferencing. Consider

(6) John is an elephant.

Let us suppose our initial logical representation for this is

elephant (J).

There are a number of things we might infer from the fact that some entity is an elephant. Among the axioms allowing such inferences would be

elephant(x) \longrightarrow large(x)
 elephant(x) \longrightarrow has-trunk(x)
 elephant(x) \longrightarrow good-memory(x)
 elephant(x) \longrightarrow thick-skinned(x)
 elephant(x) \longrightarrow clumsy(x)

The problem we are faced with in interpreting (6) is the problem we are always faced with in interpreting a text—determining which inferences it is appropriate to draw from what we have been told. Depending on the situation, we may want to infer “large(J)” or “good-memory(J).” The inference that John has a trunk is presumably rejected because of strong reasons to believe the contrary.

Which inferences *are* appropriate will depend on context. Example (6) contains insufficient context to allow precise interpretation. But we can embed it in a text in which discourse operations become decisive. For example, in

(7) Mary is graceful, but John is an elephant.

coherence considerations force the interpretation. In order to recognize the contrast coherence relation (see [46]) indicated by “but,” we must draw the inferences that John is clumsy, and thus not graceful. Other possible inferences about elephants are not drawn, not so much because they would result in an inconsistency, but because no discourse problem requires them to be drawn. Other texts would force other inferences. Consider

Patricia is small, but James is an elephant.
 Susan forgets everything, but Paul is an elephant.
 Jenifer is subtle, but Roger is an elephant.

The inferences associated with the explicit predication in the metaphor (7) are of three classes. There are those inferences that are definitely intended—for example, the inference “clumsy(J).” These “ground,” or establish a firm basis, for the metaphor; they are what warrant it. Then there are those inferences that are definitely not intended and are inappropriate to draw, the disparities, such as “has-trunk(J).” Finally, there are inferences that lie in-between, such as “large(J),” which may or may not be intended by the speaker and may or may not occur to the listener. Much of the power of a metaphor derives from this third class of inferences—the other things that are suggested by the metaphor beyond its ground or firm basis. In fact, even the inappropriate inferences of the second class lend power to the metaphor, since the very denial of something suggests its possibility. The calling up and rejection of the image of an elephant in interpreting (7) may leave its trace.

A Spatial Metaphor Schema

Metaphors that tap into our spatial knowledge are especially powerful since our knowledge of spatial relationships is so extensive, so rich, and so heavily used. As soon as the basis for the spatial metaphor is established, then in our thinking about a new domain we can begin to borrow the extensive machinery we have for reasoning about spatial relationships. For example, once I say that

(8) The variable N is at zero,

and interpret it as

(9) The value of the variable N is equal to zero,

then I have tapped into a large network of other possible uses. I can now say

N goes from 1 to 100

to mean

The value of N successively equals integers from 1 to 100.

I can say

N approaches 100

to mean

The difference between 100 and the value of N becomes smaller.

N can now *stay at* a number, *move from* one number to another *through* several others, be *between* two numbers, be *here*, be *there*. Variables can be *scattered along* an interval, they can *follow* one another *along* the number scale, they can be *switched*. In short, by means of the simple identification of (8) and (9) we have brought into the whole complex of spatial terminology.

In terms of a system for selective inferencing, what we mean when we say that our spatial terminology is an intricate network is that there are a great many axioms that relate the various spatial predicates. The concept of location—the predicate “at”—is at the heart of this network because so many of the axioms refer to it. For example, associated with the predicate “go” we might have an axiom like

$$\text{go}(x, y, z) \ \& \ \text{at}'(w_1, x, y) \ \& \ \text{at}'(w_2, x, z) \ \longrightarrow \ \text{change}(w_1, w_2)$$

that is, if x goes from y to z and w₁ is the condition of x being at y and w₂ is the condition of x being at z, then there is a change of state from w₁ to w₂. Similarly, part of the meaning of “switch” could be encoded in the axiom

$$\text{switch}(x, y_1, y_2) \ \& \ \text{at}'(w_{11}, y_1, z_1) \ \& \ \text{at}'(w_{12}, y_1, z_2) \ \& \ \text{at}'(w_{21}, y_2, z_1) \ \& \ \text{at}'(w_{22}, y_2, z_2) \ \longrightarrow \ \text{change}(w_{11}, w_{12}) \ \& \ \text{change}(w_{22}, w_{21})$$

That is, if x switches y₁ and y₂ and w_{ij} is the condition of y_i being at z_j, then there is a change from condition w₁₁ to condition w₁₂ and a change from condition w₂₂ to condition w₂₁.

We were able to establish the metaphor “a variable as an entity at a location” simply by identifying (8) and (9). In our formalism we can establish the metaphor with similar simplicity by proposing the following axiom:

$$(10) \ \text{variable}(x) \ \& \ \text{value}'(w, y, x) \ \longrightarrow \ \text{at}'(w, x, y)$$

That is, if x is a variable and w is the condition of y being its value, then w is also the condition of x being at y.

Axiom (10), identifying “is the value of” with “is at,” gives us entry into an entire metaphor schema and enables us to transfer to one domain the structure of another, more thoroughly understood domain.

A discourse operation, which in Hobbs [23] was called *predicate interpretation*, uses axioms like (10) to arrive at interpretations of certain metaphorical expressions. The idea behind it is that most utterances make very *general* or ambiguous sorts of predications and that part of the job of comprehension is

to determine the very *specific* or unambiguous meaning that was intended. Thus, someone might make the general statement

I went to London,

expecting us to be able to interpret it as

I flew to London in an airplane,

rather than interpreting the going as swimming, sailing, walking, or any of the myriad other manners of going. In the case of (8), we are expected to determine which of the many ways one thing can be *at* another is intended in this particular case. That is, rather than determining what we can infer from what is said, we try to determine what the speaker had in mind from which he inferred what he said. In terms of our notation, suppose G is a general proposition and S a specific one and

$$S \longrightarrow G$$

is an axiom expressing a fact that a speaker and a listener mutually know. The speaker utters G in the expectation that the listener will interpret it as S. The listener must locate and use the axiom to determine the specific interpretation.

In this manner, axiom (10) provides one possible interpretation of (8), in that it specifies one of the many ways in which one thing can be *at* another, which the speaker may have meant. When a metaphorical use of “go” or “switch” or any of the other spatial predicates is encountered, axiom (10) combines with the axioms defining the spatial predicate in terms of “at” to give us the correct interpretation.

