On primitives, prototypes, and other semantic anomalies

Terry Winograd
Stanford University

Over the past few years, there have been a number of papers arguing the relative merits of *primitives* and *prototypes* as representations for the meaning of natural language. Much of the discussion has been both pugnacious and confused, with each author setting up one or another straw-man to knock down. Much of the confusion has resulted from a lack of agreement as to what it would mean for a system to use primitives or prototypes. There are several different dimensions along which semantic formalisms vary, and many of the arguments have blurred these into a single distinction.

In this paper, I propose a framework within which to compare a variety of semantic formalisms which have been proposed in linguistics and artificial intelligence. The paper lays out three dimensions (called *ontological, logical,* and *relational),* describing the relevant options along each and the implications of making alternative choices in the design of a formalism. It does not attempt to demonstrate that one or another alternative is right, but instead tries to clearly state the advantages and disadvantages of each in a non-partisan way. It is more in the style of a text-book than of a research paper. Its contribution will, I hope, be in dissolving some non-issues which have occupied previous discussion, and in focussing attention on the real distinctions between alternative proposals. My own prejudices are set forth in Winograd (1976) and Bobrow and Winograd (1977). In addition to citing primary sources, I will make particular reference to the discussion by Wilks (1977) since it is recent and sets out a number of the same issues.

The ontological dimension

The formalisms we want to compare are all based on the use of *symbol structures* to represent meaning. There are deep philosophical questions as to how much of meaning can be captured in a formal system, but such questions are outside the scope of this paper. We will take it for granted that meaning is to be characterized in terms of structured relationships between discrete symbols. The first question, then, is just what these symbols are. There are three basic positions which have been taken:

LINGUISTIC. In many older accounts of meaning, the only entities which take part in the formal structure are the entities of language: words, morphemes, phrases, and sentences. The dictionary is an account of meaning within this tradition. The meaning of a word is expressed in terms of structures made up of other words, without any direct appeal to concepts which lie outside the language.

PSYCHOLOGICAL. Most current work in AI and psycholinguistics assumes that the entities which are manipulated in the formal theory represent some sort of *concepts* which underlie language use, but are not themselves part of the language. These concepts have psychological reality, in that they correspond to functional components in the memory and language activity of a person. Words and sentences are seen as corresponding to structures of underlying concepts. A psycholinguistic theory includes an account of the processes by which language is translated into conceptual structures, and generated from them. In the case of AI systems (such as the *conceptual dependency* formalism of Schank (1972)), the commitment to PSYCHOLOGICAL entities is a global assumption which plays little role in the methodology of the work. In the case of psychological experimentation (for example, much of the work described by Clark and Clark (1977)), it is a hypothesis to be tested explicitly. Some theoretical psychologists (such as Miller and Johnson-Laird (1976) and Fodor (1975)) have characterized it is a private "language of thought".

THEORETICAL. A more cautious stance is taken by most theorists who work within the generative linguistics paradigm. They argue that the symbols of their formal semantic theories need not correspond to functional psychological entities. The symbols and structures play a role similar to that of postulated theoretical entities in physics, such as neutrinos and probability waves. A system based on them is justified in terms of its resulting overall simplicity and ability to account for the observable phenomena, not by finding psychological correlates for its individual terms. This view shares with the psychological view the notion of *lexical decomposition.* Words and sentences of the language correspond to structures built up of non-linguistic symbols.

There has been a certain amount of confusion within both syntactic and semantic theory about whether there is any psychological reality to the formal constructs postulated by linguists. In the 60's, experiments were carried out (e.g., Miller, 1962) looking for psychological correlates of transformations, with generally negative results. Chomsky has repeatedly reiterated his official stance that the validity of transformational theory is not based on any assumption as to whether transformations play a functional role in language comprehension or production. Similarly, as Wilks (1977) points out, Katz's view of semantic markers shifted from PSYCHOLOGICAL (in Katz and Fodor, 1965) to THEORETICAL (in Katz, 1972).

