# Metaphoric Interpretation by Ontological Categories

Sylvia J. Weaver
and
Sarah A. Douglas

CIS-TR-88-08
June 21, 1988

Department of Computer and Information Science
University of Oregon

# Metaphoric Interpretation by Ontological Categories

Sylvia J. Weaver
and
Sarah A. Douglas

Department of Computer and Information Sciences
University of Oregon
Eugene, Oregon 97403
(503) 686-4408

## 1. ABSTRACT

The problem of metaphoric interpretation has been a neglected topic in computational systems for natural language interpretation. Part of the reason for this is a presumption of semantic well-formedness which excludes metaphors as semantically ill-formed. We present a theory of metaphorical interpretation based on lexical knowledge representation by ontological category attribution. As evidence of support for the theory we found empirical studies of English speakers interpreting classifier/noun pairs predicted as different types of interpretationas predicted. We sketch the design of a computational system based on the theory. Finally, we discuss the implications of the theory for semantic interpretation.

## 2. INTRODUCTION

Computational systems for natural language understanding have typically assumed that semantic interpretation should be approached as an issue of semantic well-formedness similar to syntactic well-formedness. Consider the following sentence:

The idea slept quietly in her brain for weeks.

Certainly we, as readers, can ascribe a meaningful interpretation to such a sentence used to describe the process of an idea developing slowly over a period of time. However, most systems would label it as semantically anomalous, and therefore meaningless, because the verb "sleep" should have an animate surface subject. Yet we, as humans, recognize this sentence as a metaphor and proceed to try to make sense out of it. A reasonable theory of semantic interpretation must be capable of processing metaphor comprehension.

In this paper we will propose a computational process model of metaphoric interpretation based on ontological category structures. We test the theory with empirical studies of English speakers interpreting classifier/noun phrases. These studies provide suggestive evidence that the theory is cognitively viable and capable of predicting the interpretation given to lexical collocations. After the empirical studies, we outline the design of a computational system that incorporates the theory. Finally, we discuss the theory in terms of its implications for metaphoric interpretation in particular and semantic analysis in general. Our analysis leads us to the conclusions that the use of semantic well-formedness in interpretation is wrong and should be questioned in any natural language processing system, and that semantic features such as animacy are not adequate without a categorical structure based on language analysis.

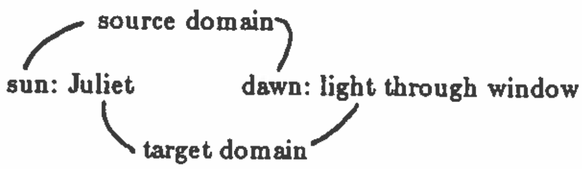
## 3. SEMANTIC INTERPRETATION OF METAPHOR

What exactly is a metaphor? Lakoff and Johnson claim the "[the] essence of metaphor is understanding and experiencing one kind of thing in terms of another." (1980, p. 5). Through metaphor, we can learn about the unknown within the context of that which we already know. Everyday metaphor helps us to grasp new concepts, to view an idea in a fresh, vivid way. Poetic metaphor is a powerful literary device, creating rich imagery. Metaphors expand the power of language.

Humans interpret metaphor as part of the process of understanding language. First of all, we try to make sense of sentences that we encounter. If it's obvious that we can't interpret the sentence literally, then we try to give it a metaphorical interpretation. Consider Romeo's declaration in Shakespeare's "Romeo and Juliet"

> But soft! what light through yonder window breaks?
> It is the east, and Juliet is the sun!

We know that Shakespeare is not asserting the fact that Juliet is a fiery gaseous body in the solar system. We use analogical reasoning instead to infer the author's meaning and temporarily consider "Juliet" a NONLIVING THING. As the sun is to the dawn, so is Juliet to the light coming through the window from her room, as illustrated in the following diagram:




Just as the sun is the source of the dawn, so is Juliet the source of the light. The process of such inference is quite complicated as indicated above since the similarity depends on inferring that the sun's dawn occurs in the east. However, the richness of the metaphor does not depend on making the single most salient inference but also on having many possible inferences. For example, Juliet is also the center of Romeo's universe; their passion burns like the heat of the sun, etc. During metaphor comprehension we search our mental encyclopedias, trying to find similarities between the two domains of the metaphor, the target domain and the source domain.

In summary, metaphoric interpretation appears to have two separable phases. First, a recognition that literal interpretation is not possible: the sentence tends to "sound funny." Second, a complicated inference procedure that generates possible similarities, including that which the rest of the sentence context renders most salient.

## 3.1. The Standard Theory of Semantic Interpretation

Since human language understanders easily comprehend metaphor, it seems mandatory that any system of language analysis be capable of the same. Further, it is possible that metaphoric interpretation should not be considered a "special case" of interpretation. Rather it suggests key issues in the process of semantic interpretation in general. In order to better understand the requirements for a theory of metaphor interpretation, we first review the standard process of semantic interpretation by selection restriction rules.

Most systematic accounts of semantic interpretation in natural language use some form of selection restriction rules incorporating semantic markers or features. While the notion of selection restriction rules does not necessarily require use of a set of primitive features, the theory is usually identified with the semantic theory of componential analysis (Lyons, 1977, p. 422). Componential analysis postulates that the sense of every lexeme can be analysed in terms of a set of more general sense components (or semantic features), some or all of which will be shared by several lexemes in the language. This sharing of atomic sense components can explain synonymity and antonymity of words as well as figure in the more general process of semantic interpretation. Selection restriction rules using these semantic features can be used to differentiate semantically meaningful sentences from those which are semantically anomalous in a manner similar to (and as a subset of) well-formedness in syntactic interpretation. This notion comes principally from the work of Katz and Fodor (1963) and generative semantics. Fillmore (1968) fully developed the notion of semantic case frames wherein selection restriction rules could define the suitability of case fillers using feature agreement for particular verbs. For example, the verb "hit" as in:

John hit the ball.

must have a subject component which is capable of agency, or marked with the feature +ANIMATE.