An alternative to this approach might seem to be to infer intended meaning from what was said. We would use axioms not of the form “S \longrightarrow G” but of the form

$$G \ \& \ C_1 \ \& \ \dots \ \& \ C_n \ \longrightarrow \ M$$

where G is the general proposition that is explicitly conveyed, the C_i's are conditions determinable from context, and M is the intended meaning. For interpreting (8), this would require an axiom like

$$(11) \ \text{at}'(w, x, y) \ \& \ \text{variable}(x) \ \longrightarrow \ \text{value}'(w, y, x),$$

that is, if w is the condition of x being at y and x is a variable, then w is also the condition of y being the value of x. To interpret (8) we would search through all axioms for axioms that, like (11), have “at” in the antecedent,

check whether the other conjuncts in the antecedent were true, and if so, conclude that the axiom's consequent was the intended meaning. This would be equivalent to a "discrimination-net" approach to word-sense disambiguation [e.g., 47], in which one travels down a tree-like structure, branching one way or the other according to whether some condition holds, until arriving at a unique specific interpretation at the bottom. The difficulty with this approach is that it supposes we could anticipate at the outset all the ways the meaning of a word could be influenced by context. For metaphors we would have to be able to decide beforehand on all the precise conditions leading to each interpretation. It is highly implausible that we could do this for familiar metaphors, and for novel metaphors the whole approach collapses.

As always, there are a number of inferences involving "at" that we would not want to draw in the case of (8). For example, in the blocks world, if BLOCK1 is at location (2,3,0), then it is impossible for BLOCK2 to be at (2,3,0) at the same time. Yet there is no difficulty whatever in two variables being "at" the same value.⁶ Similarly, if a block is at a location, it is probably being held there by friction and gravity. But with variables there is no need to concern ourselves with what holds them at their values. It is probably the case in general that facts of a "topological" character lend themselves to spatial metaphors, and facts of a "physical" character do not.

ACKNOWLEDGMENTS

I have profited from discussions with Armar Archbold, Jaime Carbonell, Gary Hendrix, George Lakoff, and Bob Moore, and also with the participants of the Artificial Intelligence Workshop at the Electrotechnical Laboratory, Tokyo, Japan, in August, 1979. None of these people are responsible for any errors in this paper.

REFERENCES

1. I. A. Richards, *The Philosophy of Rhetoric*, Oxford University Press, Oxford, 1936.
2. D. Knuth, *The Art of Computer Programming*, Vol. 1, Addison-Wesley, Reading, Massachusetts, 1973.
3. G. Lakoff and M. Johnson, *Metaphors We Live By*, University of Chicago Press, Chicago, 1980.
4. G. Vico, *The New Science of Giambattista Vico*, T. Bergin and M. Frisch (trans.), Cornell University Press, Ithaca, 1968. (Orig. published, 1744.)

⁶Even in our casual talk about physical reality, the inference is highly dependent on specific circumstances. We are quite comfortable saying that John and Bill are both *at* the post office.

5. C. Ogden, *Bentham's Theory of Fictions*, Routledge & Kegan Paul, London, 1932.
6. S. Langer, *Philosophy in a New Key: A Study in the Symbolism of Reason, Rite, and Art*, Harvard University Press, Cambridge, 1942.
7. O. Jespersen, *Language: Its Nature, Development and Origin*, George Allen & Unwin, London, 1922.
8. B. Whorf, The Relation of Habitual Thought and Behavior to Language, J. Carroll (ed.), *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, Wiley, New York, 1956.
9. W. Urban, *Language and Reality*, G. Allen and Unwin, London, 1939.
10. H. Clark, Space, Time, Semantics, and the Child. T. Moore (ed.), *Cognitive Development and the Acquisition of Language*, Academic Press, New York, 1973.
11. J. Hobbs, A Model for Natural Language Semantics, Part I: The Model, Research Report 36, Department of Computer Science, Yale University, October, 1974.
12. J. Hobbs, A Computational Approach to Discourse Analysis, Research Report 76-2, Department of Computer Sciences, City College, City University of New York, December, 1976.
13. R. Jackendoff, Toward an Explanatory Semantic Representation, *Linguistic Inquiry*, 7, pp. 89-150, 1976.
14. A. Joshi and S. Rosenschein, Some Problems of Inferencing: Relation of Inferencing to Decomposition of Predicates, *Statistical Methods in Linguistics*, 1:1, 1976.
15. B. Grosz, The Representation and Use of Focus in Dialogue Understanding, Stanford Research Institute Technical Note 151, Stanford Research Institute, Menlo Park, Calif., July, 1977.
16. R. Schank and R. Abelson, *Scripts, Plans, Goals, and Understanding*, Lawrence Erlbaum Associates, Hillsdale, N.J., 1977.
17. A. Robinson, Investigating the Process of Natural-Language Communication: A Status Report, SRI Technical Note 165. SRI International, Menlo Park, Calif., July, 1978.
18. A. Collins and M. R. Quillian, How to Make a Language User, Tulving and Donaldson (eds.), *Organization of Memory*, Academic Press, New York, 1971.
19. D. McDermott and J. Doyle, Non-Monotonic Logic I, *Artificial Intelligence*, 13, pp. 41-72, 1980.
20. R. Reiter, A Logic for Default Reasoning, *Artificial Intelligence*, 13, pp. 81-132, 1980.
21. W. Mann, J. Moore, and J. Levin, A Comprehension Model for Human Dialogue, *Proceedings, International Joint Conference on Artificial Intelligence*, Cambridge, Massachusetts, August, 1977.
22. J. Hobbs, Interpreting Natural Language Discourse, *Final Report, National Science Foundation Research Grant No. MCS 78-07121*, July, 1980.
23. J. Hobbs, Coherence and Interpretation in English Texts, *Proceedings, International Joint Conference on Artificial Intelligence*, Cambridge, Massachusetts, August, 1977.