In doing AI research, the issue can be finessed. In building a program, one must develop a set of symbolic structures which are used functionally—they play a direct role in the memory and reasoning of the system. In this sense they are purely psychological (the psychology of the computer program, not of a person). When the program is viewed as a 'theory of human language use', two routes can be taken. If *strong psychological equivalence* is claimed, there is an assumption that the internal organization and objects of the program correspond to the organization and objects in the mind of a human language user. An alternative position of *weak psychological equivalence* is similar to that of the generative linguists. The program as a whole is justified by its ability to match human performance, but no claims are made about the ways in which its organization maps onto psychological phenomena. Since programs can be built without confronting this issue, there has been a tendency by AI researchers to handwave about it, taking whichever viewpoint seems most advantageous in a given discussion.

Begging the fundamental question of semantics

A persistent cause of misunderstanding in arguments about semantics has been a lack of agreement over what a 'semantic theory' should achieve. From a philosophical standpoint, the issue centers around what *meaning* is. The fundamental question is that of the relationship between symbols (words) and a world about which they speak. From an AI standpoint, the question is operational—how can a symbolic system be organized which accounts for the phenomena of language use. As pointed out by Fodor (1978), no answer to the second question, no matter how clever or elegant, is an answer to the first. In creating a system which accepts text, answers questions, or enters into a dialog, we have not created a theory of semantics, we have created another class of objects for which such a theory is needed.

This observation applies regardless of which of the three choices is taken along the ontological dimension. In taking words as the formal objects, we leave the semantic problem completely unaddressed. In relying on psychological entities, we transform the question into the equally difficult one "How are concepts related to the world which they are concepts about?". Similarly, with theoretical objects we beg the question by pushing it into a different domain. As many people have argued, (e.g. Lewis (1972) in discussing Katz and Fodor's theory of semantic markers), translating English into 'Markerese' doesn't illuminate the fundamental nature of meaning any more than translating it into French.

Wilks (1977) describes several papers which argue for the necessity of a semantic theory along the general lines of Tarski and recent work in model-theoretic semantics for formal languages. He characterizes them as criticisms of semantic primitives and argues that they are based on weak 'escape arguments'. He is correct in concluding that the concerns of these authors are orthogonal to the specific technical debate about primitives, but wrong in assuming that they are arguments in the same domain at all. In creating formal systems for representing and manipulating structures corresponding to meaning, we are not forced to answer the fundamental question of what meaning is. As Wilks points out, this question has been asked for thousands of years, and technical progress does not seem to depend on clearing it up.

There are valid doubts about whether adequate semantic formalisms (in the AI/operational sense) can be developed without more careful thought about the basic questions. In particular, our unexamined assumptions about the nature of meaning can lead us down paths in the problems we choose to look at, which may in the long run conceal other more fruitful paths. However, this sort of question has not been addressed in current AI work, and for the purposes of setting up a clear framework for understanding that work, we will continue to ignore it. A characterization of a semantic formalism in terms of the dimensions of this paper has nothing to say about the fundamental nature of semantics.

The logical dimension

As implied in the previous section, we are primarily concerned with the operational implications of different formalisms—the ways in which they can be used in language comprehension and production. Each symbol or structure of symbols plays a role in reasoning processes which underlie language activities, and there are a number different approaches to dealing with them. There are three basically different views of the logical status of the individual concepts (or words):

ABSTRACTION. The tradition drawn from logic and linguistics is to view the elements of a semantic formalism as logical abstractions—predicates and constants within a logical system. The meaning of a word is a structure of semantic elements which express the logical truth conditions determining its applicability. For example, if we analyze one sense of "bachelor" as having the semantic components HUMAN, MALE, and UNMARRIED, it is implied that any object to which that sense of the word could be properly applied will fit the truth conditions corresponding to those terms. If "kill" is analyzed as a structure of the form CAUSE(X, DIE(Y)), then we can safely deduce from the fact that "A killed B" that, among other things, B died.

