Commonsense Metaphysics and Lexical Semantics

Jerry R. Hobbs, William Croft, Todd Davies,

Douglas Edwards, and Kenneth Laws

Artificial Intelligence Center

SRI International

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. Of central importance in this effort is the axiomatization of what might be called "commonsense metaphysics". This includes a number of areas that figure in virtually every domain of discourse, such as granularity, scales, time, space, material, physical objects, shape, causality, functionality, and force. Our effort has been to construct core theories of each of these areas, and then to define, or at least characterize, a large number of lexical items in terms provided by the core theories. In this paper we discuss our methodological principles and describe the key ideas in the various domains we are investigating.

1. Introduction

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. Of central importance in this effort is the axiomatization of what might be called "commonsense metaphysics". This includes a number of areas that figure in virtually every domain of discourse, such as scalar notions, granularity, time, space, material, physical objects, causality, functionality, force, and shape. Our approach to lexical semantics is to construct core theories of each of these areas, and then to define, or at least characterize, a large number of lexical items in terms provided by the core theories. In the TACITUS system, processes for solving pragmatics problems posed by a text will use the knowledge base consisting of these theories, in conjunction with the logical forms of the sentences in the text, to produce an interpretation. In this paper we do not stress these interpretation processes; this is another, important aspect of the TACITUS project, and it will be described in subsequent papers (Hobbs and Martin, 1987).

This work represents a convergence of research in lexical semantics in linguistics and efforts in artificial intelligence to encode commonsense knowledge. Over the years, lexical semanticists have developed formalisms of increasing adequacy for encoding word meaning, progressing from simple sets of features (Katz and Fodor, 1963) to notations for predicate-argument structure (Lakoff, 1972; Miller and Johnson-Laird, 1976), but the early attempts still limited access to world knowledge and assumed only very restricted sorts of processing. Workers in computational linguistics introduced inference (Rieger, 1974; Schank, 1975) and other complex cognitive processes (Herskovits, 1982) into our understanding of the role of word meaning. Recently linguists have given greater attention to the cognitive processes that would operate on their representations (e.g., Talmy, 1983; Croft, 1986). Independently, in artificial intelligence an effort arose to encode large amounts of commonsense knowledge (Hayes, 1979; Hobbs and Moore, 1985; Hobbs et al. 1985). The research reported here represents a convergence of these various developments. By constructing core theories of certain fundamental phenomena and defining lexical items within these theories, using the full power of predicate calculus, we are able to cope with complexities of word meaning that have hitherto escaped lexical semanticists. Moreover, we can do this within a framework that gives full scope to the planning and reasoning processes that manipulate representations of word meaning.

0362-613X/ 87 /030241-250$03.00

In constructing the core theories we are attempting to adhere to several methodological principles:

1. One should aim for characterization of concepts, rather than definition. One cannot generally expect to find necessary and sufficient conditions for a concept. The most we can hope for is to find a number of necessary conditions and a number of sufficient conditions. This amounts to saying that a great many predicates are primitives, but they are primitives that are highly interrelated with the rest of the knowledge base.

2. One should determine the minimal structure necessary for a concept to make sense. In efforts to axiomatize an area, there are two positions one may take, exemplified by set theory and by group theory. In axiomatizing set theory, one attempts to capture exactly some concept that 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 it turns out that there are unexpected models, this is a serendipitous discovery of a new phenomenon that we can reason about using an old theory. The pervasive character 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. By seeking minimal structures in axiomatizing concepts, we optimize the possibilities of using the theories in metaphorical and analogical contexts. This principle is illustrated below in the section on regions. One consequence of this principle is that our approach will seem more syntactic than semantic. We have concentrated more on specifying axioms than on constructing models. 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. As an example of the last point, many of the spatial and temporal theories we construct are intended at least to have Euclidean space or the real numbers as one model, and a subclass of graph-theoretical structures as other models.

3. A balance must be struck between attempting to cover all cases and aiming only for the prototypical cases. In general, we have tried to cover as many cases as possible with an elegant axiomatization, in line with the two previous principles, but where the formalization begins to look baroque, we assume that higher processes will block some inferences in the marginal cases. We assume that inferences will be drawn in a controlled fashion. Thus, every outré, highly context-dependent counterexample need not be accounted for, and to a certain extent, definitions can be geared specifically to a prototype.

4. Where competing ontologies suggest themselves in a domain, one should try to construct a theory that accommodates both. Rather than commit oneself to adopting one set of primitives rather than another, one should show how either set can be characterized in terms of the other. 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.

5. The theories one constructs should be richer in axioms than in theorems. In mathematics, one expects to state half a dozen axioms and prove dozens of theorems from them. In encoding commonsense knowledge, it seems to be just the opposite. The theorems we seek to prove on the basis of these axioms are theorems about specific situations that are to be interpreted, in particular, theorems about a text that the system is attempting to understand.

6. One should avoid falling into "black holes". There are a few "mysterious" concepts that crop up repeatedly in the formalization of commonsense metaphysics. Among these are "relevant" (that is, relevant to the task at hand) and "normative" (that is, conforming to some norm or pattern). To insist upon giving a satisfactory analysis of these before using them in analyzing other concepts is to cross the event horizon that separates lexical semantics from philosophy. On the other hand, our experience suggests that to avoid their use entirely is crippling; the lexical semantics of a wide variety of other terms depends upon them. Instead, we have decided to leave them minimally analyzed for the moment and use them without scruple in the analysis of other commonsense concepts. This approach will allow us to accumulate many examples of the use of these mysterious concepts, and in the end, contribute to their successful analysis. The use of these concepts appears below in the discussions of the words "immediately", "sample", and "operate".