24. J. Hobbs, Coherence and Coreference, *Cognitive Science*, 3, pp. 67-90, 1979.
25. J. Hobbs, Selective Inferencing, *Proceedings, Third National Conference of Canadian Society for Computational Studies of Intelligence*, Victoria, British Columbia, May, 1980.
26. M. Black, More About Metaphor, A. Ortony (ed.), *Metaphor and Thought*, Cambridge University Press, Cambridge, 1979.
27. Aristotle, "Art" of Rhetoric, J. Freese (trans.), *Aristotle: "Art" of Rhetoric*. Harvard University Press, Cambridge, 1975.
28. S. W. Russell, Computer Understanding of Metaphorically Used Verbs, *American Journal of Computational Linguistics*, microfiche 44, 1976.
29. S. Levin, *The Semantics of Metaphor*, The Johns Hopkins University Press, Baltimore, 1977.
30. R. Matthews, Concerning a 'Linguistic Theory' of Metaphor, *Foundations of Language*, 7, pp. 413-425, 1971.
31. E. Kahn, Frame Semantics for Motion Verbs with Application to Metaphor, C. Cogen et al. (eds.), *Proceedings of the First Annual Meeting of the Berkeley Linguistic Society*, 1975.
32. G. Nunberg, *The Pragmatics of Reference*, Ph.D. Thesis, City University of New York, 1978.
33. G. Miller, Images and Models, Similes and Metaphors, A. Ortony (ed.), *Metaphor and Thought*, Cambridge University Press, Cambridge, 1979.
34. A. Tversky, Features of Similarity, *Psychological Review*, 84, pp. 327-352, 1977.
35. A. Ortony, The Role of Similarity in Similes and Metaphors, A. Ortony (ed.), *Metaphor and Thought*, Cambridge University Press, Cambridge, 1979.
36. J. Carbonell, Metaphor: An Inescapable Phenomenon in Natural-Language Comprehension, W. Lehnert and M. Ringle (eds.), *Strategies for Natural Language Processing*, Lawrence Erlbaum Associates, Hillsdale, N.J., 1982.
37. T. Evans, A Program for the Solution of a Class of Geometric-Analogy Intelligence-Test Questions, M. Minsky (ed.), *Semantic Information Processing*, The MIT Press, Cambridge, 1968.
38. R. Kling, A Paradigm for Reasoning by Analogy, *Artificial Intelligence*, 2, pp. 147-178, 1971.
39. E. Artin, *Galois Theory*, Notre Dame Mathematical Lectures, No. 2, Notre Dame, Indiana, 1959.
40. N. Dershowitz and Z. Manna, The Evolution of Programs: A System for Automatic Program Modification, *IEEE Transactions on Software Engineering*, Vol. SE-3, No. 4, 1977.
41. R. Moll and J. Ulrich, The Synthesis of Programs by Analogy, *Proceedings, International Joint Conference on Artificial Intelligence*, Tokyo, August, 1979.
42. J. McDermott, Learning to Use Analogies, *Proceedings, International Joint Conference on Artificial Intelligence*, Tokyo, August, 1979.
43. P. Winston, Learning by Creating and Justifying Transfer Frames, *Artificial Intelligence*, 10, pp. 147-172, 1978.
44. D. Gentner, Studies of Metaphor and Complex Analogies, Paper presented at the A.P.A. Symposium on Metaphor as Process, Montreal, September, 1980.
45. A. Ortony, R. Reynolds, and J. Arter, Metaphor: Theoretical and Empirical Research, *Psychological Bulletin*, 85, pp. 919-943, 1978.
46. J. Hobbs, Why is Discourse Coherent? SRI Technical Note 176, SRI International, Menlo Park, Calif., November, 1978.
47. C. Reiger, Viewing Parsing as Word Sense Discrimination, W. Dingwall (ed.), *A Survey of Linguistic Science*, Greylock, London, 1978.

Direct reprint requests to:

Jerry R. Hobbs
SRI International
Menlo Park, CA

METAPHOR INTERPRETATION AS SELECTIVE
INFERENCING: COGNITIVE PROCESSES IN
UNDERSTANDING METAPHOR* (PART 2)**

JERRY R. HOBBS
SRI International

ABSTRACT

In Part 1, the importance of spatial and other metaphors was demonstrated. An approach to interpreting metaphor in a computational framework was described, based on the idea of selective inferencing, in which a processor draws or refrains from drawing certain inferences in a controlled fashion. Two examples of metaphors were examined in detail in this light—a simple metaphor and a spatial metaphor schema. In Part 2 a novel metaphor is examined and there is a discussion, from this perspective, of some classic issues concerning metaphor, including the analogical processes that underlie metaphor, the stages in the life of a metaphor, and the definition of metaphor.

In Part 1, the importance of spatial and other metaphors was demonstrated. An approach to interpreting metaphor in a computational framework was described, based on the idea of selective inferencing, in which a natural language processing system draws or refrains from drawing certain inferences in a controlled fashion. Two examples of metaphor were examined in detail in this light. The first was a simple metaphor:

(6) John is an elephant.

Among the inferences one could draw, such as “has trunk,” “large,” “has good memory,” and “clumsy,” only the last is drawn in the context “Mary is graceful, but. . .” The second example was a spatial metaphor schema from computer science, the use of “the variable N is at zero” to mean that “the

*This work was supported by the National Science Foundation under Grant No. MCS-78-07121, by the Advanced Research Projects Agency under Contract No. N00039-79-C-0118 with the Naval Electronics Systems Command, and by the National Library of Medicine under Grant No. 2R01 LM03611-03.

**Part 1 of this article appeared in *Empirical Studies of the Arts*, Vol. 1(1).

value of the variable N is equal to zero." The identification was established by means of the axiom

$$(10) \text{ variable}(x) \ \& \ \text{value}'(w,y,x) \longrightarrow \text{at}'(w,x,y)$$

That is, if w is the condition of y being the value of a variable x , then w is also the condition of x being at y . Sentences in which variables go, approach something, or are switched can then be interpreted using axiom (10) and axioms decomposing "go," "approach" and "switch" in terms of "at."

In Part 2 a novel metaphor is examined and there is a discussion, from this perspective, of some classic issues concerning metaphor, including the analogical processes that underlie metaphor, the stages in the life of a metaphor, and the definition of metaphor.

A NOVEL METAPHOR

Our final example of metaphor illustrates how we can represent a metaphor that depends on an elaborate analogy between two complex processes. The metaphor comes from a *Newsweek* article (July 7, 1975) about Gerald Ford's vetoes of bills Congress had passed. A Democratic congressman complains:

(12) We insist on serving up these veto pitches that come over the plate the size of a pumpkin.

It is clear from the rest of the article in which this appears that this means that Congress has been passing bills that the President can easily veto without political damage. There are a number of problems raised by this example, but the only ones we will address are the questions of how to represent and interpret "veto pitches that come over the plate."

The analogy here is between Congress sending a bill to the President to sign or veto and a pitcher throwing a baseball past a batter to miss or hit. Let us encode each of the processes first and establish the links between them, and then show how a natural language processing system might discover them.

