

# World Knowledge and Word Meaning

Jerry R. Hobbs  
Artificial Intelligence Center  
SRI International

We use words to talk about the world. Therefore, to understand what words mean, we must have a prior explication of how we view the world. In a sense, efforts in the past to decompose words into semantic primitives were attempts to link word meaning to a theory of the world, where the set of semantic primitives constituted the theory of the world. With the advent of naive physics and research programs to formalize commonsense knowledge in a number of areas in predicate calculus or some other formal language, we now have at our disposal means for building much richer theories of various aspects of the world, and consequently, we are in a much better position to address the problems of lexical semantics.

In the TACITUS project for using commonsense knowledge in the understanding of texts about mechanical devices and their failures, we have been developing various commonsense theories that are needed to mediate between the way we talk about the behavior of such devices and causal models of their operation (Hobbs et al., 1986). The theories cover a number of areas that figure in virtually every domain of discourse, such as scalar notions, granularity, structured systems, time, space, material, physical objects, causality, functionality, force, and shape. Our approach has been to construct *core theories* of each of these areas. These core theories may use English words as their predicates, but the principal criterion for adequacy of the core theory is elegance, whatever that is, and this can usually be achieved better using predicates that are not lexically realized. It is easier to achieve elegance if one does not have to be held responsible to linguistic evidence. Predicates that *are* lexically realized are then pushed to the periphery of the theory. A large number of lexical items can be defined, or at least characterized, in terms provided by the core theories. The hypothesis is that once these core theories have been formulated in the right way, it will be straightforward to explicate the meanings of a great many words.

The phrase “in the right way” is key in this strategy. The world is

complex and can be viewed from many different perspectives. Some of these will lend themselves well to the investigation of problems of word meaning, whereas others will only lead us into difficulties. We could, for example, axiomatize space as Euclidean 3-space, with  $x$ ,  $y$ , and  $z$ -coordinates for every point. We could then attempt to define what the various prepositions and verbs of motion mean in this framework. I am quite sure such an attempt would fail. Such a theory of space would be too foreign to the way we talk about space in everyday life. Even if we were to succeed in this limited task, we would not have advanced at all toward an understanding of metaphorical uses of these words.

In contrast, we view our core theories not so much as theories about *particular* aspects of the world, but rather as abstract frameworks that have proven useful in interpreting, generally, a number of different kinds of phenomena. Thus, at the very center of our knowledge base is an axiomatization of “systems”, where a system is a set of elements and a set of relations among them. An abstract, “primitive” relation *at* places entities at locations within a system, encoding the basic figure-ground relation. A large number of things in the world can be understood as systems, and a large number of relations can be understood as *at* relations. When we apply the theory to a particular phenomenon, we buy into a way of thinking about the phenomenon, and, more to the present purposes, of talking about it. It is in this way that the metaphorical usages that pervade natural language discourse are accommodated. Once we characterize some piece of the world as a system, and some relation as an *at* relation, we have acquired the whole locational way of talking about it. Once this is enriched with a theory of time and change, we can import the whole vocabulary of motion. For example, in computer science, a data structure can be viewed as a system, and we can stipulate that if a pointer points to a node in a data structure, then the pointer is *at* that node. We have then acquired a spatial metaphor, and we can subsequently talk about, for example, the pointer *moving around* the data structure. Space, of course, is itself a system and can be talked about using a locational vocabulary.

Also central in the knowledge base is an axiomatization of “scales”, which is a particular kind of system whose relations are a partial ordering and an indistinguishability relation (encoding granularity). Once we develop a core theory of scales, we can use the predicates it provides to characterize a large number of lexical items, such as “range”, “limit”, and the comparative and superlative morphemes. For  $x$  to range from  $y$  to  $z$ , for example, is for  $y$  and  $z$  to be endpoints of a subscale  $s$  of a scale, and for  $x$  to be a set of

entities which are located *at* elements of  $s$ . By choosing different scales, we can get such uses as

The buffalo ranged from northern Texas to southern Saskatchewan.

The students' SAT scores range from 1100 to 1550.

The hepatitis cases range from moderate to severe.

His behavior ranges from sullen to vicious.

Our desire to optimize the possibilities of using core theories in metaphorical and analogical contexts leads us to adopt the following methodological principle: For any given concept we wish to characterize, we should determine the minimal structure necessary for that concept to make sense. In efforts to axiomatize some domain, there are two positions one may take, one exemplified by set theory and the other by group theory. In axiomatizing set theory, one attempts to capture exactly some concept one has strong intuitions about. If the axiomatization turns out to have unexpected models, this exposes an inadequacy. In group theory, by contrast, one characterizes an abstract class of structures. If there turn out to be unexpected models, this is a serendipitous discovery of a new phenomenon that we can reason about using an old theory. The pervasiveness of metaphor in natural language discourse shows that our commonsense theories of the world ought to be much more like group theory than set theory.