In addition to defining semantic well-formedness, selection restriction rules can account for the ability of language users to infer propositions that are presupposed or implied, rather than asserted. If, for example, the following sentence fragment is heard:

After a long sleep,....

most hearers will assume that the entity sleeping is animate. This results from a selection restriction rule that requires the verb "sleep" to have an animate surface subject.

Lastly, selection restriction rules can discriminate between multiple senses of a polysemous word. For example,

The man hit the ball.

The restriction of the object of the verb "hit" to a physical object forces the interpretation of the word "ball" to the sense of round object used in a game rather than social dancing.

Semantic selection restriction rules have been incorporated into many computational systems of linguistic parsing, both those traditionally performing a distinct syntactic analysis as well as those claiming to be semantic only. (The best known examples would include Schank & Riesbeck, 1981; Simmons, 1984; Wilks, 1975; Winograd, 1972.) The purpose of using these rules ranges, as described above, from determining semantic well-formedness to determining noun phrase roles to word sense disambiguation. In many so-called semantic parsers, the use of selection restriction rule agreement is crucial for determining the roles (in the Fillmorean sense) of noun phrases.

However, as Hirst (1983) has pointed out, semantic interpretation has frequently been the most ad hoc part of natural language understanding systems. Many systems incorporate the use of selection restriction rules based on semantic features such as +ANIMATE without subscribing to any theory of semantics based on componential analysis and without considering the possible contradictions between it and other semantic theories implicit in their systems. For example, the feature set approach of componential analysis seems at odds with conceptual dependency theory, which proposes a highly relational representation for propositions and a rather different type of semantic primitives. However, all conceptual dependency parsers use some form of feature set, selection restriction rules to determine the case relations of noun phrases.

### 3.2. Computational Systems for Metaphoric Interpretation

The predominance of semantic well-formedness using selection restriction rules for parsing has all but eliminated the possibility of metaphoric interpretation, specifically because metaphors violate those notions. Only a few computational systems for metaphoric interpretation have been proposed. Two of these systems use selection restriction rules in the recognition step just prior to the inferencing of semantic properties and the third does not propose any recognition process but proposes instead a detailed model of metaphoric inference.

Preference semantics, a NLP system for natural language analysis and generation, accommodates metaphor by allowing relaxation of the selection restrictions which determine which words may be semantically grouped together in a sentence (Wilks, 1975). Sentences are reduced to a string of formulas containing semantic and syntactic information. The formulas themselves are composed of strings of elements selected from approximately 70 primitive semantic units. The following example shows the formula for "drink" (from Wilks, 1975)

"drink" (action) --> ((*ANI SUBJ)(((FLOW STUFF)OBJ)
((SELF IN)(((*ANI(THRU PART))TO)(BE CAUSE)))))

The most salient information is rightmost, BE CAUSE. In the sentence

Cars drink gasoline

the most crucial information can be obtained by selecting the head of each formula, yielding

((CONTAINER THING)( BE CAUSE) LIQUID)

The model looks at the head of each formula and matches it onto a basic template. The basic template, using the BE CAUSE verb formula head and LIQUID would be

((MAN (BE CAUSE) LIQUID)

The basic templates are then expanded and the model looks for agreement between the expanded template and the actual formula. Since "drink" prefers an animate object, the basic template and the formula string do not match. The selection restriction rules must then be relaxed to allow for an inanimate subject, and metaphor is recognized.

There is no way, with Wilks' model, to determine aptness of metaphor. Nor does it explain the richness of metaphor by allowing inferences to be drawn between the source domain and the target domain of the metaphor. A metaphor such as "the national economy is ill" would not have the attending metaphors inferred, such as "a cure for the economy" or "a slow recovery for the economy".

In more recent work on metaphor, Carbonell proposed a model for metaphor comprehension based on Lakoff's and Johnson's theory of metaphor (Carbonell, 1982). They propose that there are about 50 general metaphor structures which are the basis for most of the metaphors in the English language. These general metaphor structures include concepts such as time-is-money ("You're wasting her time") and the-mind-is-a-machine ("I've processed that information now").

First, the model attempts to interpret the input literally. If this fails due to a case grammar selection restriction violation, then metaphor is recognized. Otherwise, the input is considered anomalous. "Don't waste time" would fail due to a case selection restriction violation ("waste" requires a physical object) and be recognized as a metaphor. After a metaphor is recognized, it is tested to see if it falls into one of the general categories of metaphor, such as time-is-money. If so, the inferences of the metaphor domain are mapped onto the target domain. Further inferences from the source (metaphor) domain are used to yield additional information about the metaphor. Then the program looks at the reason why this particular metaphor was chosen. What was the intent of the writer? This particular instantiation of a general metaphor can be then stored for later reference in the discourse.

This model does not readily account for metaphors which do not fall into one of the general categories. It deals best with metaphors that are already at least partially idiomatic, also called "dead" metaphors. It does not reflect the ease with which humans seem to be able to comprehend new, unusual metaphors.

Indurkhya (1985) presents a formalized and computable model of metaphor comprehension as a special case of a generalized cognitive function, namely analogical reasoning. His model allows inferences to be drawn by mapping between the source and target domains, but does not address the problem of metaphor recognition during parsing.

## 4. THE PROPOSED THEORY

The metaphor recognition theory proposed in this paper is based on lexical category membership that broadens the notion of selection restriction to allow literal as well as metaphoric interpretation in a highly predictable way. The notion of semantic features is redefined as a hierarchical structure of ontological category definition. Because the theory can cope with literal and non-literal interpretations, literal interpretations are given to those word groups which follow selection restriction rules, while selection restriction violations are those word groupings which can be given a metaphorical or physical transformation interpretation. Theories of metaphor based on selection restriction violations (of subject/verb semantic type agreement) are thus accommodated. Essentially the theory does away with the notion of failed selection restriction rules since an interpretation is always possible.