A remark about notation is necessary first, however. It will be convenient to represent a sentence like "Congress sends the bill to the President" not in the most obvious way as "send(C,B,P)," but as a statement about the existence of a condition or action SD, which is the sending by Congress of the bill to the President (compare [1]). We will represent this by

$$\text{send}'(\text{SD},\text{C},\text{B},\text{P}).$$

The single quote may be thought of as a nominalization operator turning the sentence "Congress sends the bill to the President" into the corresponding noun

phrase "the sending by Congress of the bill to the President." There are two reasons for using this notational convention: it allows us to express certain higher predications in the schemata, and it allows us to express the mapping between the schemata with greater precision. (The notation is also used in the second example, but there I thought I could slip it past the reader.)

The facts about a bill are as follows: The participants are Congress, the bill, and the President. Congress sends a bill to the President, who then either signs it or vetoes it. We will assume there is an entity C, Congress. To encode the fact that C is Congress, again we could write simply

$$\text{Congress}(\text{C}).$$

But here also it will prove more useful to assume there is a condition, call it CC, which is the condition of C being Congress. We will represent this

$$\text{Congress}'(\text{CC},\text{C}).$$

CC is thus the entity referred to by the noun phrase "being Congress." Similarly, there are entities B, CB, P, and CP, with the properties

$$\text{bill}'(\text{CB},\text{B});$$

i.e., CB is the condition of B being a bill; and

$$\text{President}'(\text{CP},\text{P}),$$

i.e., CP is the condition of P being the President. There are three relevant actions, call them SD, SG, and VT, with the following properties:

$$\text{send}'(\text{SD},\text{C},\text{B},\text{P});$$

i.e., SD is the action by Congress C of sending the bill B to the President P;

$$\text{sign}'(\text{SG},\text{P},\text{B}),$$

i.e., SG is the action by the President P of signing the bill B; and

$$\text{veto}'(\text{VT},\text{P},\text{B}),$$

i.e., VT is the action by the President P of vetoing the bill B. There is the condition—call it OSV—in which either the signing SG takes place or the vetoing VT takes place:

$$\text{or}'(\text{OSV},\text{SG},\text{VT}).$$

Finally, there is the situation or condition, TH, of the sending SD happening followed by the alternative actions OSV:

then'(TH,SD,OSV).

The corresponding facts about baseball are as follows:¹ There are a pitcher x, a ball y, and a batter z, and there are the conditions cx, cy, and cz, of x, y, and z being what they are:

pitcher'(cx,x)
ball'(cy,y)
batter'(cz,z).

The actions are the pitching p by the pitcher x of the ball y to the batter z,

pitch'(p,x,y,z);

the missing m of the ball y by the batter z,

miss'(m,z,y);

and the hitting h of y by z,

hit'(h,z,y).

Let omh represent the condition of one or the other of m and h occurring,

or'(omh,m,h),

and th the situation of the pitching p followed by either m or h,

then'(th,p,omh).

The linkage established by the metaphor is among other things, between the bill and the ball. But it is not enough to say that B, in addition to being the bill, is also in some sense a ball, just as B has other properties, say, being concerned with federal housing loans, being printed on paper, and containing seventeen subsections. The metaphor is stronger. What the metaphor tells us is that the

¹Where individual constants, C, CC, B, . . . , were used in the Congressional bill schema, universally quantified variables, x, cx, y, . . . , are used here. This is because the baseball schema is general knowledge that will be applied to the specific situation involving Congress and the President. It is a collection of axioms that get instantiated in the course of interpreting the metaphor.

condition of B being the bill is indeed the condition of B being a ball. Similar links are established among the other participants, actions, and situations. That is, the baseball schema is instantiated with the entities of the Congressional bill schema, leading to the following set of propositions:

- | | |
|-----------------------|-------------------|
| (13) Congress' (CC,C) | pitcher' (CC,C) |
| bill' (CB,B) | ball' (CB,B) |
| President' (CP,P) | batter' (CP,P) |
| send' (SD,C,B,P) | pitch' (SD,C,B,P) |
| sign' (SG,P,B) | miss' (SG,P,B) |
| veto' (VT,P,B) | Hit' (VT,P,B) |
| | or' (OSV,SG,VT) |
| | then' (TH,SD,OSV) |

The two schemata and their links are shown more graphically in Figure 1.

Although all of this has been described in terms of schemata, a schema in this framework is simply a collection of possibly very complex axioms that are interrelated by the co-occurrence of some of the same predicates, perhaps together with some meta-knowledge for controlling the use of the axioms in inferencing. The linkage between the two schemata does not require some special "schema-mapping" operation, but only the assumption of identity

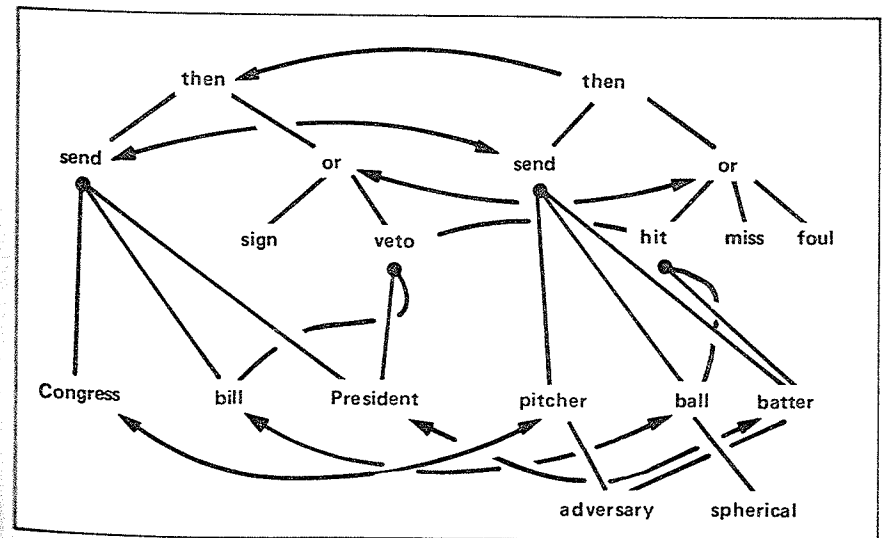


Figure 1. Mapping from Baseball Schema to Congress Schema.

between the corresponding conditions, just as in the second example we identified “is the value of” with “is at.” The difference between a conventional metaphor and a novel metaphor is that in the case of the former the identity is encoded in an axiom like (10), whereas in the latter the identity must be drawn as an implicature. Thus, to represent the metaphor, we do not have to extend our formalism beyond what was required for the first two examples, nor indeed beyond what is required for nonmetaphorical discourse.