There are many old and unsettled debates about the status of such knowledge as *analytic* or *synthetic*. The issue here is not that distinction, but the status of the semantic analysis as leading to logical consequences which can be drawn from the the application of a given word.

PROTOTYPE. One of the currently fashionable trends in AI is the development of languages and systems based on some kind of *frame* or *prototype* representation. The basic motivation comes from the observation that much of what we know about the world is not in the form of simple logical statements, but in knowledge about what is *typical* or *expected*. If we represent the meaning of "buy" and "sell" in terms of a COMMERCIAL-TRANSACTION scenario which includes the transfer of money, we also want to be able to apply it to cases which involve the exchange of valued objects other than money. However, we do not want to do this by creating an abstraction (e.g. the exchanged object is a VALUED-OBJECT) and thereby lose the information that it is usually money.

Many papers have been written on the advantages and problems of including prototypical information as a fundamental part of a semantic representation. Formally, such systems are distinct from those based on logical abstraction only if issues of computational order and resources are taken into account (See Winograd (1976), for a discussion of these issues). However, it is important not to focus too narrowly on form rather than use: there is a clear difference in approach between the adherents of the alternate views. Some systems (such as Schank's (1972) system of primitives) are clearly based on prototypes even though they may not appear as such in the formal characterization. The inferences they draw from semantic decomposition are based on typical expectation, rather than logical certainty.

Prototype-based systems have often gone along with a psychological view of the status of the symbols they use. Some of the motivation has come from psycholinguistic experiments which indicate that in many cases people are uncertain about the applicability of words to 'borderline cases', although they have a clear notion of the 'prototypical case'. This applies to areas of the vocabulary as varied as color terms (Berlin and Kay, 1969) and simple nouns such as "cup", "glass", and "bowl" (Labov, 1973). The implication is that the semantic representation of words is organized around a set of 'most typical' cases rather than around a checklist of logical criteria which must be met for the word to be applied.

EXEMPLAR. Extending the prototype notion one step further, some psychologists have suggested that our understanding of words is based on having *exemplars* which are drawn from experience. Rather than having a semantic prototype for "fruit", we may have an exemplary fruit (e.g. a red apple) and understand the use of the word by comparison to what we know about this apple. The line between prototypes and exemplars is not sharp, but there is a difference in emphasis. Prototypes emphasize the presence of information which is typical to the class of objects described by a word, while exemplars emphasize the ability to reason by comparing one specific object to another specific object, which may have its own peculiarities which are not general to the class.

Although there has been some discussion of reasoning by analogy (e.g. Moore and Newell, 1973), no system I know of has really made use of exemplars in a substantial way. There are many difficult issues surrounding the selection of the 'important' or 'invariant' aspects of the exemplar in a specific context. Critics of AI (e.g. Dreyfus, 1972) see this as being impossible to adequately represent in a formal system. Whether this turns out to be ultimately true or not, we are far from having explored the potential for such reasoning within AI programs.

What is a primitive?

Before going on to the third dimension—the way in which the symbols within a semantic formalism are interrelated—it is useful to examine the notion of *primitive* which plays a central role in arguments on semantics. In understanding the properties of semantic primitives, it is helpful to look at two other domains where primitives have played an important role: chemistry and mathematics. Much of the thinking and discussion about primitives draws on conscious or unconscious comparisons with these two domains, often without recognition that they differ in some critical ways.

Chemistry. One exemplar of a system based on primitives is the analysis of physical substances as structures made up of elements. There are atomic elements (note how much of the abstract vocabulary comes from this exemplar), and well-defined rules for the ways they can be combined into structures. Every substance, no matter how complex, can be analyzed as a compound of these primitive elements. The set of elements is experimentally determined and dealt with as a fact of nature—no two chemists would imagine postulating different sets of elements in their theories. Similarly, the structural analysis of a substance is not a matter of theoretical choice, but can be determined empirically.