We chose as an initial target the problem of encoding the commonsense knowledge that underlies the concept of "wear", as in a part of a device wearing out. Our aim was to define "wear" in terms of predicates characterized elsewhere in the knowledge base and to be able to infer some consequences of wear. For something to wear, we decided, is for it to lose imperceptible bits of material from its surface due to abrasive action over time. One goal, which we have not yet achieved, is to be able to prove as a theorem that, since the shape of a part of a mechanical device is often functional and since loss of material can result in a change of shape, wear of a part of a device can cause the failure of the device as a whole. In addition, as we have proceeded, we have characterized a number of words found in a set of target texts, as it has become possible.

We are encoding the knowledge as axioms in what is for the most part a first-order logic, described by Hobbs (1985a), although quantification over predicates is sometimes convenient. In the formalism there is a nominalization operator " ′ " for reifying events and conditions, as expressed in the following axiom schema:

$$(\forall x)p(x) \equiv (\exists e)p'(e,x) \wedge Exist(e)$$

That is, p is true of x if and only if there is a condition e of p's being true of x and e exists in the real world.

In our implementation so far, we have been proving simple theorems from our axioms using the CG5 theorem-prover developed by Mark Stickel (1982), and we are now beginning to use the knowledge base in text processing.

2 Requirements on Arguments of Predicates

There is a notational convention used below that deserves some explanation. It has frequently been noted that relational words in natural language can take only certain types of words as their arguments. These are usually described as selectional constraints. The same is true of predicates in our knowledge base. The constraints are expressed below by rules of the form

$$p(x,y) : r(x,y)$$

This means that for p even to make sense applied to x and y, it must be the case that r is true of x and y. The logical import of this rule is that wherever there is an axiom of the form

$$(\forall\, x,y) p(x,y) \supset q(x,y)$$

this is really to be read as

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

The checking of selectional constraints, therefore, emerges as a by-product of other logical operations: the constraint $r(x,y)$ must be verified if anything else is to be proved from $p(x,y)$.

The simplest example of such an $r(x,y)$ is a conjunction of sort constraints $r_1(x) \wedge r_2(y)$. Our approach is a generalization of this, because much more complex requirements can be placed on the arguments. Consider, for example, the verb "range". If x ranges from y to z, there must be a scale s that includes y and z, and x must be a set of entities that are located at various places on the scale. This can be represented as follows:

$$range(x,y,z) : (\exists\, s)\,[scale(s) \wedge y \in s \wedge z \in s \wedge set(x) \wedge (\forall\, u)[u \in x \supset (\exists\, v)\, v \in s \wedge at(u,v)]]$$

3 The Knowledge Base

3.1 Sets and Granularity

At the foundation of the knowledge base is an axiomatization of set theory. It follows the standard Zermelo-Fraenkel approach, except that there is no axiom of infinity.

Since so many concepts used in discourse are grain-dependent, a theory of granularity is also fundamental (see Hobbs 1985b). A grain is defined in terms of an indistinguishability relation, which is reflexive and symmetric, but not necessarily transitive. One grain can be a *refinement* of another, with the obvious definition. The most refined grain is the identity grain, i.e., the one in which every two distinct elements are distinguishable. One possible relationship between two grains, one of which is a refinement of the other, is what we call an "Archimedean relation", after the Archimedean property of real numbers. Intuitively, if enough events occur that are imperceptible at the coarser grain g_2 but perceptible at the finer grain g_1, the aggregate will eventually be perceptible at the coarser grain. This is an important property in phenomena subject to the heap paradox. Wear, for instance, eventually has significant consequences.

3.2 Scales

A great many of the most common words in English have scales as their subject matter. This includes many prepositions, the most common adverbs, comparatives, and many abstract verbs. When spatial vocabulary is used metaphorically, it is generally the scalar aspect of space that carries over to the target domain. A scale is defined as a set of elements, together with a partial ordering and a granularity (or an indistinguishability relation). The partial ordering and the indistinguishability relation are consistent with each other:

$$(\forall\, x,y,z)\; x < y \wedge y \sim z \supset x < z \vee x \sim z$$

That is, if x is less than y and y is indistinguishable from z, then either x is less than z or x is indistinguishable from z.

It is useful to have an adjacency relation between points on a scale, and there are a number of ways we could introduce it. We could simply take it to be primitive; in a scale having a distance function, we could define two points to be adjacent when the distance between them is less than some ϵ; finally, we could define adjacency in terms of the grain size for the scale:

$$(\forall\, x,y,s)\; adj(x,y,s) \equiv (\exists z)\; z \sim_s x \wedge z \sim_s y \wedge \neg [x \sim_s y],$$

That is, distinguishable elements x and y are adjacent on scale s if and only if there is an element z which is indistinguishable from both.

Two important possible properties of scales are connectedness and denseness. We can say that two elements of a scale are connected by a chain of adj relations:

$$(\forall x,y,s) connected(x,y,s) \equiv adj(x,y,s) \vee (\exists z) adj(x,z,s) \wedge connected(z,y,s)$$

A scale is connected (*sconnected*) if all pairs of elements are connected. A scale is dense if between any two points there is a third point, until the two points are so close together that the grain size no longer allows us to determine whether such an intermediate point exists. Cranking up the magnification could well resolve the continuous space into a discrete set, as objects into atoms.

$$(\forall s) dense(s) \equiv (\forall x,y) x \in s \wedge y \in s \wedge x <_s y \supset (\exists z)(x <_s z \wedge z <_s y) \vee (\exists z)(x \sim_s z \wedge z \sim_s y)$$

This expresses the commonsense notion of continuity.

A subscale of a scale has as its elements a subset of the elements of the scale and has as its partial ordering and its grain the partial ordering and the grain of the scale.

$$(\forall s_1,s_2)subscale(s_2,s_1) \equiv \text{subset}(s_2,s_1)$$
$$\wedge (\forall x,y)[[x <_{s1} y \equiv x <_{s2} y] \wedge [x \sim_{s1} y \equiv x \sim_{s2} y]]$$

An interval can be defined as a connected subscale:

$$(\forall i)interval(i) \equiv (\exists s)scale(s)$$
$$\wedge\ subscale(i,s) \wedge sconnected(i)$$

The relations between time intervals that Allen and Kautz (1985) have defined can be defined in a straightforward manner in the approach presented here, but for intervals in general.