Keil (1979) has proposed a tree structure with which to represent the way humans categorize ontological knowledge, based on the work of the philosopher Fred Sommers (1971). (See FIGURE 1.) Ontological categories represent major organizational characteristics of things, such as PHYSICAL or NON-PHYSICAL, ANIMATE or INANIMATE, or LIVING or NONLIVING. These categories can be structured hierarchically, so that some categories include other categories as subsets. For instance, the category LIVING THINGS would include the categories ANIMATE (animals) and INANIMATE (plants). Membership in a given category would imply membership in the parent node above the given category. An item is in any category described at nodes above the item's position on the tree, but an item may not be in a category which can only be reached by traveling downwards on the tree at some point. For instance, "milk" may not be in the category SOLID OBJECTS, because to reach that category, one must travel not only upwards to PHYSICAL OBJECTS, but downwards again to SOLID OBJECTS. This theory of restriction on category membership is known as the M constraint theory, and reflects the tree's hierarchical ordering of the categories. An item may not have simultaneous membership in two categories, one of which is not a subset of the other.

Keil has constructed a predicability tree corresponding to the ontological tree. Predicability determines what can be sensibly asserted of a term. There are predicates which can be categorized on the ontological tree according to the nouns with which they can be paired. For instance, the predicate "is thoughtful" can be paired with "the woman", but not with "the penny". This is because SENSIBLE BEINGS are capable of thought. An ontological attribute does not determine truth or falsity, only whether a predicate is verifiably true or false. In fact, the woman may not be thoughtful. This is quite different though from determining the nonsensical truth or falsity of whether a penny is thoughtful.

Keil has empirically tested the validity of the predicability tree with numerous experiments (Keil, 1979, 1981). His experiments with young children are based on the hypothesis that ontological categories are learned by young children as part of language acquisition.

In order for Keil's subjects (both adults and children) to effectively test his predicability tree, they had to be able to distinguish between literal and metaphorical interpretation of a phrase. Keil did not test for metaphor; instead he chose to lump it together with anomaly for the purpose of constructing the tree. He instructed his subjects to consider a sentence anomalous if a metaphorical interpretation was necessary for it to make sense.

The notion of a predicability tree can be extended to include other word groupings used to convey a specific concept, such as classifier/noun phrases. Classifiers are semantic units (morphemes) which give semantic category information about the nouns which they modify. In English, classifiers are words, but in other languages, they may be smaller units of meaning (bound morphemes) attached to other words. English classifiers are optional, but in many languages, they are obligatory. Their descriptive function serves to give information about the category in which the nouns which they modify belongs. For instance, the English classifier "blade" used with the noun "grass" in "a blade of grass" tells us that the piece of grass is a long, thin, flat shape. In the phrase "a drop of milk", "drop" gives the information that milk is in the category LIQUIDS and that there is only a small bit of liquid.

Classifiers can be categorized on Keil's ontological tree, with a classifier's span limited to the category in which it is placed. When a classifier is used in a literal sense, it may only modify nouns which appear in the same category or a subset thereof. For example, the classifier "pile" is associated with the category NONLIVING SOLID OBJECTS. Intuitively, we do not refer to "a pile of pigs" when we see a group of pigs. The classifier "group", however, spans both living and nonliving categories, so it can be used to modify both "pigs" and "stones". Note that since "group" spans more categories than "pile", it has less semantic distinction than "pile".

When a classifier is paired with a noun from its category or a category below it on the tree (which implies that it is a subset of the classifier's category), the phrase can be interpreted literally. The inherent semantics of the noun remain the same, with the classifier providing some additional information about the noun. What happens, however, when there is a violation of the M constraint theory? If the classifer is from a category which doesn't span the category of the noun which it modifies, then another interpretation is called for.

When a classifier from a given category on the ontological tree is used to modify an item from a category reached by traveling upwards on the tree and then down a right branch, metaphor is implied. Consider the following situation: a logjam is floating down a river. If we refer to this logjam as a "school of logs", we are creating a metaphor. "School" is a classifier used for the category of FISH farther down on the ontological tree. We don't infer here that the logs are propelling themselves, or that they are motivated to exhibit schooling behavior: "school of logs" is to be interpreted metaphorically. This means that the classifier has one or more features which the noun does not have. In the above example, the ontological category for "school" has the attributes LIVING and ANIMATE. Only in a fantasy situation could one imagine that logs could actually be animate beings.

If a classifier is used with a noun reached by traveling up from the classifier's category and then down a left branch, metaphor is not implied. In these cases, the classifier must be in a category on a different branch of the tree than the noun. In the example "pile of trees", one gets to the classifier "pile" by traveling upwards on the tree from "trees" and down again on another branch. Instead of metaphor, in this case the noun's feature which distinguishes it from the classifier's category changes value in order to resolve the feature differences between the noun and classifier. The noun must undergo physical transformation (as opposed to the fantasy situation implied in metaphor) to change a feature value. This phenomenon of physical change interpretation is a specialized instance of a literal interpretation. In our example "pile of trees", "pile" comes from NONLIVING THINGS and "trees" from LIVING THINGS. "Trees" should take on here the attribute of NONLIVING. Intuitively, we can say that this is exactly what "pile of trees" implies. In this way, physical change interpretation functions as a marking system, highlighting a feature value not usually associated with the noun.

Note that in the above example, the noun categories do not actually change their place on the ontological tree. Rather, the noun temporarily takes on the salient attribute from the classifier's category which distinguishes it from sibling categories.

## 5. EMPIRICAL VALIDATION

In order to test out the principle ideas elaborated by this theory of metaphor recognition, we decided to first determine the appropriate ontological membership of selected classifiers to give a predictable meaningful interpretation with a compatible noun and then to explicitly violate this "normal" semantic relationship by pairing the classifier with nouns predicted to give both metaphoric and physical change interpretations. These two empirical studies will be referred to as the "Classifier Membership Study" and the "Predicted Interpretation Study."