Mathematics. One of the methodological advances in the foundations of mathematics at the beginning of this century was the understanding of how complex mathematical systems could be constructed in a systematic way from small sets of primitive concepts. Beginning with a primitive basis (such as the notions of *set*, *inclusion*, and *the null set*), one can define complex constructions, and use these in still further definitions to build up ever-widening circles of complexity. In doing this, each new term is defined in terms of previous terms and simple rules of composition. The meaning of a complex term like "abelian group" or "divisor field" can be reduced step by step to primitives through these definitions. The choice of primitives is not determined by the domain to be covered. For any field of mathematics, there are alternative axiomatizations which take different things as primitive, and define others in terms of them. Even with the same set of primitives, there are alternative ways of defining

higher order concepts. For example, there are different ways of embedding the real numbers in the rational numbers for which it is quite difficult to prove equivalence.

These two examples illustrate some typical features of primitives listed below (the terms used here are somewhat expanded from those in Wilks, 1977). Not every system based on primitives exhibits all of them, but they form a part of our understanding of what it is to be 'primitive':

1. Finitude. A system contains a relatively small closed set of primitives. As it is applied to a wider range of things (substances, mathematical constructs, vocabulary items), the set of primitives remains fixed. The number of primitives should be substantially smaller than the number of things which can be reduced to combinations of primitives.

2. Comprehensiveness. The set of primitives covers the range of phenomena. Every entity of interest can be expressed as a structure of primitives. For example, a chemist would be upset by a new substance which was not built of the available elements, and a mathematician would reject a new definition which was not in terms of the primitives of his or her axiomatization.

3. Completeness. A description of an entity in terms of primitives is sufficient for generating all of the information about the entity. There are no 'hidden properties'. This does not mean that the information must be explicit—a set of mathematical definitions does not provide all of the theorems, but it does provide a basis for proving all those which could be proved. In the case of substances, this criterion does not apply. Information other than the chemical structure (for example energy, phase, crystalline structure, etc.) is needed for determining the properties of a substance.

4. Independence. Primitives should not be definable in terms of one another. This is clear in the case of chemical elements, and in mathematics it provides a strong metric for judging axiomatizations. There is a high value placed on reducing the primitives to an absolutely minimal set.

5. Canonicality. The analysis of an entity as a structure of primitives should be unique and unambiguous. Chemists agree on the structure of a compound as a unique formula. Within a particular axiomatization of a mathematical system, there is one and only one way a term such as "integer" is defined in terms of the primitives.

6. Irreducibility. The meaning of a primitive cannot be expanded within the same level of theory. There are many issues here as to what a 'level of theory' is, but the application is clear in chemistry. The primitive elements can indeed be described as composite structures made up of even more primitive sub-atomic particles. But in doing so, we move from chemistry to atomic physics. For the purposes of doing normal chemistry, it is more useful to treat them as primitives. It is important to recognize that 'primitivity' is always relative to an overall choice of the scope of the theory.

In comparing the various forms of semantic primitives, we will look at the ways in which they match these criteria.

The relational dimension

The notion of primitive makes sense only within a system of interrelated terms. The basic idea of composition from primitives is only one of several possible ways of organizing such sets of relationships:

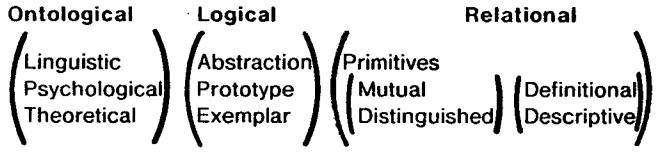
PRIMITIVES. The most straightforward use of semantic primitives would be a system in which the full meaning of any word or phrase could be expressed as a structure whose components are chosen from a small set of primitives, combined according to a well-defined set of rules. No existing system is pure in this sense, as discussed below.