A concept closely related to scales is that of a "cycle". This is a system that has a natural ordering locally but contains a loop globally. Examples are the color wheel, clock times, and geographical locations ordered by "east of". We have axiomatized cycles in terms of a ternary *between* relation whose axioms parallel those for a partial ordering.

The figure-ground relationship is of fundamental importance in language. We encode it with the primitive predicate *at*. It is possible that the minimal structure necessary for something to be a ground is that of a scale; hence, this is a selectional constraint on the arguments of *at*.[1]

$$at(x,y) : (\exists s)y \in s \wedge scale(s)$$

At this point, we are already in a position to define some fairly complex words. As an illustration, we give the example of "range" as in "x ranges from y to z":

$$(\forall x,y,z)range(x,y,z) \equiv$$
$$(\exists s,s_1,u_1,u_2)scale(s) \wedge subscale(s_1,s)$$
$$\wedge\ bottom(y,s_1) \wedge top(z,s_1)$$
$$\wedge u_1 \in x \wedge at(u_1,y) \wedge u_2 \in x \wedge at(u_2,z)$$
$$\wedge (\forall u)[u \in x \supset (\exists v)v \in s_1 \wedge at(u,v)]$$

That is, x ranges from y to z if and only if y and z are the bottom and top of a subscale s_1 of some scale s and x is a set which has elements at y and z and all of whose elements are located at points on s_1.

A very important scale is the linearly ordered scale of numbers. We do not plan to reason axiomatically about numbers, but it is useful in natural language processing to have encoded a few facts about numbers. For example, a set has a cardinality which is an element of the number scale.

Verticality is a concept that would most properly be analyzed in the section on space, but it is a property that many other scales have acquired metaphorically, for whatever reason. The number scale is one of these. Even in the absence of an analysis of verticality, it is a useful property to have as a primitive in lexical semantics.

The word "high" is a vague term asserting that an entity is in the upper region of some scale. It requires that the scale be a *vertical* one, such as the number scale. The verticality requirement distinguishes "high" from the more general term "very"; we can say "very hard" but not "highly hard". The phrase "highly planar" sounds all right because the high register of "planar" suggests a quantifiable, scientific accuracy, whereas the low register of "flat" makes "highly flat" sound much worse.

The test of any definition is whether it allows one to draw the appropriate inferences. In our target texts, the phrase "high usage" occurs. Usage is a set of using events, and the verticality requirement on "high" forces us to coerce the phrase into "a high or large number of using events". Combining this with an axiom stating that the use of a mechanical device involves the likelihood of abrasive events, as defined below, and with the definition of "wear" in terms of abrasive events, we should be able to conclude the likelihood of wear.

3.3 Time: Two Ontologies

There are two possible ontologies for time. In the first, the one most acceptable to the mathematically minded, there is a time line, which is a scale having some topological structure. We can stipulate the time line to be linearly ordered (although it is not in approaches that build ignorance of relative times into the representation of time (e.g., Hobbs, 1974) nor in approaches employing branching futures (e.g., McDermott, 1985)), and we can stipulate it to be dense (although it is not in the situation calculus). We take *before* to be the ordering on the time line:

$$(\forall t_1,t_2)before(t_1,t_2) \equiv$$
$$(\exists T)Time\text{-}line(T) \wedge t_1 \in T \wedge t_2 \in T \wedge t_1 <_T t_2$$

We allow both instants and intervals of time. Most events occur at some instant or during some interval. In this approach, nearly every predicate takes a time argument.

In the second ontology, the one that seems to be more deeply rooted in language, the world consists of a large number of more or less independent processes, or histories, or sequences of events. There is a primitive relation *change* between conditions. Thus,

$$change(e_1,e_2) \wedge p'(e_1,x) \wedge q'(e_2,x)$$

says that there is a change from the condition e_1 of p's being true of x to the condition e_2 of q's being true of x.

The time line in this ontology is then an artificial construct, a regular sequence of imagined abstract events (think of them as ticks of a clock in the National Bureau of Standards) to which other events can be related. The change ontology seems to correspond to the way we experience the world. We recognize relations of causality, change of state, and copresence

[1] However, we are currently examining an approach in which a more abstract concept, "system", discussed in Section 3.6.3, is taken to be the minimal structure for expressing location.

among events and conditions. When events are not related in these ways, judgments of relative time must be mediated by copresence relations between the events and events on a clock and change of state relations on the clock.

The predicate *change* possesses a limited transitivity. There has been a change from Reagan's being an actor to Reagan's being president, even though he was governor in between. But we probably do not want to say there has been a change from Reagan's being an actor to Margaret Thatcher's being prime minister, even though the second event comes after the first.

In this ontology, we can say that any two times, viewed as events, always have a *change* relation between them.

$$(\forall t_1,t_2) before(t_1,t_2) \supset change(t_1,t_2)$$

The predicate *change* is related to before by the axiom

$$(\forall e_1,e_2)\ change(e_1,e_2) \supset (\exists t_1,t_2)\ at(e_1,t_1) \wedge at(e_2,t_2) \wedge before(t_1,t_2)$$

That is, if there is a change from e_1 to e_2, then there is a time t_1 at which e_1 occurred and a time t_2 at which e_2 occurred, and t_1 is before t_2. This does not allow us to derive change of state from temporal succession. For this, we would need axioms of the form

$$(\forall e_1,e_2,t_1,t_2,x)\ p'(e_1,x) \wedge at(e_1,t_1) \wedge q'(e_2,x) \wedge at(e_2,t_2) \wedge before(t_1,t_2) \supset change(e_1,e_2)$$

That is, if x is p at time t_1 and q at a later time t_2, then there has been a change of state from one to the other. This axiom would not necessarily be true for all p's and q's. Time arguments in predications can be viewed as abbreviations:

$$(\forall\ x,t)p(x,t) \equiv (\exists e)p'(e,x) \wedge at(e,t)$$

The word "move", or the predicate *move*, (as in "x moves from y to z") can then be defined equivalently in terms of change,

$$(\forall\ x,y,z)move(x,y,z) \equiv (\exists e_1,e_2)\ change(e_1,e_2) \wedge at'(e_1,x,y) \wedge at'(e_2,x,z)$$

or in terms of the time line,

$$(\forall\ x,y,z)\ move(x,y,z) \equiv (\exists t_1,t_2)\ at(x,y,t_1) \wedge at(x,z,t_2) \wedge before(t_1,t_2)$$

(The latter definition has to be complicated a bit to accommodate cyclic motion. The former axiom is all right as it stands, provided there is also an axiom saying that for there to be a change from a state to the same state, there must be an intermediate different state.)