### 5.1. Classifier Membership Study

*Procedure.* Fourteen classifiers were chosen from various ontological categories. Each classifier was initially placed in an ontological category, determined by the experimenter's intuition. Classifiers were then paired with nouns (chosen randomly) from their own categories and nouns (chosen randomly) from other categories, with the prediction that the classifier/noun pair would make less sense if it was not from the same category. Subjects were given 32 classifier/noun pairs. Their task consisted of rating the phrases on a scale from one to five according to how much sense the phrases made to them. A score of one indicated that the phrase made no sense, and a score of five indicated that the phrase made sense. Subjects were asked to work quickly, relying on their first impression as to how much sense each phrase made. For most subjects, the task required only a few minutes. Thirteen subjects participated in the experiment. Ten were graduate students in computer science. Two were undergraduates and one was a high school graduate. All were naive to the experimental hypothesis.

*Analysis.* A t-test for a difference between a sample mean and the population mean, compared the difference between the mean for a classifier used with a noun from a given category and the sample mean of 3, with 3 representing a score which is indeterminate, neither indicating exclusion or inclusion in a category. It should be noted here that the actual mean was 3.94, showing the subjects' bias towards a higher score (toward making sense). If the subjects determined that a classifier/noun phrase made sense, then the classifier could be considered to span at least that noun's category, showing inclusion in that category. If a classifier/noun pair did not make sense, then the classifier had to be considered excluded from that category. If there was no significant difference (for $p<.05$, $df = 12$, a significant difference of 2.179 was required), then the classifier's membership status for the given category was indeterminate and all noun phrases using that classifier were excluded from further experimentation. This rather stringent requirement guaranteed that only clearly agreed upon literal meanings were used for the more unusual tasks subjects would be required to perform in the second empirical study.

*Results.* TABLE 1 summarizes the findings of this first study and FIGURE 2 illustrates the accepted classifer/noun pairs as arranged on the ontological tree. Seven of the classifiers remained in the pre-experimental categories. Three of the classifiers were shown by the experimental results to belong in categories other than the pre-experimental categories. (So much for experimenter intuition!) Four classifiers had to be eliminated because their category memberships were shown to be indeterminate.

### 5.2. Predicted Interpretation Study

The theory of metaphor recognition claims that subjects can distinguish between metaphorical, literal, and physical change interpretations. The theory also claims that these interpretations are predictable, according to the categories of the classifiers and nouns in the phrases. The problem with testing the theory originally seemed to be one of ensuring that the phrases were entirely context-free, so that the subject would only look at the semantic content of the given phrase.

Three pilot experiments were given to a few subjects to see if the subjects could understand and perform the task as intended by the experimental design. Different subjects were used for each version of the pilot experiment. The experimental task was revised three times before a final format was chosen.

None of the three pilot experiments seemed to get at the heart of the information which the experiments were supposed to reveal. The post-experimental interviews were often more enlightening regarding the subjects' intuitions than the pilot experiments themselves. None of the pilot experiments were able to provide the subjects with a completely context-free environment for each phrase being examined. This makes sense, because in the everyday real world, people always have a context which they operate within. The answer for the experimental design seemed to lie in discovering the elaborated context which the subject used for interpretation. At this point, it seemed more appropriate to explore the process of interpretation used by the subjects. Thus, this second study provides strong support for the theory but does not experimentally test it.

*Procedure.* Four subjects were used. Three were native English speakers and one was a non-native English speaker. The exploratory data analysis consisted of a series of 28 tasks. It was conducted within a protocol format. Each session was videotaped. In each session, the subject was orally presented with the series of tasks. Each task consisted of the following: the interviewer read a classifier/noun phrase to the subject. The subject was then asked to make up a sentence using that phrase. This task forced the subject to try to makes sense of the phrase and provide a context for it. The interviewer then asked the subject to paraphrase the sentence. This part of the task ensured that any ambiguities would be resolved. If the subject's sentence and paraphrase did not clearly indicate if particular features were present, the interviewer asked more questions to elicit this information. Seven nouns were used in constructing the 28 phrases. Each noun was paired with two classifers from its own category (predicted literal), one classifier a predicted metaphor category, and one classifier from a predicted physical change category. The order of these phrases were scrambled to minimize the possibility of the subject recognizing a pattern to the tasks.

*Analysis.* Subjects' responses were evaluated by the interviewer. If the features and transformations matched up with the predicted ones, the response was scored as a success; otherwise it was scored as a failure. The scoring was not done by an independent judge, so there was a bias.

*Results.* There were 94 successes out of a possible 100% match of 112 items (28 tasks x 4 subjects). This represents a success rate of 84%. The success rate for the predicted metaphor categories was 96% (27 successful matches out of a possible 28), for the literal category it was 86% (48 successful matches out of a possible 56), and for the physical change category, it was 68% (19 successful matches out of a possible 28). The scores for the stimulus phrases are displayed in TABLE II.

There were so many unknown factors, such as the probability of a success, that a more complex statistical analysis was not appropriate. Some of the failures were not complete failures; there was a partial match with the predicted response, but because it was not a complete match, it had to be scored as a failure.

The cases where the subject's response did not match with the predicted response can be attributed to a number of different reasons. The main discrepancy occurred in the initial categorization of either the classifier or the noun. Within the context created by different subjects, classifier and nouns were sometimes placed in categories different from the categories predicted by the first empirical study.

In some instances, there was a failed match between the minor predicted features of either a classifier or noun and the features ascribed by the subject. These differences did not result in a change in category, but they had to be scored as a failed match nonetheless. With other failed matches, the predicted physical transformation which the noun would have to undergo was too great for the subjects to easily imagine. For instance, with the phrase "drop of radio", the radio would have to liquify (i.e. melt) to reconcile with the classifier "drop."

One of the difficulties in testing the theory lies in the inherent non-static quality of language. The constant change which language undergoes is reflected in the categorization of nouns and classifiers. For example, "pile" usually modifies a heap of nonliving objects, but its use can be informally extended to simply mean "many" and "disorderly".