However, a shortcoming of this representation, as it stands, is that there is no explicit separation of the two parts of the metaphor. Thus, C is both Congress and a pitcher and P is both the President and a batter. But there is no explicit indication that the properties “Congress” and “President” belong to one side of the metaphor and “pitcher” and “batter” to the other. We could remedy this by being more careful about the difference between a condition and a description of the condition. For then we could say that the condition CC of C being Congress is identical to the condition of C being a pitcher, while the descriptions involving “Congress” and “pitcher” are distinct. We would then make assertions about the descriptions that they belong to one domain or the other. But the details of this hastily sketched idea cannot be worked out here.

No natural language processing system existing today could derive (13) from (12). Nevertheless, we can make a reasonable guess as to the basic outline of a solution. The congressman said, “We insist on serving up these veto pitches. . . .” For someone to serve up a pitch is for him to pitch. This leads to the identification of Congress with the pitcher. To interpret the compound nominal “veto pitch,” we must find the most salient, plausible relation between a veto and a pitch. From our knowledge about vetoes, we know that Congress must first *send* the bill to the President. From our knowledge about pitching, we know that for the Congress/pitcher to pitch, it must *send* a “ball” to a “batter.” We have a match on the predicate “send” and on the agents of the sendings, Congress. We can complete this match by assuming, or drawing as an implicature, that the bill is the ball and the President is the batter.²

We have almost a complete match between the two situations. The analogy will be completed when we determine which of the various possible actions that a batter can perform corresponds to the President’s veto. But this is just what we need to complete the relation between “veto” and “pitch” in the compound nominal. By some means well beyond the scope of this paper to discuss, “pitches that come over the plate the size of a pumpkin” must be interpreted to mean that the ball is easy for the batter to hit. If we assume maximum redundancy—that a veto pitch and a pitch that comes over the plate the size of a pumpkin are roughly the same thing—then we assume that the pitch is a bill/ball that the

Congress/pitcher sends to the President/batter which he then finds easy to veto/hit. The analogy is complete.

As with all metaphorical expressions, as indeed with any expression, there will be a number of inferences that should not be drawn in this case—for example, that B is spherical and has stitching. But this metaphor invokes other inferences that we do accept, inferences that would not necessarily follow from the facts about the American government. It suggests, for example, that Congress and the President are adversaries in the same way that a pitcher and a batter are, and that from the President’s perspective it is good for him to veto a bill Congress had passed and bad for him to sign it. What we know about the adversary relationship in baseball is vivid and unambiguous, and herein lies the power of the metaphor.

This example involves the identification of two highly structured portions of our knowledge base. It raises a question of whether our approach can handle metaphors in which one domain has much less structure, especially metaphors which impart structure to a domain that it would not otherwise have. Lakoff and Johnson demonstrate this effect by inventing a “love as a collaborative work of art” metaphor and showing some of the things that can be concluded about love as a result [3]. I see no fresh difficulties that this would cause for my approach. Corresponding to the numerous basic links between the existing Congressional bill and baseball schemata, there would be only a few links between our knowledge of love and of collaborative works of art. If this new metaphor is productive, then corresponding to the suggestion from baseball of an adversary relationship in government, there will be numerous suggestions from the nature of collaborative works of art about the nature of love. Therefore, the effect of the new metaphor may be quite different from the effect of the ones we have examined, but the mechanisms involved in interpreting it are the same.

SOME CLASSICAL ISSUES

Metaphor and Analogy

In all three examples, we have seen the same broad processes at work. They can be summarized as follows: there are two domains, which we may call the *new domain*, or the domain which we are seeking to understand or explicate, and the *old domain*, or the domain in terms of which we are trying to understand the new domain and which provides the metaphor. These are Richards’ tenor and vehicle, respectively [4]. In our examples, the new domains are John’s nature, computer science, and the workings of the American government. The old domains are an elephant’s nature, spatial relationships, and baseball. For each old domain, we can distinguish between what may be called the *basic concepts and relationships* and *complex concepts and relationships*. For spatial

²Such assumptions are common in interpreting discourse. In fact, they constitute one of the principal mechanisms for resolving pronouns and implicit arguments [2].

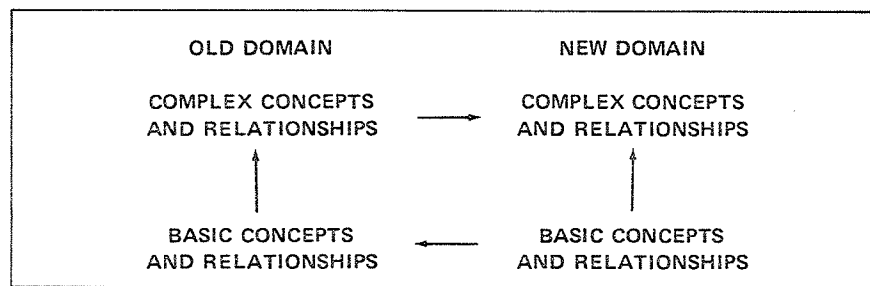


Figure 2. Analogical Processes Underlying Metaphor.

relationships, “at” is a basic concept; “go,” “approach,” and “switch” are complex concepts. For baseball, “pitcher” and “batter” are basic, their adversary relationship is complex. In the elephant metaphor, “elephant” is basic, “has-good-memory,” “clumsy” and “large” are complex. What is basic and what is complex in a particular domain are not necessarily fixed beforehand, but may be determined in part by the metaphor itself.

Each of the examples can be viewed as setting up a link between the basic concepts of a new domain and an old domain, in order that complex concepts or relationships will carry over from the old to the new. Figure 2 illustrates this.

Figure 2 is familiar from Galois theory, algebraic topology, and category theory [5-7]. One can prove theorems in one domain—for example, the category of fields—by constructing a “functor” to map its objects and relations into the objects and relations of another domain—for example, the category of groups—proving the theorem in the second domain, and using the inverse functor to map it back into the original domain.

The diagram illustrates a general paradigm for analogical reasoning. To reason in a new domain about which we may know little, we map it into an old domain, do the reasoning in the old domain, and map the results back into the new domain. To make use of this paradigm, in our framework, for understanding the process of metaphor, we have had to specify the nature of the links in the diagram. The horizontal links are realized by means of explicit statements like (6), or by axioms like (10) in the case of frozen metaphors, or by means of implicatures like (13) in the case of novel metaphors. The vertical links in the diagram are realized by the collections of axioms encoding the relationships between basic and complex concepts.³

But there is a problem. In category theory, once the functor maps the new domain into the old domain, then everything we can conclude in the old domain must carry over to the new. However, in most kinds of analogical reasoning and

³It is of course also important to specify what we mean by “domain.” This issue is addressed below.

in interpreting metaphors, only a subset of what can be concluded in the old domain will carry over to the new. The major problem for us, then, is how to determine precisely what from the old domain *does* carry over to the new. Let us elaborate on this.