Our approach to space and dimensionality illustrates this. Rather than defining dimension in the classical manner of linear algebra, in a way that requires a measure and arithmetic operations, we have sought to be able to build spaces out of less structured components. Thus, we have defined a two-dimensional space as a set of elements that can be located on two different scales that are independent in the sense that the order of two elements on one scale cannot be predicted from their order on the other. A space can then be defined corresponding to any set of scales. Real space is an instantiation of this theory, and so are various idealizations of it. But metaphorical spaces are also instantiations. We can, for example, talk about salary and quality of life as different dimensions relevant to job choice.

We have concentrated more on specifying axioms than on constructing models. Thus, our approach is more syntactic than semantic, in the logical sense. Our view is that the chief role of models in our effort is for proving the consistency and independence of sets of axioms, and for showing their adequacy. Many of the spatial and temporal theories we construct are intended at least to have Euclidean space or the real numbers as one model,

but they are also intended to have discrete, finite, and less highly structured models as well.

Not only do people seem to have single theories for multiple phenomena, they also seem to have multiple theories for single phenomena. Where this is so, where for example several competing ontologies suggest themselves, we attempt to construct a theory that accommodates both. Rather than commit ourselves to adopting one set of primitives rather than another, we try to show how each set of primitives can be characterized in terms of the other. Then one need not make claims of primacy for either. Generally, each of the ontologies is useful for different purposes, and it is convenient to be able to appeal to both. Our treatment of time illustrates this. One possible approach is to take the time line as basic, and to say that events and conditions have associated time instants or intervals. In this view, there is a change in the world if an entity is in one state at one point in time and in another state at another point in time. This view is reflected in language in the clock and calendar vocabulary. Another approach, one I think corresponds better with the way we really view the world most of the time, is to say that there is a primitive relation *change* between conditions or situations, that these conditions and changes can co-occur, and that the time line is just an idealized sequence of changes that many other events co-occur with. This view seems to be deeply embedded in language, in, for example, verbs describing changes of state. Rather than be forced into one ontology or the other, we have shown how each can be defined in terms of the other.

In addition to being cavalier about the match between the core theories and the way the world really is, we are being cavalier about whether the axiomatizations fit into the classical mold of a few undefined, “primitive” predicates and a large number of predicates defined in terms of these primitives. We take it that one can rarely expect to find necessary and sufficient conditions for some concept  $p$ . There will be few axioms of the form

$$(\forall x)p(x) \equiv Q$$

The most we can hope for is to find a number of necessary conditions and a number of sufficient conditions, that is, a number of axioms of the form

$$(\forall x)p(x) \supset Q$$

and a number of axioms of the form

$$(\forall x)R \supset p(x)$$

It is generally hopeless to aim for *definitions*; the most we can expect is *characterizations*. This amounts to saying that virtually every predicate is a primitive, but a primitive that is highly interrelated with the rest of the knowledge base.

One way this can happen is illustrated by the predicate *at*. There are very few facts that one can conclude from the fact that one entity is *at* another in an arbitrary system. The predicate is used first as a way of relating many other concepts, especially concepts involving change, with each other. So there are axioms that say that when something moves from one point to another, it is no longer at the first and is now at the second. Its second use is as an entry point into spatial metaphors. There are a number of axioms of the form

$$(\forall x, y, s)p(x, y) \wedge q(y, s) \supset at(x, y, s)$$

When we see a spatial metaphor and ask what would imply such a usage, axioms like these enable an interpretation.

The predicate *cause* is another illustration of the roles of primitive predicates in the knowledge base. We do not attempt to define causality in terms of other, more basic concepts. There are a few things we know about causality in general, such as the transitivity of *cause* and the relation between *cause* and temporal order. But otherwise almost all we know about causality is particular facts about what kinds of particular events cause what other kinds of particular events. We should not expect to have a highly developed theory of causality *per se*. Rather we should expect to see causal information distributed throughout the knowledge base.

Another example of characterization rather than definition is provided by natural kind terms, like “metal”. We all know from Putnam that we can’t hope to define such terms in ways that will survive future scientific discovery. Even if we were able to define them in ways consistent with current science, the definitions would be very distant from common sense. Nevertheless, we know a great many properties of metals, and this knowledge plays a role in the interpretation of many texts we encounter. Therefore, the knowledge base contains a number of axioms encoding things like the fact that metals behave in a certain way when subjected to increasing forces.