In English and apparently all other natural languages, both ontologies are represented in the lexicon. The time line ontology is found in clock and calendar terms, tense systems of verbs, and in the deictic temporal locatives such as "yesterday", "today", "tomorrow", "last night", and so on. The change ontology is exhibited in most verbs, and in temporal clausal connectives. The universal presence in natural languages of both classes of lexical items and grammatical markers requires a theory that can accommodate both ontologies, illustrating the importance of methodological principle 4.

Among temporal connectives, the word "while" presents interesting problems. In "e_1 while e_2", e_2 must be an event occurring over a time interval; e_1 must be an event and may occur either at a point or over an interval. One's first guess is that the point or interval for e_1 must be included in the interval for e_2. However, there are cases, such as

> The electricity should be off while the switch is being repaired.

which suggest the reading "e_2 is included in e_1". We came to the conclusion that one can infer no more than that e_1 and e_2 overlap, and any tighter constraints result from implicatures from background knowledge.

The word "immediately", as in "immediately after the alarm", also presents a number of problems. It requires its argument e to be an ordering relation between two entities x and y on some scale s.

$$immediate(e) : (\exists x,y,s) less\text{-}than'(e,x,y,s)$$

It is not clear what the constraints on the scale are. Temporal and spatial scales are acceptable, as in "immediately after the alarm" and "immediately to the left", but the size scale is not:

> * John is immediately larger than Bill.

Etymologically, it means that there are no intermediate entities between x and y on s. Thus,

$$(\forall e,x,y,s)\ immediate(e) \wedge less\text{-}than'(e,x,y,s) \supset \neg\ (\exists z) less\text{-}than(x,z,s) \wedge less\text{-}than(z,y,s)$$

However, this will only work if we restrict z to be a relevant entity. For example, in the sentence

> We disengaged the compressor immediately after the alarm.

the implication is that no event that could damage the compressor occurred between the alarm and the disengagement, since the text is about equipment failure.

3.4 SPACES AND DIMENSION: THE MINIMAL STRUCTURE

The notion of dimension has been made precise in linear algebra. Since the concept of a region is used metaphorically as well as in the spatial sense, however, we were concerned to determine the *minimal* structure a system requires for it to make sense to call it a space of more than one dimension. For a two-dimensional space, there must be a scale, or partial ordering, for each dimension. Moreover, the two scales must be independent, in that the order of elements on one scale can not be determined from their order on the other. Formally,

$$(\forall\ sp)space(sp) \equiv (\exists s_1,s_2)\ scale_1(s_1,sp) \wedge scale_2(s_2,sp)$$

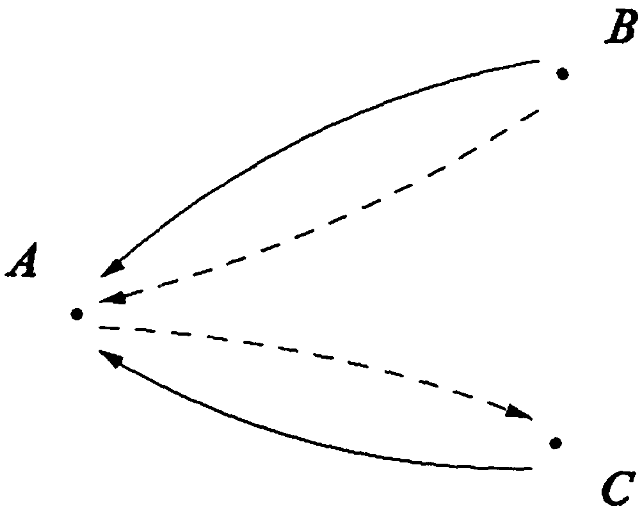

Figure 1.1 The Simplest Space.

$$\wedge (\exists x)[(\exists y_1)\ [x <_{s_1} y_1 \wedge x <_{s_2} y_1]$$
$$\wedge (\exists y_2)[x <_{s_1} \wedge y_2 \wedge y_2 <_{s_2} x]]$$

Note that this does not allow $<_{s_2}$ to be simply the reverse of $<_{s_1}$. An unsurprising consequence of this definition is that the minimal example of a two-dimensional space consists of three points (three points determine a plane), e.g., the points A, B, and C, where

$$A <_1 B,\ A <_1 C,\ C <_2 A,\ A <_2 B.$$

This is illustrated in Figure 1.

The dimensional scales are apparently found in all natural languages in relevant domains. The familiar three-dimensional space of common sense can be defined by the three scale pairs "up-down", "front-back", and "left-right"; the two-dimensional plane of the commonsense conception of the earth's surface is represented by the two scale pairs "north-south" and "east-west".