People also try to make sense from what they read and hear. When a phrase doesn't seem to match the world as they know it, they try to make it fit, using metaphorical or physical change interpretations. Sometimes, however, it's easier to change or modify the meaning of the words involved (as in "drop of radio") then to only imagine a context for a metaphorical or physical change interpretation.

The support for the theory provided by the protocol analysis is strong, particularly when there were so many ways in which the subjects' interpretations could fail to match the theory's predictions. A more tightly structured experiment could be designed if it were first determined what the subjects' categorizations of the classifiers and nouns were, and theory predictions made on the basis of their classifications.

## 6. COMPUTATIONAL IMPLEMENTATION

In this section, we will briefly sketch a design for implementing this theory in a computer parser and interpreter. Earlier, we postulated that there are two separate processes for interpretation: a recognition phase and an inferencing phase. Our theory has briefly sketched this process which is elaborated in the following algorithm.

```
begin
        determine category of classifier;
        determine category of noun;
        if classifier does not span the noun's category
                then if noun is from a category higher on the tree
                        then interpret metaphorically
                        else interpret with physical change interpretation
                else interpret literally;
end.
```

Although we have used the components of classifier/noun modification to illustrate the theory, other modification or collocation components could be used, e.g. agent/verb pairs. This accommodates within the theory the notion of selection restriction rules and semantic markers and provides a more powerful recognition process.

Both recognition and interpretation phases require that all open-class lexical items be grouped on the ontological tree. We imagine that different senses of a word may have a different positions. The organization of this tree as a data structure is the standard object-oriented or frame representation. What the theory demands, though, is a principled organization to the structure of knowledge. This structure is that of ontological categorization.

During the second phase of interpretation, the theory predicts that certain features will be present or absent, depending on inherent features of the noun used and the relative positions of the lexeme categories on the ontological tree. The classifier provides additional information about the noun in all three cases of metaphorical, literal, or physical change interpretation. These different interpretations are illustrated in FIGURE 3 with specific examples of represented semantics for example classifier/noun pairs.

In order to resolve the differences between the features of the classifier's category and the noun's category, the noun must undergo certain transformations. In the case of physical change, these are physical changes; a certain feature of the noun must change its value, such as SOLID transforming to NON-SOLID. This change is indicated in FIGURE 3 as *physical trans.* In the case of metaphor, the change a noun must undergo to match up with the classifier's features is of a different nature. The noun must acquire a feature which it previously did not have. The only way this could take place is in a fantasy world, not in the real physical world defined by Western culture. Fantasy transformation is indicated in FIGURE 3 as *fantasy trans.* In all cases, the additional semantic information supplied by the classifier is shown below as *add info.*

This brief sketch of interpretation is based on a simple feature matching concept, and needs much more definition. It does provide a workable method for representing word senses and building collocative meanings from them. We feel that these issues of interpretation need much more work integrating them into a propostional representation. The method presented is much simpler than the complex analogical mapping methods proposed by Indurkhya (1985). In many ways, the problem of focus becomes crucial as the number of possible mappings could increase exponentially with the number of attributes and relations. It is our feeling that metaphorical interpretation captures the most salient surface similarities between the two items of comparison. The context of the surrounding sentence(s) is what acts as the focus.

## 7. DISCUSSION

In this section we explore some of the limitations of the theory we have presented as well some interesting implications.

### 7.1. Other Aspects of Semantic Interpretation

How does this theory deal with anomaly? If the putative anomaly is only a category mistake (i.e., a pairing of a noun and a classifier from nonintersecting ontological categories), then it is treated as either a metaphor or an instance of physical change interpretation. This theory cannot cope with other forms of anomaly, such as logical contradictions (colorless green, for example) or tautology. An system which uses this theory to detect and interpret metaphor would also need to have a logical inference system to detect anomaly in the form of logical contradictions and tautologies, although we would probably argue that even contradictions and tautologies have meaning. Consider the following sentences:

Business is business.

Is Mary married? She is and she isn't.

While the first is a logical tautology and the second a contradiction, they both convey meaning. Thus, we conclude that the notion of semantic anomaly is very sticky and that a system should attempt an interpretation under all possible circumstances.

This theory also does not reflect the influence of context. If a particular classifier/noun phrase refers to previously described situation, then the phrase might require a literal interpretation even though metaphor or physical change interpretation seem to be implied. A phrase such as "pile of moonbeams" might not be a metaphor, but might refer instead to a heap of foil-covered cardboard cutouts used in a school play.

### 7.2. Diachronic Change

The scope of a lexeme may also change over time. For instance, the classifier "tribe" once was used to refer only to humans (specifically to groupings of Roman peoples). Today its use has been extended to include nonhuman groups, such as primates. It has changed categories from HUMAN to ANIMALS. The change in usage is neatly reflected by an upwards movement on the ontological tree, usually a change in category to the parent node. The indeterminate results for some of the classifiers studied in the first empirical study which could not be definitely categorized may also reflect a diachronic change currently taking place in the classifiers' usage.

This upward movement of the classifier on the ontological tree also illustrates the death of a metaphor. As a classifier becomes more commonly associated with a noun from a category higher up on the tree, the metaphor loses its ability to surprise the listener. Eventually the classifier may move to a higher node to accommodate the literalization of the phrase (the new category now allows the classifer to span the noun's category).

The notion of lexical change moving from the more specific to the more abstract was postulated by Porzig (as discussed in Lyons, 1977, pp. 262-266). He theorized that the original meaning of words corresponds to concrete objects, activities, qualities, etc. Over time most words will come to be applied to a wider range of things in a wider range of situations. While this is a nice explanation of how lexical change incorporates the theory of metaphor, it is not a universal explanation of semantic change in a language. There are equally many words which move from a more general to a more specific meaning. This also seems to be the direction of lexical acquisition in language learning.