The TACITUS project is fairly new, and we have not yet characterized a large number of words or axiomatized very many core theories. But already the range of words we have been able to handle indicates the promise of our approach. Here are some examples. The word “range” has already been discussed. Assemblies and environments are both systems of particular

kinds, and we can say that an assembly “operates” if it engages in its normative behavior in an environment. The word “immediately”, etymologically, predicates of an ordering relation between two events that a third relevant event does not occur between them. This fact can be expressed in terms provided by the core theories of scales and time. The word “brittle” can be characterized within the same theory of materials acted upon by forces that was useful in specifying some properties of metals, mentioned above. The concept “wear”, as in “worn bearings” or “a worn-out shirt”, was one of the original targets of our research effort. Wear is the cumulative small-scale loss of material from the surface of an object due to the abrasive action of some external material. We have been able to state this formally in terms of predicates from core theories of granularity, change, force, and the topology and cohesion of pieces of material. The diversity and complexity of the set of words we have been able to handle encourages us in the belief that lexical semantics should be integrated with efforts to formalize commonsense knowledge.

An old favorite question for lexical theorists is whether one can make a useful distinction between linguistic knowledge and world knowledge. The position I have articulated leads one to an answer that can be stated briefly. There is no useful distinction. In discourse comprehension and generation, both kinds of knowledge are required and, in our work so far on interpretation, both are handled in the same way. Defining or characterizing words can only be done as an adjunct to an effort to build theories useful for understanding phenomena in the world. In fact, the only reason I can imagine for maintaining such a distinction is for preserving discipline boundaries.

There is, however, a useful, related distinction in kinds of knowledge bases one might build. The knowledge base we are building is geared toward communication. There are other efforts, such as those in qualitative physics (e.g., DeKleer and Brown, 1985), which are geared toward the prediction of physical events in the absence of complete information. In such efforts, one is less concerned about metaphor and more concerned about detailed correspondence with the world. It wouldn't disturb me if with our knowledge base we failed to predict when a valve would close, but I would be disturbed if we could not cope with spatial metaphors for, say, economic information.

So far we have spent more time developing the core theories than in characterizing words in terms of them. What we have done in the latter area has primarily been for exploratory and illustrative purposes. Moreover, the entire effort is so new that frequently when we try to characterize a word we discover another core theory or two that needs to be axiomatized first. So

we have barely scratched the surface in constructing the kind of knowledge base required for genuine text processing. What hope is there for scaling up? There are two points to make here. First of all, Maurice Gross is fond of pointing out that other fields, such as astronomy and botany, have faced just as formidable a task of classification and cataloguing as we face, and have thrived on it. When we have a better idea of what we want to do, there will be people enough to do it.

Secondly, there is promise in the recent attention given to automatic processing of already existing on-line dictionaries and other knowledge sources. I can imagine that work eventually converging in a fruitful way with our research. I like to characterize the difference between the TACITUS project and recent projects aimed at encoding all the knowledge in an encyclopedia by saying that rather than encoding the knowledge in the encyclopedia, we are trying to encode the knowledge required by someone before he even opens the encyclopedia, just to be able to read it. The same holds true of a dictionary. As we build up a larger and larger knowledge base and further implement the procedures that will use this knowledge in text comprehension, we will be more and more in the position of being able to use the information in large, on-line dictionaries. Work on extracting semantic hierarchies from on-line dictionaries (Amsler, 1980; Chodorow, Byrd, and Heidorn, 1985) will not merely reveal a set of semantic primitives for some domain. These semantic primitives will be concepts that have already been explicated in core theories in the knowledge base, so that this automatic analysis will have in turn yielded more valuable results. We will have extended the knowledge base itself using these on-line resources.

## Acknowledgements

The research described here is a joint effort with William Croft, Todd Davies, Douglas Edwards, and Kenneth Laws. The opinions expressed here are, however, my own. The research is funded by the Defense Advanced Research Projects Agency under Office of Naval Research contract N00014-85-C-0013.

## References

- [1] Amsler, Robert A., 1980. *The Structure of the Merriam-Webster Pocket Dictionary*, doctoral dissertation, TR-164, University of Texas, Austin.

- [2] Chodorow, Martin S., Roy J. Byrd, and George E. Heidorn, 1985. "Extracting Semantic Hierarchies from a Large On-line Dictionary", Proceedings, 23rd Annual Meeting of the Association for Computational Linguistics, Chicago, Illinois, pp. 299-304.
- [3] De Kleer, Johann, and John Seely Brown. 1985. "A Qualitative Physics Based on Confluences." *Formal Theories of the Commonsense World*, ed. by Jerry R. Hobbs and Robert C. Moore, Ablex Publishing Corp., pp. 109-184.
- [4] Hobbs, Jerry R., William Croft, Todd Davies, Douglas Edwards, and Kenneth Laws, 1986. "Commonsense Metaphysics and Lexical Semantics", Proceedings, 24th Annual Meeting of the Association for Computational Linguistics, New York, pp. 231-240.