The simplest, although not the only, way to define adjacency in the space is as adjacency on both scales:

$$(\forall x,y,sp)adj(x,y,sp) \equiv$$
$$(\exists s_1,s_2)\ scale_1(s_1,sp) \wedge scale_2(s_2,sp)$$
$$\wedge\ adj(x,y,s_1) \wedge adj(x,y,s_2)$$

A region is a subset of a space. The surface and interior of a region can be defined in terms of adjacency, in a manner paralleling the definition of a boundary in point-set topology. In the following, s is the boundary or surface of a two- or three-dimensional region r embedded in a space sp.

$$(\forall\ s,r,sp)surface(s,r,sp) \equiv$$
$$(\forall\ x)x \in r \supset [x \in s \equiv$$
$$(Ey)(y \in sp \wedge \neg\ (y \in r) \wedge adj(x,y,sp))]$$

Finally, we can define the notion of "contact" in terms of points in different regions being adjacent:

$$(\forall\ r_1,r_2,sp)contact(r_1,\ r_2,sp) \equiv$$
$$disjoint(r_1,r_2) \wedge (\exists\ x,y)(x \in r_1 \wedge y \in r_2 \wedge adj(x,y,sp))$$

By picking the scales and defining adjacency right, we can talk about points of contact between communication networks, systems of knowledge, and other metaphorical domains. By picking the scales to be the real line and defining adjacency in terms of ϵ-neighborhoods, we get Euclidean space and can talk about contact between physical objects.

3.5 MATERIAL

Physical objects and materials must be distinguished, just as they are in apparently every natural language, by means of the count noun-mass noun distinction. A physical object is not a bit of material, but rather is composed of a bit of material at any given time. Thus, rivers and human bodies are physical objects, even though their material constitution changes over time. This distinction also allows us to talk about an object's losing material through wear and still remaining the same object.

We will say that an entity b is a bit of material by means of the expression *material(b)*. Bits of material are characterized by both extension and cohesion. The primitive predication *occupies(b,r,t)* encodes extension, saying that a bit of material b occupies a region r at time t. The topology of a bit of material is then parasitic on the topology of the region it occupies. A *part* b_1 of a bit of material b is a bit of material whose occupied region is always a subregion of the region occupied by b. Point-like particles (*particle)* are defined in terms of points in the occupied region, disjoint bits (*disjointbit)* in terms of the disjointness of regions, and contact between bits in terms of contact between their regions. We can then state as follows the principle of non-joint-occupancy that two bits of material cannot occupy the same place at the same time:

$$(\forall\ b_1,\ b_2)(disjointbit(b_1,b_2)$$
$$\supset(\forall\ x,y,b_3,b_4)\ interior(b_3,b_1) \wedge interior(b_4,b_2)$$
$$\wedge\ particle(x,b_3) \wedge particle(y,b_4)$$
$$\supset \neg\ (\exists z)(at(x,\ z) \wedge at(y,\ z))$$

That is, if bits b_1 and b_2 are disjoint, then there is no entity z that is at interior points in both b_1 and b_2. At some future point in our work, this may emerge as a consequence of a richer theory of cohesion and force.

The cohesion of materials is also a primitive property, for we must distinguish between a bump on the surface of an object and a chip merely lying on the surface. Cohesion depends on a primitive relation *bond* between particles of material, paralleling the role of *adj* in regions. The relation *attached* is defined as the transitive closure of *bond*. A topology of cohesion is built up in a manner analogous to the topology of regions. In addition, we have encoded the relation that *bond* bears to motion, i.e., that bonded bits remain adjacent and that one moves when the other does, and the relation of bond to force, i.e. that there is a characteristic force that breaks a bond in a given material.

Different materials react in different ways to forces of various strengths. Materials subjected to force exhibit or fail to exhibit several invariance properties, proposed by Hager (1985). If the material is shape-invariant with respect to a particular force, its shape remains the same. If it is topologically invariant, particles that are adjacent remain adjacent. Shape invariance implies topological invariance. If subjected to forces of a certain strength or degree d_1, a material ceases being shape-invariant. At a force of strength $d_2 \geq d_1$, it ceases being topologically invariant, and at a force of strength $d_3 \geq d_2$, it simply breaks. Metals exhibit the full range of possibilities, that is, $0 < d_1 < d_2 < d_3 < \infty$. For forces of strength $d < d_1$, the material is "hard"; for forces of strength d where $d_1 < d < d_2$, it is "flexible"; for forces of strength d where $d_2 < d < d_3$, it is "malleable". Words such as "ductile" and "elastic" can be defined in terms of this vocabulary, together with predicates about the geometry of the bit of material. Words such as "brittle" ($d_1 = d_2 = d_3$) and "fluid" ($d_2 = 0$, $d_3 = \infty$) can also be defined in these terms. While we should not expect to be able to *define* various material terms, like "metal" and "ceramic", we can certainly characterize many of their properties with this vocabulary.

Because of its invariance properties, material interacts with containment and motion. The word "clog" illustrates this. The predicate *clog* is a three-place relation: x clogs y against the flow of z. It is the obstruction by x of z's motion through y, but with the selectional restriction that z must be something that can flow, such as a liquid, gas, or powder. If a rope is passing through a hole in a board, and a knot in the rope prevents it from going through, we do not say that the hole is clogged. On the other hand, there do not seem to be any selectional constraints on x. In particular, x can be identical with z: glue, sand, or molasses can clog a passageway against its own flow. We can speak of clogging where the obstruction of flow is not complete, but it must be thought of as "nearly" complete.

3.6 OTHER DOMAINS

3.6.1 CAUSAL CONNECTION

Attachment within materials is one variety of causal connection. In general, if two entities x and y are causally connected with respect to some behavior p of x, then whenever p happens to x, there is some corresponding behavior q that happens to y. In the case of attachment, p and q are both *move*. A particularly common kind of causal connection between two entities is one mediated by the motion of a third entity from one to the other. (This might be called a "vector boson" connection.) Photons mediating the connection between the sun and our eyes, raindrops connecting a state of the clouds with the wetness of our skin and clothes, a virus being transmitted from one person to another, and utterances passing between people are all examples of such causal connections. Barriers, openings, and penetration are all defined with respect to paths of causal connection.