### 7.3. Aptness of Metaphor

Keil (1979) suggests that aptness of metaphor might be predictable using the ontological tree. He proposes that the closer the categories for the subject/predicate pair (or classifier/noun pair), the more apt the metaphor. In more recent work Kelly & Keil (1985) have postulated that ontology is such an important property of an object's identity that a mythmaker would resist changing it across a metamorphosis. Using Grimm's *Fairy Tales* and Ovid's *Metamorphoses* as data, they show that the numbers of transformations of conscious beings into members of other ontological categories decreases as the distance between the ontological categories increases.

However, this argument seems to fall down for particular metaphors such as the "Juliet is the sun" metaphor. Indurkhya (1985) proposes, based on the work of the psychologists Tourangeau and Sternberg (1981), that the goodness of the metaphor can be measured according to the interdomain dissimilarity and the intradomain similarity. If there is much intradomain similarity, then many inferences can be drawn, as is the case between sailing and farming. For example, one can speak of a boat "plowing the waves". If the two domains (the source domain and the target domain) are dissimilar, then the novelty of the comparison adds to the value of the metaphor. The ontological tree reflects dissimilarities by the inherent properties of the different categories, so in this way, the tree could be a simple measure of one aspect of the aptness of the metaphor by noting the length of the pathway from the source domain node to the target domain node. Note that this measure of the aptness of metaphor directly contradicts Keil's prediction regarding the aptness of metaphor.

## 8. CONCLUSIONS

We have presented a theory of metaphorical interpretation based on the use of ontological category structures. Using the example of classifier/noun pair interpretation we have conducted some rudimentary empirical studies of English speakers. From this supportive evidence we have sketched the design of a computational system based on the theory. By noting the relative position on the ontological tree of the classifier and noun in the classifier/noun pair, this theory can determine if a phrase should be interpreted literally, metaphorically, or if the noun's meaning should be shaded by the classifier's category. Although only classifier/noun phrases are used to illustrate the theory, it also applies to other phrases. As long as the words can be individually categorized the ontological tree, the theory can determine appropriate interpretation. In this way, the theory can accommodate different case combinations.

While the above description of a computational implementation of the theory is by no means complete, it does provide a basic design by which we will implement this theory. In particular, we feel that structuring the lexicon as part of the knowledge representation mechanisms of most AI systems and specifically organizing it by an ontological structure make the process of recognition and interpretation less ad hoc. Further, this approach forces us to abandon the use of rules to detect semantic well-formedness and instead to focus on semantic interpretation of all propositions, including those traditionally ignored such as metaphor. The assumption that meaningfulness in a language should be approached as a notion of semantic well-formedness has been questioned by only a few. The most compelling arguments are made by Lyons, 1977. This paper has attempted to explore this notion through the development of a theory of metaphor interpretation, a process which we feel has been sorely neglected in computational linguistics.

## 9. REFERENCES

Carbonell, J. G. (1982). Metaphor: an inescapable phenomenon in natural-language comprehension. In W. G. Lehnert & M. H. Ringle (Eds.), *Strategies for Natural Language Processing.* Hillsdale, New Jersey: Erlbaum.

Fillmore, C. J. (1968). The case for case. In Bach & Harms (Eds.), *Universals in Linguistic Theory.* New York: Holt, Rinehart & Winston.

Hirst, G. (1983) A foundation for semantic interpretation. In *Proceedings of the 21st annual Meeting of the Association for Computational Linguistics..* Association for Computational Linguistics.

Indurkhya, B. (1985). A computational theory of metaphor comprehension and analogical reasoning. *Boston University Tech Report No. 85/001.* Boston, Massachusetts.

Keil, F. C. (1979). *Semantic and conceptual development: An ontological perspective.* Cambridge, Massachusetts: Harvard University.

------. (1981). Constraints on knowledge and cognitive development. *Psychological Review*, 88, 197-227.

Keil, F. C. and Kelly, M. H. (1985). The more things change...: Metamorphoses and conceptual structure. *Cognitive Science, 9*, 4.

Lakoff, G., and Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lyons, J. (1977). *Semantics*. Cambridge: Cambridge University Press.

Schank, R. & Riesbeck, C. (1981). *Inside computer understanding*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Simmons, R. (1984). *Computations from the English*. Englewood cliffs, N.J.: Prentice-Hall.

Sommers, F. (1971). Structural ontology. *Philosophia*, 1, 1, 21-42.

-------. (1963). Types and ontology. *Philosophical Review*, 72, 327-363.

Tourangeau, R., and Sternberg, R. J. (1981). Aptness in Metaphor. *Cognitive Psychology*, 13, 27-55.

-------. (1982). Understanding and appreciating metaphors. *Cognition*, 11, 3, 203-244.

Wilks, Y. (1975). An intelligent analyzer and understander of English. *Communications of the ACM*, 18, 5, 264-274.

Winograd, T. (1972). *Understanding natural language*. New York: Academic Press.