MUTUAL. Another approach is to have a web of mutually related elements, with no primitive set on which to 'bottom out'. A standard dictionary describes word meaning in this way. Words are defined using other words which are defined using others, and so on, inevitably leading to circularity. A mutually related system of terms can be either DEFINITIONAL or DESCRIPTIVE. In a DEFINITIONAL system, each item is defined by giving a structure made up of other items. The definition is complete, in that no information which is available from the term itself is lost by replacing it with the definition. In a DESCRIPTIVE system, each term is described by structures of other terms, but these do not necessarily capture its full meaning. Although the dictionary is normally thought of as being DEFINITIONAL, this is the case only for very precise technical terms. For most of the common vocabulary, the 'dictionary definition' is a quite partial account of the meaning of the word.

DISTINGUISHED. In systems based on mutual relations, it will often be the case that some terms tend be be used in definitions or descriptions much more often than others. There may be small finite *distinguished subsystems* of terms which form a standardized basis for a large number of descriptions. These terms need not be primitive in the senses discussed above—they may be further reducible, definable in terms of each other, and may provide only a partial coverage of the meanings to be expressed. However, there are organizational (and computational) advantages to granting them a privileged status in the way other definitions and descriptions are built up. In fact, most of the argument in favor of semantic primitives for AI systems has been (as we will see below) argument in favor of having one or more preferred subsystems within a mutually related system.

Some examples

The following table summarizes the dimensions and choices described above. In this section, we will use it to characterize a number of existing formalisms.



Dimensions of choice in a semantic formalism

The traditional dictionary. The traditional dictionary is clearly LINGUISTIC, based primarily on ABSTRACTION, and MUTUAL relationships. It varies between being DEFINITIONAL and DESCRIPTIVE, and at times does include some PROTOTYPE information. The popular view of the dictionary tends to ignore the PROTOTYPE and DESCRIPTIVE aspects.

Theories from generative linguistics. Semantic theories within the Chomskian tradition of generative linguistics tend to be THEORETICAL, based on ABSTRACTION and PRIMITIVES. Katz and Fodor (1964), Jackendoff (1976), and Leech (1969) all fit these categories. There is an occasional hint of PSYCHOLOGICAL relevance, but it does not play a major role in the methodology. Within the school of 'generative semantics', there are many approaches. Much of Fillmore's (1974, 1975) work is an examination of how PROTOTYPE and EXEMPLAR systems can provide insights which do not fit neatly into ABSTRACTION. Some of the earlier work on 'underlying verbs' takes a more LINGUISTIC turn, in which the underlying components are seen as closely related to actual lexical items.

Semantics based on formal logic. Much of the work on the semantics of natural language has been closely related to work on the semantics of formal languages. This includes the classical work on issues like reference, and more recent attempts to view English as a formal language, as developed in Montague grammar. On the first two dimensions, this work is clearly THEORETICAL and ABSTRACTION based. On the third, the relationship between the symbols used for semantic representation carries over that of an underlying logical system. From the point of view of the semantic theory (the relationship between words and underlying entities), each predicate or constant is a PRIMITIVE. The fact that these are related by theorems, definitions, etc. within the logical system is independent of the semantic formalism in the same sense that the representation of elements in terms of sub-atomic particles is independent of ordinary chemistry. The clarity of this distinction (between the semantic rules and the reasoning rules) is one of the advantages of this style of work, not shared by most AI programs, which use data structures and procedures which make no clear distinction.