3.6.2 FORCE

The concept of "force" is axiomatized, in a way consistent with Talmy's treatment (1985), in terms of the predications *force*(a,b,d_1) and *resist*(b,a,d_2) — a forces against b with strength d_1 and b resists a's action with strength d_2. We can infer motion from facts about relative strength. This treatment can also be specialized to Newtonian force, where we have not merely movement, but acceleration. In addition, in spaces in which orientation is defined, forces can have an orientation, and a version of the "parallelogram of forces" law can be encoded. Finally, force interacts with shape in ways characterized by words like "stretch", "compress", "bend", "twist", and "shear".

3.6.3 SYSTEMS AND FUNCTIONALITY

An important concept is the notion of a "system", which is a set of entities, a set of their properties, and a set of relations among them. A common kind of system is one in which the entities are events and conditions and the relations are causal and enabling relations. A mechanical device can be described as such a system — in a sense, in terms of the plan it executes in its operation. The *function* of various parts and of conditions of those parts is then the role they play in this system, or plan.

The intransitive sense of "operate", as in

The diesel was operating.

involves systems and functionality. If an entity x operates, there must be a larger system s of which x is a part. The entity x itself is a system with parts. These parts undergo normative state changes, thereby causing x to undergo normative state changes, thereby causing x to produce an effect with a normative function in the larger system s. The concept of "normative" is discussed below.

3.6.4 SHAPE

We have been approaching the problem of characterizing shape from a number of different angles. The classical treatment of shape is via the notion of "similarity" in Euclidean geometry, and in Hilbert's formal reconstruction of Euclidean geometry (Hilbert, 1902) the key primitive concept seems to be that of "congruent angles". Therefore, we first sought to develop a theory of "orientation". The shape of an object can then be characterized in terms of changes in orientation of a tangent as one moves about on the surface of the object, as is done in some vision research (e.g., Zahn and Roskies, 1972). In all of this, since "shape" can be used loosely and metaphorically, one question we are asking is whether some minimal, abstract structure can be found in which the notion of "shape" makes sense. Consider, for instance, a graph in which one scale is discrete, or even unordered. Accordingly, we have been examining a number of examples, asking when it seems right to say two structures have different shapes.

We have also examined the interactions of shape and functionality (see Davis, 1984). What seems to be crucial is how the shape of an obstacle constrains the motion of a substance or of an object of a particular shape (see Shoham, 1985). Thus, a funnel concentrates the flow of a liquid, and similarly, a wedge concentrates force. A box pushed against a ridge in the floor will topple, and a rotating wheel is a limiting case of continuous toppling.

3.7 HITTING, ABRASION, WEAR, AND RELATED CONCEPTS

For x to hit y is for x to move into contact with y with some force.

The basic scenario for an abrasive event is that there is an impinging bit of material m that hits an object o and by doing so removes a pointlike bit of material b_0 from the surface of o:

$$abr\text{-}event'(e,m,o,b_0) : material(m)$$
$$\wedge\ (\forall t)\ at(e,t) \supset topologically\text{-}invariant(o,t)$$

$$(\forall\ e,m,o,b_0) abr\text{-}event'(e,m,o,b_0) \equiv$$
$$(\exists\ t,b,s,e_1,e_2,e_3)\ at(e,t) \wedge consists\text{-}of(o,b,t)$$
$$\wedge\ surface(s,b) \wedge \text{particle}(b_0,s) \wedge change'(e,e_1,e_2)$$
$$\wedge\ attached'(e_1,b_0,b) \wedge not'(e_2,e_1) \wedge cause(e_3,e)$$
$$\wedge\ hit'(e_3,m,b_0)$$

That is, e is an abrasive event of a material m impinging on a topologically invariant object o and detaching b_0 if and only if b_0 is a particle of the surface s of the bit of material material b of which o consists at the time t at which e occurs, and e is a change from the condition e_1 of b_0's being attached to b to the negation e_2 of that condition, where the change is caused by the hitting e_3 of m against b_0.

After the abrasive event, the pointlike bit b_0 is no longer a part of the object o:

$$(\forall e,m,o,b_0,e_1,e_2,t_2) abr\text{-}event'(e,m,o,b_0)$$
$$\wedge\ change'(e,e_1,e_2) \wedge at(e_2,t_2)$$
$$\wedge\ consists\text{-}of(o,b_2,t_2)$$
$$\supset \neg\ part(b_0,b_2)$$

That is, if e is an abrasive event of m impinging against o and detaching b_0, and e is a change from e_1 to e_2, and e_2 holds at time t_2, then b_0 is not part of the bit of material b_2 of which o consists at t_2. It is necessary to state this explicitly since objects and bits of material can be discontinuous.

An abrasion is a large set of abrasive events widely distributed through some nonpointlike region on the surface of an object:

$$(\forall\ e,m,o)\ abrade'(e,m,o) \equiv$$
$$(\exists\ bs) large(bs)$$
$$\wedge\ [(\forall e_1)[e_1 \in e \supset (\exists\ b_0) b_0 \in bs \wedge abr\text{-}event'(e_1,m,o,b_0)]$$
$$\wedge\ (\forall b,s,t)[at(e,t) \wedge consists\text{-}of(o,b,t) \wedge\ surface(s,b)$$
$$\supset (\exists r)\ subregion(r,s) \wedge widely\text{-}distributed(bs,r)]]$$

That is, e is an abrasion by m of o if and only if there is a large set bs of bits of material and e is a set of abrasive events in which m impinges on o and removes a bit b_0, an element in bs, from o, and if e occurs at time t and o consists of material b at time t, then there is a subregion r of the surface s of b over which bs is widely distributed.