FIGURE 1

*The Ontological Tree*

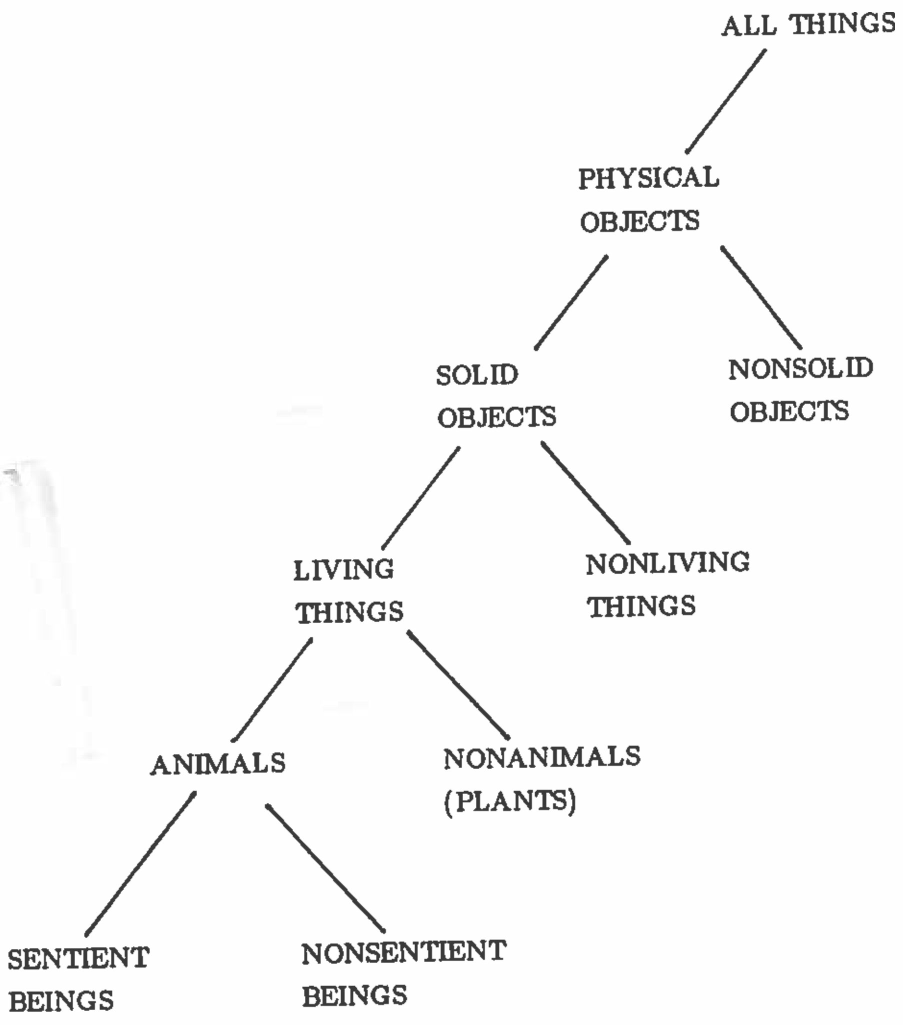
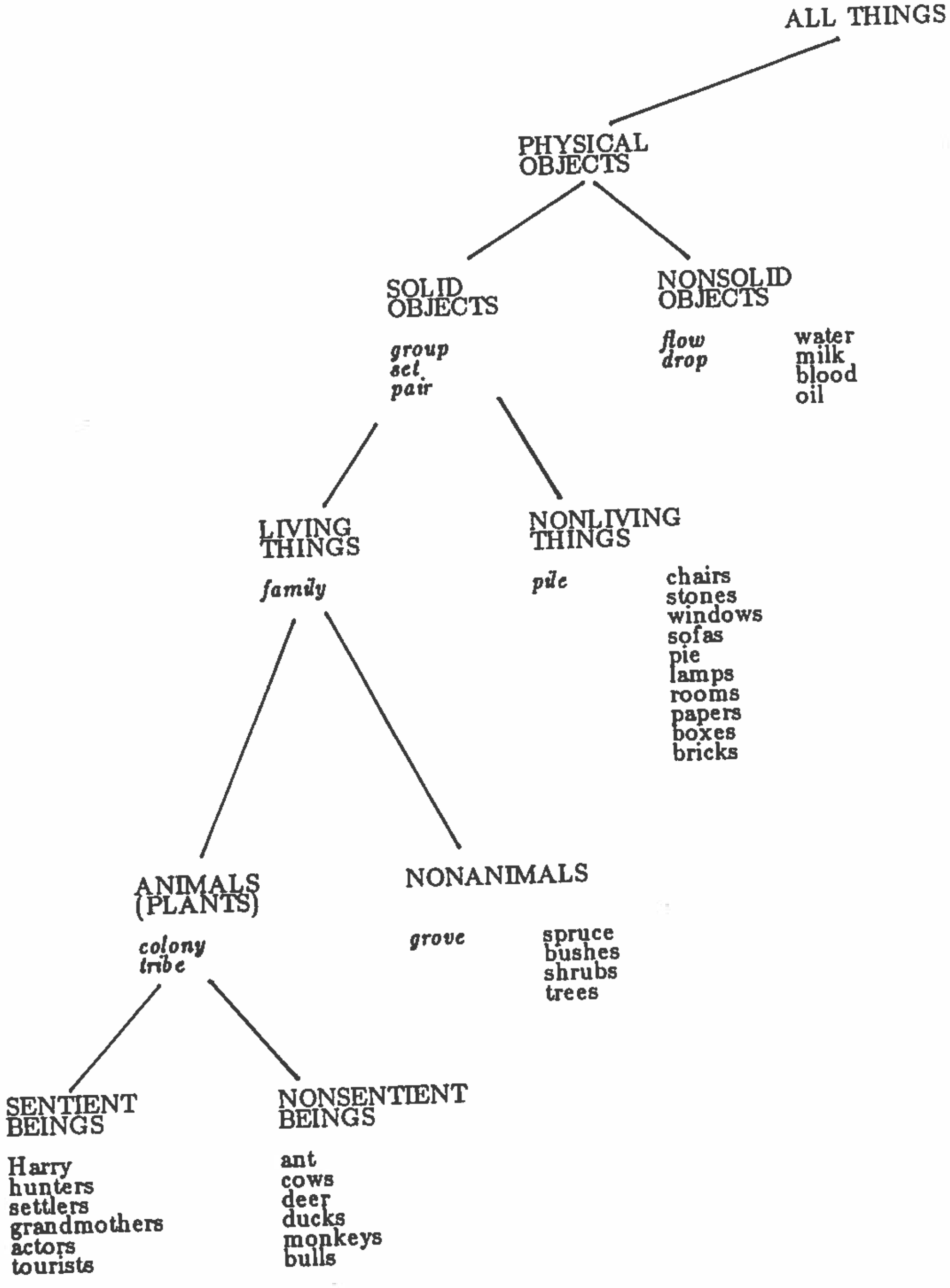


**FIGURE 2**

*Post-Experimental Positions of Classifiers and Nouns on the Ontological Tree*




NOTE: Classifiers are denoted with italics; nouns with Roman characters.