There are three kinds of inferences in the old domain that must be distinguished in interpreting the metaphor.

1. The *grounds* of the metaphor, or the inferences that must be drawn if one is to make sense of the metaphor. These are what warrant the metaphor. In our first example, the grounds may be the inference that John is clumsy; in the third example, that the bill/ball is sent to the President/batter.

Black suggests a classification of theories of metaphor that includes “substitution theories,” [8] in which a metaphor is analyzed by replacing the explicit predication with those literal propositions it is intended to convey.⁴ In our terms, it is the ground inferences that such theorists want to substitute for the metaphor.

2. *Disparities*, or the inferences that should not be drawn, whether because they are contradictory or irrelevant. In our examples, a disparity between John and an elephant that an elephant has a trunk, between the bill and a ball that a ball is spherical.

Richards points out that the disparities frequently play an important role: a significant effect of a metaphor may be the recognition that some of the criterial inferences that could be drawn from the explicit predication are not appropriate. The fact that John, though an elephant, is not a large animal, but a person, carries the implication that he should resemble a large animal even less. Ong suggests that a metaphor is effective only as long as it calls these disparities to mind [11]. “John is an elephant” strikes us in a way that “the foot of a mountain” does not.

In our approach, certain disparities are considered and actively denied, rejected when inconsistency is discovered. This active process may be compared with a cartoon in which John gradually acquires bulk, a trunk and four stocky legs while crashing along clumsily, then returns to his normal appearance. This has the flavor of a “reverse substitution” theory of metaphor, in which the *inappropriate* properties inferable from the explicit predication, for a moment, replace the metaphor.

3. *Suggestions*, a weak term for one of metaphor’s greatest powers, its suggestiveness. These are the inferences that may or may not be drawn. They are not required to interpret the metaphor, nor are they obviously inappropriate. In our first example, a suggestion is that John is large; in the third it is suggested that the President and Congress are adversaries. There are positive and negative aspects to this suggestiveness. On the positive side, it is this more than anything

⁴Beardsley refers to this as the “literalist” theory [9] and the “comparison” theory [10].

else that makes metaphor such a powerful conceptual tool. We are able to draw conclusions that we could not have anticipated.

On the other hand, Lakoff and Johnson point out the dangers of mistaking the metaphor for a true description, and thus drawing too many suggested inferences without adequately examining their appropriateness [3]. One is blinded to the limits of the metaphor, and also to alternative metaphors. Reddy discusses a specific case [12], the language-as-conduit metaphor and its influence on the study of communication; the theme is also developed at length by Turbayne with respect to metaphors of science [13].

The problem of interpreting a metaphor is to determine for the various possible inferences, into which of the three classes they fall. It is the principal thesis of this paper that much of the solution to this problem will come from knowledge-based interpretation processes that are already required for non-metaphorical discourse.

This position is in contrast with a commonly proposed account of metaphor interpretation. In this account, one first tries the literal interpretation, and then if that fails semantic constraints, one interprets the expression as a metaphor. That is, a separate initial step is postulated in which something is found to be wrong. There are several problems with this account. First of all, literal interpretation may not fail. Consider the following two statements

People are not cattle.

Whales are not fish.

Both statements are literally true biological facts. But suppose we encounter the first sentence in a political speech arguing that people cannot be herded around without consideration for their individual needs. Then it is to be interpreted as a metaphor, or if it is not a metaphor, at least it is the denial of a metaphor, and all the same interpretation processes must be called into play. Morgan gives further examples of metaphors that are or could be literally true [14].

A second difficulty is that all failures of literal interpretation are not due to metaphor. More often they result from metonymy, or indirect reference. For example, in

This restaurant accepts American Express,

we are not using "accept" metaphorically as a special kind of relation between small businesses and large corporations. Rather we are using "American Express" metonymically to refer to credit cards issued by American Express. An interesting intermediate case is

America believes in democracy.

Are we viewing America metaphorically as something which can believe, or are we using it metonymically to refer to the typical inhabitant, or the majority of inhabitants, of America?

However, the principal difficulty is that this position underestimates the task of arriving at a literal interpretation of an expression. A striking example is a clause that appeared in a paper by Wallace Chafe [15]:

Back when we were fish, . . .

The intent is that this be interpreted literally, where "we" is taken to refer to all people and their ancestors indefinitely far back. But to arrive at this interpretation we have to access what we know about evolution.

An excellent example of the difficulties in interpreting literal expressions is provided by what Black calls the "comparison" view of metaphor [8]. A metaphor is seen as an elliptical form of a simile. Thus, the metaphorical "John is an elephant" translated into the literal "John is like an elephant" or "John is like the stereotypical elephant in certain respects." But the predicate "like" is a very good example of a literal expression whose interpretation is quite problematic. Part of the literal meaning of "A is like B" is that A shares certain properties with B. Thus, in understanding "His house is like my house," we need to determine in which respects the two are alike. Similarly, in interpreting "John is like an elephant," we must discover in just what respects John is like an elephant. But this means that the problem of interpreting the literal "like" is isomorphic to the problem of interpreting the original metaphor.⁵

There is generally a large overlap in the processes of literal interpretation and metaphor interpretation, as this paper has argued and illustrated. Other writers have made or failed to make this point. Searle discusses at length the difficulties of interpreting literal utterances, but nevertheless separates these processes from the process of interpreting the utterance once the deviance is found, overlooking their likely identity [16]. Rumelhart, by contrast, shows that literal interpretation is sometimes problematic, as a way of arguing for the identity of these processes [17]. Nunberg also argues for the identity [18].

Perhaps the most detailed argument is that of Miller [19]. He shows how the interpretation of a sentence with the verb "to be" is problematic. Even if such a sentence is used literally, we have to determine at least whether it conveys entailment, as in "Trees are plants," or attribution, as in "This tree is a landmark." This can be characterized by saying that in Miller's formula (5),

$$(5) \quad G(x) \rightarrow (\exists F)(\exists y)(SIM(F(x), G(y)))$$

⁵Except of course identity is not assumed between the tenor and the vehicle. This is the standard observation about the difference between metaphor and simile.

in place of "SIM," there would be the relation "ENTAIL" or "ATTRIBUTE." Thus the general problem, Miller argues, is to determine which of these relations R is appropriate. That is, he proposes an interpretation process in which the first step is to determine R, and then, depending on what R is, the relevant inferences are drawn.