Wear can result from a large collection of abrasive events distributed over time as well as space (so that there may be no instant at which enough abrasive events occur to count as an abrasion). Thus, the link between wear and abrasion is via the common notion of abrasive events, not via a definition of wear in terms of abrasion.

$$(\forall\ e,m,o)\ wear'(e,m,o) \equiv$$
$$(\exists bs)\ large(bs)$$
$$\wedge\ [(\forall e_1)[e_1 \in e$$
$$\supset (\exists\ b_0) b_0 \in bs \wedge abr\text{-}event'(e_1,m,o,b_0)]$$
$$\wedge\ (\exists i)[interval(i) \wedge widely\text{-}distributed(e,i)]]$$

That is, e is a wearing by x of o if and only if there is a large set bs of bits of material and e is a set of abrasive events in which m impinges on o and removes a bit b_0, an element in bs, from o, and e is widely distributed over some time interval i.

We have not yet characterized the concept "large", but we anticipate that it would be similar to "high". The concept "widely distributed" concerns systems. If x is distributed in y, then y is a system and x is a set of entities which are located at components of y. For the distribution to be wide, most of the elements of a partition of y, determined independently of the distribution, must contain components which have elements of x at them.

The word "wear" is one of a large class of other events involving cumulative, gradual loss of material — events described by words like "chip", "corrode", "file", "erode", "sand", "grind", "weather", "rust", "tarnish", "eat away", "rot", and "decay". All of these lexical items can now be defined as variations on the definition of "wear", since we have built up the axiomatizations underlying "wear". We are now in a position to characterize the entire class. We will illustrate this by defining two different types of variants of "wear" — "chip" and "corrode".

"Chip" differs from "wear" in three ways: the bit of material removed in one abrasive event is larger (it need not be point-like), it need not happen because of a material hitting against the object, and "chip" does not require (though it does permit) a large collection of such events: one can say that some object is chipped even if there is one chip in it. Thus, we slightly alter the definition of *abr-event* to accommodate these changes:

$$(\forall\ e,m,o,b_0) chip'(e,m,o,b_0) \equiv$$
$$(\exists t,b,s,e_1,e_2,e_3) at(e,t) \wedge consists\text{-}of(o,b,t)$$

$\wedge$ *surface(s,b)* $\wedge$ *part*(b_0,s) $\wedge$ *change'*(e,e_1,e_2)

$\wedge$ *attached'*(e_1,b_0,b) $\wedge$ *not'*(e_2,e_1)

That is, e is a chipping event by a material m of a bit of material b_0 from an object o if and only if b_0 is a part of the surface s of the bit of material material b of which o consists at the time t at which e occurs, and e is a change from the condition e_1 of b_0's being attached to b to the negation e_2 of that condition.

"Corrode" differs from "wear" in that the bit of material is chemically transformed as well as being detached by the contact event; in fact, in some way the chemical transformation causes the detachment. This can be captured by adding a condition to the abrasive event that renders it a (single) corrode event:

corrode-event(m,o,b_0) : *fluid(m)*

$\wedge$ *contact*(m,b_0)

$(\forall\ e,m,o,b_0)$ *corrode-event'*$(e,m,o,b_0) \equiv$

$(\exists\ t,b,s,e_1,e_2,e_3)$ *at(e,t)* $\wedge$ *consists-of(o,b,t)*

$\wedge$ *surface(s,b)* $\wedge$ *particle*(b_0,s) $\wedge$ *change'*(e,e_1,e_2)

$\wedge$ *attached'*(e_1,b_0,b) $\wedge$ *not'*(e_2,e_1) $\wedge$ *cause*(e_3,e)

$\wedge$ *chemical-change'*(e_3,m,b_0)

That is, e is a corrosive event by a fluid m of a bit of material b_0 with which it is in contact if and only if b_0 is a particle of the surface s of the bit of material b of which o consists at the time t at which e occurs, and e is a change from the condition e_1 of b_0's being attached to b to the negation e_2 of that condition, where the change is caused by a chemical reaction e_3 of m with b_0.

"Corrode" itself may be defined in a parallel fashion to "wear", by substituting *corrode-event* for *abr-event*.

All of this suggests the generalization that abrasive events, chipping and corrode events all detach the bit in question, and that we may describe all of these as detaching events. We can then generalize the above axiom about abrasive events that result in loss of material to the following axiom about detaching:

$(\forall\ e,m,o,b_0,e_1,e_2,t_2)$ *detach'*(e,m,o,b_0)

$\wedge$ *change'*(e,e_1,e_2) $\wedge$ *at*(e_2,t_2) $\wedge$ *consists-of*(o,t_2,b_2)

$\supset \neg$ *part*(b_0,b_2)

That is, if e is a detaching event by m of b_0 from o, and e is a change from e_1 to e_2, and e_2 holds at time t_2, then b_0 is not part of the bit of material b_2 of which o consists at t_2.

4 Relevance and the Normative

Many of the concepts we are investigating have driven us inexorably to the problems of what is meant by "relevant" and by "normative". We do not pretend to have solved these problems. But for each of these concepts we do have the beginnings of an account that can play a role in analysis, if not yet in implementation.

Our view of relevance, briefly stated, is that something is relevant to some goal if it is a part of a plan to achieve that goal. (A formal treatment of a similar view is given in Davies, forthcoming.) We can illustrate this with an example involving the word "sample". If a bit of material x is a sample of another bit of material y, then x is a part of y, and moreover, there are *relevant* properties p and q such that it is believed that if p is true of x then q is true of y. That is, looking at the properties of the sample tells us something important about the properties of the whole. Frequently, p and q are the same property. In our target texts, the following sentence occurs:

We retained an oil sample for future inspection.

The oil in the sample is a part of the total lube oil in the lube oil system, and it is believed that a property of the sample, such as "contaminated with metal particles", will be true of all the lube oil as well, and that this will provide information about possible wear on the bearings. It is therefore relevant to the goal of maintaining the machinery in good working order.