**FIGURE 3**

*Representation of Three Types of Interpretation*

**Example 1: Literal Interpretation**

"Grove of trees" (literal)

"grove"
(phys obj (solid (living (-animal)))) / (whole (plural)) (together)

*add info*

"trees"
(phys obj (solid (living (-animal)))) / (whole (plural))

**Example 2: Metaphorical Interpretation**

"Tribe of carrots"

"tribe"
(phys obj (solid (living (animal)))) / (whole (plural)) (together)

*fantasy trans* *add info*

"carrots"
(phys obj (solid (-living))) / (whole (plural))

**Example 3: Physical Change Interpretation**

"Pile of trees"

"pile"
(phys obj (solid (-living))) / (whole (plural)) (heaped)

*physical trans* *add info*

"trees"
(phys obj (solid (living (-animal)))) / (whole (plural))

NOTE: The information appearing before the slashes represents the category membership of the noun or classifier. The information after the slash represents, in the case of the classifiers, semantic information highlighted by that particular classifier. In the case of the noun, this information can be gleaned by the singular or plural information given by the noun, and unmarked features associated with the noun (e.g., a carrot is usually whole unless otherwise marked). With a literal interpretation, no transformation occurs, because there are no feature differences to resolve, but the classifier does supply additional information about the noun.

TABLE I

*Summary of Classifier Membership Study*

| Phrase | Range | Mean | SD | t-stat | Same Category |
|---|---|---|---|---|---|
| Group of bushes | 4 | 4.15 | 1.14 | 3.94 | + |
| Group of milk | 2 | 1.15 | 0.38 | -17.76 | - |
| Group of chairs | 4 | 4.46 | 0.88 | 6.00 | + |
| Set of lamps | 4 | 4.15 | 1.14 | 3.63 | + |
| Set of trees | 4 | 3.85 | 1.14 | 2.68 | + |
| Bunch of water | 4 | 2.77 | 1.17 | -0.64 | indet. |
| Bunch of papers | 3 | 4.69 | 0.63 | 9.67 | + |
| Bunch of cows | 3 | 4.54 | 0.78 | 7.15 | + |
| Pair of blood | 4 | 1.62 | 0.96 | -6.18 | - |
| Pair of stones | 3 | 4.54 | 0.78 | 4.96 | + |
| Pair of ducks | 4 | 4.62 | 0.87 | 6.72 | + |
| Flow of water | 4 | 4.38 | 0.96 | 5.18 | + |
| Flow of sofas | 3 | 1.23 | 0.83 | -7.67 | - |
| Drop of oil | 2 | 4.92 | 0.28 | 24.76 | + |
| Drop of boxes | 5 | 1.69 | 1.25 | -3.78 | - |
| Pile of bricks | 4 | 4.46 | 1.13 | 4.87 | + |
| Pile of grandmothers | 5 | 1.46 | 1.13 | -4.93 | - |
| Suite of rooms | 5 | 3.77 | 1.69 | 1.64 | indet. |
| Suite of bulls | 2 | 1.15 | 0.38 | -14.08 | - |
| Piece of pie | 2 | 4.92 | 0.28 | 24.96 | + |
| Piece of Harry | 5 | 3.15 | 1.72 | 0.31 | indet. |
| Family of tourists | 4 | 4.23 | 1.01 | 4.23 | + |
| Family of shrubs | 4 | 3.85 | 1.07 | 2.87 | + |
| Family of windows | 4 | 1.85 | 0.90 | -4.61 | - |
| Colony of settlers | 4 | 4.69 | 0.85 | 7.13 | + |
| Colony of ants | 4 | 4.54 | 0.88 | 6.33 | + |
| Grove of trees | 4 | 4.38 | 1.12 | 4.44 | + |
| Grove of deer | 3 | 1.77 | 0.83 | -5.33 | - |
| Stand of spruce | 5 | 3.54 | 1.71 | 1.14 | indet. |
| Stand of actors | 3 | 2.08 | 0.76 | -0.44 | indet. |
| Tribe of monkeys | 3 | 4.31 | 0.95 | 4.99 | + |
| Tribe of hunters | 4 | 4.31 | 1.11 | 4.26 | + |

Note. For the t- test, df = 12; $p < .05$; significant difference = 2.179

TABLE II

*Summary of Predicted Interpretation Study*

| Phrase | Predicted Type | Prediction Success |
|---|---|---|
| Pair of carrots | literal | 3/4 |
| Tribe of carrots | metaphor | 3/4 |
| Pile of carrots | literal | 3/4 |
| Drop of carrot | physical change | 3/4 |
| Pile of stoves | literal | 4/4 |
| Tribe of stoves | metaphor | 4/4 |
| Group of stoves | literal | 4/4 |
| Flow of stoves | physical change | 4/4 |
| Grove of trees | literal | 3/4 |
| Colony of trees | metaphor | 4/4 |
| Family of trees | literal | 4/4 |
| Pile of trees | physical change | 3/4 |
| Pile of rugs | literal | 4/4 |
| Tribe of rugs | metaphor | 4/4 |
| Pair of rugs | literal | 2/4 |
| Flow of rugs | physical change | 4/4 |
| Set of radios | literal | 3/4 |
| Family of radios | metaphor | 4/4 |
| Pile of radios | literal | 3/4 |
| Drop of radio | physical change | 0/4 |
| Pair of rosebushes | literal | 4/4 |
| Tribe of rosebushes | metaphor | 4/4 |
| Family of rosebushes | literal | 4/4 |
| Pile of rosebushes | physical change | 2/4 |
| Pile of chairs | literal | 3/4 |
| Grove of chairs | metaphor | 3/4 |
| Set of chairs | literal | 4/4 |
| Flow of chairs | physical change | 3/4 |