There are two difficulties with Miller's approach. First, he does not specify how R would be determined, at least at a level of detail that would satisfy a computational linguist. It is likely that whatever processes determine that similarity is intended simultaneously determine what the similarity is. In the approach I have been presenting, the mechanisms of selective inferencing first determine what inferences should be drawn, and then it may or may not be determined what relation R best characterizes this set of inferences.

The second difficulty with Miller's approach is that it seems to imply that there is always an explicit recognition that a metaphor is being used—whenever $R = \text{SIM}$. Most examples of metaphors are not explicitly recognized as such. The reader can test this for himself: the previous paragraph depends on at least four metaphors. We have seen in this paper that frequently the discourse operations result in a metaphor being interpreted, and that the operations themselves do not depend on the metaphor-nonmetaphor distinction. They are just the ordinary processes of deciding which inferences to draw and which to refrain from drawing.

This is not to say, however, that metaphors are never recognized. In many cases their recognition is just part of our general awareness of discourse, like the recognition that the speaker has used a French word, an uncommon syntactic construction, a particularly apt expression, or whatever. In other cases, the recognition might contribute to the interpretation of the sentence. For example, if someone tells me

John is a clock,

I may have to recognize explicitly that a metaphor is being used before I can get any interpretation at all. From a more computational point of view, it may be that once the grounds of a metaphor are discovered, knowledge that it *is* a metaphor often plays a role in directing further inferencing. But metaphor recognition is by no means a computationally necessary part of metaphor interpretation. It is an inference about the speaker, not the spoken.

However, not all metaphors are interpreted alike. There are various processes that might be invoked, and there are several degrees of awareness that a metaphor is being used. We can clarify this issue by using the picture of metaphor and analogy presented in Figure 2 to tell the life story of a metaphor. This will also throw light on another classical issue concerning metaphor: what should we count as a metaphor—is metaphor ornament or omnipresent?

The life story of a metaphor has four stages.

Think of a novel metaphor as a complex term from the old domain used in a context that requires a concept from the new domain. To interpret it we must decompose the complex term into basic concepts in the old domain, and either use available links between new and old basic concepts or surmise such links for the first time. This enables us to project the complex concept from the old to the new domain. For novel metaphors, we might expect this to require quite a bit of computing, and involve following a number of false leads.

The second stage is when the metaphor has become "familiar." The same path is followed in interpreting it, but now the salience of the required inferences is such that the computation is direct and fast. The path that had to be reconnoitered with some care when the metaphor was novel is now worn into a broad avenue that is difficult not to follow.

In the third stage, the metaphor becomes "tired." A direct link is established between the basic and complex levels in the new domain. That is, the expression acquires a new sense, it becomes technical terminology in the new domain. Nevertheless, at this stage, the metaphor can be reactivated. We can be forced to compute anew the path whose computation is no longer ordinarily necessary. For instance, if someone tells me

I live at the foot of a mountain,

I do not see this as a metaphor. But if he then says,

Right next to the big toe.

the comparison is placed squarely before me.

Finally the metaphor dies. Because of changes in the language user's knowledge base or because of the way he learned the expression, he can not recover the path that makes sense of the metaphor. It exists only as an expression in the new domain. Yet at this stage we can still ask, as linguists, what processes "motivate" this expression in this domain [22]—why does the expression make sense—even though as psychologists we do not believe the person uses or could use the processes. Suppose for example someone learns the expression

set a variable *to* a value,

purely as technical terminology, without ever learning the underlying spatial metaphor of, say, setting a dial to a location. A text that would reactivate the metaphor if it were merely tired—"twist a little more" to mean "increase its value"—only baffles him. The metaphorical nature of the expression cannot be said to play a role in his interpretation of it. Nevertheless, its technical sense is not arbitrary. The technical use of "set to" was originally *motivated* by the metaphor. The processes used to interpret it when it was novel can be said to motivate it now.

In summary, the four stages can be described thus. In stage 1, the interpretation is computed. In stage 2, it is computed easily. In stage 3, it is computable, though no longer computed; at this stage, reactivation of the metaphor causes it to be computed again. In stage 4, it is neither computed nor computable, but there is nevertheless a "historical" motivation.

It is controversial whether the so-called "tired" or "dead" metaphors should count as metaphors at all, or whether we should reserve the term for novel examples. Extremes have been argued. Isenberg urges that "metaphor" be reserved for examples that are not just novel, but have artistic intent [23]. Black wants to exclude the example "that no longer has pregnant metaphorical use" [21]. On the other hand, Richards [4] and Whorf [24] see metaphor everywhere—the "fundamental insight" of Part 1. On the far left, Lakoff and Johnson [3] even view nominalizations of verbs as examples of an "event-as-object" metaphor.

Which stages are entitled to be called metaphor? Where should the line be drawn? The above account provides reasons enough for drawing the line anywhere. But in terms of the processes involved, there is simply no point in drawing a line, for they are the same at every stage. What differs is how and when they are used. The reason not to exclude the more decrepit metaphors from our investigation is that they require the same processes to be explicated as do livelier metaphors. But here the processes appear as the processes that *motivated* the expression, not the processes used to *interpret* it.

What are Metaphors and Why Do We Use Them?

I have not argued in this paper that there is no difference between metaphorical and nonmetaphorical usage. Rather I have argued that frequently the interpretation processes for both are identical. There is a distinct thing called metaphor. It is a special and very powerful way of exploiting a knowledge base in the production of discourse. This leads us to the question of what, precisely, is metaphor.

It might seem more appropriate to ask this at the beginning of a paper on metaphor rather than at the end. But in fact what counts as a metaphor is determined by our theory of it. Of course there are central cases of metaphor—statements that are novel and literally false, function effectively in the discourse to make us see one thing in light of another, and involve a mapping between clearly distinct domains—and one's theory of metaphor must encompass these, or one is simply not talking about the same phenomenon as other writers on metaphor. But what else counts as a metaphor is theory-dependent. What one should do then is what I have done in this paper—present the theory and then say what kinds of expressions must be considered metaphors as a consequence.

In the framework presented here, a metaphor is a linguistic expression that involves in its interpretation a mapping (computed, computable, or historical)

from one domain to another via identity for the purpose of making available a new, otherwise unavailable set of inferences. Thus, "people are not cattle" and "set a variable to a value" would both count as metaphors to me.