We have arrived at the following provisional account of what it means to be "normative". For an entity to exhibit a normative condition or behavior, it must first of all be a component of a larger system. This system has structure in the form of relations among its components. A pattern is a property of the system, namely, the property of a subset of these stuctural relations holding. A norm is a pattern established either by conventional stipulation or by statistical regularity. An entity behaves in a normative fashion if it is a component of a system and instantiates a norm within that system. The word "operate", discussed in Section 3.6.3, illustrates this. When we say that an engine is operating, we have in mind a larger system — i.e., the device the engine drives — to which the engine may bear various possible relations. A subset of these relations is stipulated to be the norm — the way it is supposed to work. We say it is operating when it is instantiating this norm.

5 Conclusion

The research we have been engaged in has forced us to explicate a complex set of commonsense concepts. Since we have done it in as general a fashion as possible, we expect to be able, building on this foundation, to axiomatize a large number of other areas, including areas unrelated to mechanical devices. The very fact that we have been able to characterize words as diverse as "range", "immediately", "brittle", "operate", and "wear" shows the promising nature of this approach.

Acknowledgements

The research reported here was funded by the Defense Advanced Research Projects Agency under Office of Naval Research contract N00014–85-C-0013. It builds on work supported by NIH Grant LM03611 from the National Library of Medicine, by Grant IST-8209346 from the National Science Foundation, and by a gift from the Systems Development Foundation.

References

Allen, James F., and Henry A. Kautz. 1985. A Model of Naive Temporal Reasoning. In: Jerry R. Hobbs and Robert C. Moore, Eds., *Formal Theories of the Commonsense World,* Ablex Publishing Corp., Norwood, New Jersey: 251-268.

Croft, William. 1986. *Categories and Relations in Syntax: The Clause-Level Organization of Information.* Ph.D. dissertation, Department of Linguistics, Stanford University, Stanford, California.

Davies, Todd R. Forthcoming. Determination Rules for Generalization and Analogical Inference. In: David H. Helman, Ed., *Analogical Reasoning.* D. Reidel, Dordrecht, Netherlands.

Davis, Ernest. 1984. Shape and Function of Solid Objects: Some Examples. Computer Science Technical Report 137, New York University, New York, New York.

Hager, Greg. 1985. Naive Physics of Materials: A Recon Mission. In: *Commonsense Summer: Final Report,* Report No. CSLI-85-35, Center for the Study of Language and Information, Stanford University, Stanford, California.

Hayes, Patrick J. 1979. Naive Physics Manifesto. In: Donald Michie, Ed., *Expert Systems in the Micro-electronic Age,* Edinburgh University Press, Edinburgh, Scotland: 242–270.

Herskovits, Annette. 1982. *Space and the Prepositions in English: Regularities and Irregularities in a Complex Domain.* Ph.D. dissertation, Department of Linguistics, Stanford University, Stanford, California.

Hilbert, David. 1902. *The Foundations of Geometry.* The Open Court Publishing Company.

Hobbs, Jerry R. 1974. A Model for Natural Language Semantics, Part I: The Model. Research Report #36, Department of Computer Science, Yale University, New Haven, Connecticut.

Hobbs, Jerry R. 1985a. Ontological Promiscuity. *Proceedings, 23rd Annual Meeting of the Association for Computational Linguistics,* Chicago, Illinois, 61-69.

Hobbs, Jerry R. 1985b. Granularity. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence,* Los Angeles, California, 432-435.

Hobbs, Jerry R. and Robert C. Moore, Eds. 1985. *Formal Theories of the Commonsense World.* Ablex Publishing Corp., Norwood, New Jersey.

Hobbs, Jerry R., Tom Blenko, Bill Croft, Greg Hager, Henry A. Kautz, Paul Kube, and Yoav Shoham. 1985. *Commonsense Summer: Final Report,* Report No. CSLI-85–35, Center for the Study of Language and Information, Stanford University, Stanford, California.

Hobbs, Jerry R., and Paul A. Martin. 1987. Local Pragmatics. *Proceedings of the Tenth International Joint Conference on Artificial Intelligence,* Milano, Italy, 520-523.

Katz, Jerrold J. and Jerry A. Fodor. 1963. The Structure of a Semantic Theory. *Language,* Vol. 39: 170–210.

Lakoff, George. 1972. Linguistics and Natural Logic. In: Donald Davidson and Gilbert Harman, Eds., *Semantics of Natural Language*: 545-665.

McDermott, Drew. 1985. Reasoning about Plans. In: Jerry R. Hobbs and Robert C. Moore, Eds., *Formal Theories of the Commonsense World,* Ablex Publishing Corp., Norwood, New Jersey: 269-318.

Miller, George A. and Philip N. Johnson-Laird. 1976. *Language and Perception,* Belknap Press.

Rieger, Charles J. 1974. Conceptual Memory: A Theory and Computer Program for Processing and Meaning Content of Natural Language Utterances. Stanford AIM-233, Department of Computer Science, Stanford University, Stanford, California.

Schank, Roger. 1975. *Conceptual Information Processing.* Elsevier Publishing Company.

Shoham, Yoav. 1985. Naive Kinematics: Two Aspects of Shape. In: *Commonsense Summer: Final Report,* Report No. CSLI-85-35, Center for the Study of Language and Information, Stanford University, Stanford, California.

Stickel, Mark E. 1982. A Nonclausal Connection-Graph Resolution Theorem-Proving Program. *Proceedings of the AAAI-82 National Conference on Artificial Intelligence,* Pittsburgh, Pennsylvania: 229-233.

Talmy, Leonard. 1983. How Language Structures Space. In: Herbert Pick and Linda Acredolo, Eds., *Spatial Orientation: Theory, Research, and Application,* Plenum Press.

Talmy, Leonard. 1985. Force Dynamics in Language and Thought. In: William H. Eilfort, Paul D. Kroeber, and Karen L. Peterson, Eds., *Proceedings from the Parasession on Causatives and Agentivity, 21st Regional Meeting, Chicago Linguistic Society,* Chicago, Illinois.

Zahn, C. T., and R. Z. Roskies. 1972. Fourier Descriptors for Plane Closed Curves. *IEEE Transactions on Computers,* Vol. C-21, No. 3: 269-281.