Conceptual Dependency. Schank has been one of the most insistent advocates of primitives, and his early (1972) work was clearly PSYCHOLOGICAL based on PRIMITIVES. As mentioned above, his attention to 'typical' inferences places it closer to PROTOTYPE than to ABSTRACTION. In trying to expand his theory beyond the set of simple actions for which it was initially developed, he has gradually shifted away from a strong PRIMITIVES based view, and has been one of the major developers of systems based on DISTINGUISHED subsystems. Schank and Abelson (1977), provide subsystems for actions, scales reflecting a person's state, causes, scripts, goals, plans, goal outcomes, interpersonal themes, and life themes. Their students have carried out the same kind of activity in other areas, such as the uses and classification of physical objects. In all of this work, the emphasis is on finding a plausible and useful set of terms, rather than on justifying their primitive status. Most of the arguments are based on the pragmatics of doing language comprehension and reasoning within the system.

KRL. KRL provides a language for representation within computer systems. As such, it is neutral between a PSYCHOLOGICAL and THEORETICAL stance, but the authors lean heavily towards the PSYCHOLOGICAL in developing their formalism. It is clearly based on PROTOTYPES, and much of the discussion (see Bobrow and Winograd, 1977) centers around this aspect. It is based on a MUTUAL DESCRIPTIVE set of relationships. DISTINGUISHED subsystems have been developed within specific applications (see Bobrow, Winograd, et. al., 1977), but these have not been a part of the basic formalism.

Preference Semantics. Wilks' system of 'preference semantics' is one of the hardest to understand, since he seems to combine many different (and often incompatible) views. He insists that his system is based on PRIMITIVES, but it has few of the characteristics described above. In fact, his discussion argues strongly for the possibility of a MUTUAL DEFINITIONAL system, and he provides an interesting set of DISTINGUISHED subsystems (1977, Appendix A). In stating that "primitives are to be found in all natural language understanding systems" (1977, p. 19) he seems to be using the term 'primitive' to cover any formal symbol used in a semantic system. He argues against the PSYCHOLOGICAL basis, but alternates between the other two possibilities along the ontological dimension. He is LINGUISTIC in stating that his formalism is consistent with the view that "Every semantic primitive can appear as a surface word in a natural language", and THEORETICAL in arguing that the primitives are part of an interlingual "primitive language" which is a "useful organizing hypothesis" which has no independent justification in psychological terms, and "has no *correct* vocabulary, any more than English has". His formulas generally contain only ABSTRACTION information in their structure, but have PROTOTYPE information (or in his terms, 'preferences') in the assignment of types of objects to the nodes.

OWL. The OWL representation is much closer to a LINGUISTIC base than any of the others listed here. It is described as a system of 'concepts', but its developers (Szolovits, Hawkinson, and Martin, 1977) have paid a good deal of attention to the way that natural language words and collocations can be preserved in the representation. It has a MUTUAL DESCRIPTIVE organization, which focuses on ABSTRACTION sorts of information,

although the semantics of the reasoning process are not clearly enough specified to distinguish between this and other choices on the logical dimension. The term 'exemplar' is used in OWL to refer to sub-classes of a larger class, a concept related to but not the same as the one described above.

Semantic networks. There are many versions of semantic networks, and it is hard to say anything which applies across the board. The majority have been argued on PSYCHOLOGICAL grounds, have focussed on ABSTRACTION information, although with some PROTOTYPE, and have been a web of MUTUAL DESCRIPTION. The network notation is well suited to MUTUAL (as opposed to PRIMITIVE), but is general enough to be used for almost anything.

Properties of semantic systems

The purpose of the classification given above is to provide a basis for comparing the merits and problems of alternative formalisms. Rather than arguing whether primitives are right or wrong, we will examine some desirable properties for semantic systems and see what they imply for the choices to be made along the three dimensions. This paper cannot hope to cover the full range of important issues, but as examples we will consider the following properties:

The ability to state significant generalizations
Criteria for deciding on a set of semantic entities
Coverage of relevant semantic phenomena
Canonicality and its effects on memory form
Possibilities for dealing with extended meaning and metaphor

The ability to state significant generalizations. The *raison d'etre* of a semantic theory is the desire to find regularities in the way language conveys meaning. Rather than enumerating the relationships among every possible set of texts, we can assign formal semantic structures to texts in a regular way, and systematically describe relationships between these structures. The theory is interesting to the extent that the formal semantic system allows us to find regularities and state broader generalizations than we could at the surface level.