There is still some indeterminacy in this definition, however: what is meant by "domain"? A rough first cut at this might be that a domain is a collection of predicates and axioms in a knowledge base such that the predicates are richly connected with each other by means of the axioms and are only sparsely connected with other predicates in the knowledge base. But let us look at a range of examples that illustrates the fuzziness of the notion of "domain." In

People are not cattle,

used as a political statement, we are appealing to a mapping from the domain of people and how one interacts with them, to the domain of domesticated animals and how one interacts with them. These are clearly different domains, and thus the sentence contains a metaphor. The sentence

Whales are not fish,

can also be used as a political statement in an argument against the whaling industry. Do whales and fish belong to sufficiently different domains for this to be considered a metaphor. What about

Chimpanzees are not monkeys

in an argument against the use of chimpanzees as experimental animals? Suppose someone asks me if he can borrow one hundred dollars, and I reply

I'm not David Rockefeller

Do David Rockefeller and I belong to sufficiently different domains for this to count as a metaphor?⁶

Consider another range of examples. Suppose my car is a real gas guzzler. I might say one of the following.

My car is the Queen Mary.

My car is a tank.

My car is a truck.

The first is clearly a metaphor. The last is quite dubious. It is perhaps an argument in favor of my definition of metaphor that certain fuzziness in what counts as a metaphor is reduced to the fuzziness in what counts as a domain.

⁶This example is due to Bob Moore.

In the framework presented here, we can also begin to understand why metaphors are used and why they are so pervasive. Any discourse is built on a shared knowledge base of possible inferences. By means of his utterances, the speaker triggers certain of these inferences in the listener's head. The richer the shared knowledge base, the more economical, or equivalently, the more suggestive, the discourse can be. Metaphor is a deceptively simple device for enlarging the knowledge base. By using an apt metaphor to map a new, uncertainly understood domain into an old, well-understood domain, such as spatial relationships, we gain access to a more extensive collection of axioms connecting the basic and complex levels, thereby securing a more certain grasp on the new domain conceptually and providing it with a richer vocabulary linguistically. A metaphor is good to the extent that it taps into a domain that allows a rich collection of inferences to be drawn that otherwise could not be, or equivalently, allows us to see something in a new light. When we learn a new domain, we must learn not just the logical structure of its objects, but also its basic metaphors, generally spatial, and their limits, for by this means we acquire a large chunk of knowledge about the new domain very quickly.

The interpretation problem posed by this very powerful device is that the inferences in the old domain must be sorted out properly. It has been the argument of this paper that the ordinary context-dependent discourse operations, based on selective inferencing, will frequently insure that the right inferences are drawn and the wrong ones aren't.

ACKNOWLEDGMENTS

I have profited from discussions with Armar Archbold, Jaime Carbonell, Gary Hendrix, George Lakoff, and Bob Moore, and also with the participants of the Artificial Intelligence Workshop at the Electrotechnical Laboratory, Tokyo, Japan, in August 1979. None of these people are responsible for any errors in this paper.

REFERENCES

1. D. Davidson, The Logical Form of Action Sentences, in *The Logic of Decision and Action*, N. Rescher (ed.), University of Pittsburgh Press, Pittsburgh, 1967.
2. J. Hobbs, Coherence and Coreference, *Cognitive Science*, 3:1, 67-90, 1979.
3. G. Lakoff and M. Johnson, *Metaphors We Live By*, University of Chicago Press, Chicago, 1980.
4. I. A. Richards, *The Philosophy of Rhetoric*, Oxford University Press, Oxford, 1936.
5. E. Artin, *Galois Theory*, Notre Dame Mathematical Lectures, No. 2, Notre Dame, Indiana, 1959.

6. E. Spanier, *Algebraic Topology*, McGraw-Hill, New York, 1966.
7. S. MacLane, *Categories for the Working Mathematician*, Springer-Verlag, New York, 1971.
8. M. Black, Metaphor, in *Models and Metaphors: Studies in Language and Philosophy*, M. Black (ed.), pp. 25-47, Cornell University Press, Ithaca, New York, 1962.
9. M. Beardsley, *Aesthetics: Problems in the Philosophy of Criticism*, Harcourt Brace, New York, 1958.
10. _____, Metaphor, in *Encyclopedia of Philosophy*, Vol. 5, P. Edwards (ed.), pp. 284-289, 1967.
11. W. Ong, Metaphor and the Twinned Vision, *Sewanee Review*, 63, pp. 193-201, 1955.
12. M. Reddy, The Conduit Metaphor—A Case of Frame Conflict in Our Language about Language, in *Metaphor and Thought*, A. Ortony (ed.), pp. 284-324, Cambridge University Press, Cambridge, 1979.
13. C. Turbayne, *The Myth of Metaphor*, University of South Carolina Press, Columbia, S.C., 1962.
14. J. Morgan, Observations on the Pragmatics of Metaphor, in *Metaphor and Thought*, A. Ortony (ed.), pp. 136-147, Cambridge University Press, Cambridge, 1979.
15. W. Chafe, The Deployment of Consciousness in the Production of a Narrative, in *The Pear Stories: Cognitive and Linguistic Aspects of Narrative Production*, W. Chafe (ed.), Ablex Publishing Company, Norwood, New Jersey, 1980.
16. J. Searle, Metaphor, in *Metaphor and Thought*, A. Ortony (ed.), pp. 92-123, Cambridge University Press, Cambridge, 1979.
17. D. Rumelhart, Some Problems with the Notion of Literal Meanings, in *Metaphor and Thought*, A. Ortony (ed.), Cambridge University Press, Cambridge, 1979.
18. G. Nunberg, "The Pragmatics of Reference," Ph.D. thesis, City University of New York, New York, 1978.
19. G. Miller, Images and Models, Similes and Metaphors, in *Metaphor and Thought*, A. Ortony (ed.), pp. 202-250, Cambridge University Press, Cambridge, 1979.
20. C. Brooks, Metaphor, Paradox, and Stereotype, *British Journal of Aesthetics*, 5, pp. 315-328, 1965.
21. M. Black, More about Metaphor, in *Metaphor and Thought*, A. Ortony (ed.), pp. 19-43, Cambridge University Press, Cambridge, 1979.
22. C. Fillmore, Innocence: A Second Idealization for Linguistics, *Proceedings, Fifth Annual Meeting, Berkeley Linguistics Society*, pp. 63-76, Berkeley, California, 1979.
23. A. Isenberg, On Defining Metaphor, *The Journal of Philosophy*, 60, pp. 609-622, 1963.
24. B. Whorf, The Relation of Habitual Thought and Behavior to Language, in *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, pp. 134-159, John Wiley & Sons, (originally published, 1939), 1956.