There are many possible views as to what kinds of generalizations are most interesting. Linguists look for generalizations which predict the judgements of native speakers as to whether sentences are well-formed. Some, like Jackendoff (1976) also look for generalizations as to the entailment relations between sentences. AI work, such as that of Rieger (1975) emphasizes inferential generalizations—that certain inferences will be made whenever a given underlying semantic structure appears. AI systems in general are based on 'reasoning' programs which make use of semantic representations to do reasoning which is independent of the specific linguistic form in which the knowledge was stated.

In some discussions of primitives, it is implied that it is necessary to have a system based on primitives in order to make significant generalizations. It should be clear from the discussion above that this is a confusion of categories. Any system of formal semantics is based on generalization. The specific choice to base it on primitive decomposition may lead to a different set of generalizations, but not a necessarily better one.

Criteria for deciding on a set of semantic entities. The main factor influencing the choice and justification of semantic entities within a formalism is the choice along the ontological dimension. Those who take a LINGUISTIC position need make no choice—the words of the language are themselves the entities of the semantic theory. There is work to be done in determining the relations between them, but the set of entities is given from the beginning. Those who take a THEORETICAL stance are free to create semantic entities at will, but must justify them by demonstrating that the set chosen leads to generalizations and simplifications which are not shared by alternative sets. In the generative grammar tradition, a good deal of attention is given to finding a highly valued set. Through careful work, one can construct tests in the form of sentences whose acceptability would be predicted by one possible set, and not by another. Simplicity of stating the semantic theory is used to choose between sets with equal coverage.

In the AI tradition, the selection of entities is more intuitive and less careful. A system as a whole is claimed to 'work', and there is little precise evaluation of which aspects of the formalism were critical, and what might be done with alternatives. In this context, there are only vague intuitions and heuristics to guide the choice of entities and their relationships. Wilks accepts this, in noting that "no direct justification of the vocabulary [of primitives] makes any sense."

The most interesting problems arise if the formalism is intended as a PSYCHOLOGICAL theory. In this case, the determination of a set of semantic entities is an empirical question. There is an implicit claim that there are functional equivalents to the elements of the semantic theory within the psychological activities of comprehending and generating language. It is possible to invent experiments which can choose between alternative theories according to the detailed predictions they make about human performance. Some of the distinctions above (such as that between ABSTRACTION, PROTOTYPE and EXEMPLAR) grew out of experiments of this type. However, there is a large gap between the isolated examples handled in experiments and the kind of coverage needed in a comprehensive semantic formalism. Those people in AI who have built large-scale systems have not looked to detailed psychological justifications, even though they often informally describe their formalism as a psychological theory. When Schank (1972) calls his formalism '*conceptual* dependency', or Jackendoff describes his system as using '*cognitive* primitives' the appeal to psychology is suggestive, not of direct relevance to the methodologies they follow.

Within a PSYCHOLOGICAL viewpoint, there are many further issues as to the generality of the postulated semantic entities. Are they idiosyncratic, or shared by all competent speakers of a language? Are they language-specific, or do they represent a more basic experiential knowledge which cuts across cultures and languages? If they are not language-specific, then are they innate or learned? There has been some interesting work done on these questions in very specific semantic domains such as the lexicon for describing colors, but once we move outside of these limited domains, most of what can be said is anecdotal or purely speculative.

Coverage of relevant semantic phenomena. In developing a comprehensive semantic theory, there are many aspects of meaning which must be taken into account. A formalism which is developed for one aspect of meaning (for example, the hierarchical relationships between the classes named by common nouns) may be inadequate or completely irrelevant for others (for example, the ways in which participants are related to events). In some cases, a general approach cuts across several aspects. Much of the discussion of primitives and prototypes above can be applied both to classification (for example, Schank's (1972) classification of acts vs. Lakoff's (1977) 'gestalts') and to the case relationships between participants and an act (Fillmore's (1968) notion of a primitive set of cases vs. the Bobrow and Winograd (1977) notions of hierarchies of prototypes with named 'slots').

Existing semantic formalisms are all partial, and many of the arguments in the literature are of the "I can do something you can't do" style. It is clear, for example, that PRIMITIVES are not well suited for handling the broad vocabulary of nouns and verbs describing the objects and actions of our world in all their variety. As Wilks says, "No representation in primitives could be expected to distinguish by its structure *hammer*, *mallet*, and *axe*." Formalisms based on ABSTRACTION are problematic when we attempt to deal with lexical fields where there are no clear criteria for whether a word applies. This includes the naming of simple objects, such as "cup" and "bowl" (Labov, 1973), as well as the more obvious areas of metaphor. On the other hand, alternatives, such as PROTOTYPE systems based on MUTUAL relations have been far less developed in the details of the generalizations they allow, and the specification of how they would deal with any specific semantic domains.

It is clear that no formalism at this point has a claim to "Anything you can do, I can do better." Intuitions as to which aspects of language are most central play the leading role in determining which of the competing theories seems most promising.

Canonical form and its effects on memory and reasoning. In early work on semantic primitives, there was a good deal of debate about the advantages provided by a *canonical form* for the representation of meaning. Two words or sentences with the same meaning have identical semantic representations in a formalism based on canonical form. In other formalisms, they may have *equivalent* representations (anything inferrable from one would be inferred from the other) which nevertheless differ in form. Typically, PRIMITIVE systems tend to support a canonical form, while MUTUAL organizations do not. However, DISTINGUISHED subsystems can be used to create a canonical form for their particular aspect of meaning in a system which does not depend on primitives. By choosing to always expand into the terms of this subsystem in the same way, all of the properties of canonical form apply.

In evaluating the benefits of canonical form, it is important to take into account the procedural aspects. In its simplest usage, each piece of input text is converted immediately to canonical form and stored that way. Inferences are based on the elements of this expanded form, and memory search depends on finding the form corresponding to the query as a subset of what is stored. In a more sophisticated use, the canonical form is available for *potential* expansion, but memory can include unexpanded structures built up out of a vocabulary of non-primitive semantic entities. Expansion is done only when needed for a specific task such as matching a new input to previous knowledge in answering a question. The advantages and disadvantages of canonical form are somewhat different for these two organizations. The primary ones can be summarized:

1. Absence of ambiguity and vagueness. This property applies to the canonical form after expansion. It is a global property of systems based on expansion at input—since meanings are expanded into canonical structures of primitives at the time they are analyzed, there is no remaining uncertainty about their meaning. This is viewed as an advantage by those who emphasize the use of the formalism in abstract reasoning, and as a disadvantage by those (like Martin, 1976) who emphasize the importance of context and interpretation in using knowledge. Martin argues that a semantic representation for natural language must share its ability to represent imprecise meanings.

2. Reasoning activity at input time. The process of expansion to canonical form can be used as a procedural driver for carrying out inference. Much of the work on conceptual dependency makes use of this organization. The advantage is a uniform way of triggering standard inferences. The disadvantages come from the problems of triggering too much—of drawing inferences far below the level of detail relevant to the particular context because the canonical form demands expansion to that level.

3. Uniqueness for indexing and search. A canonical form can be stored and indexed in a uniform way which makes it possible to use straightforward algorithms for memory search and consistency checking. These have the advantages and disadvantages of most uniform procedures for dealing with complex structures—they are easy to write and understand, but they suffer from combinatorially explosive inefficiency and tend to bog down for all but tiny toy bodies of knowledge. One of the fundamental technical differences among existing systems is in whether they emphasize uniformity (as in most logic-based systems, and in early versions of conceptual dependency) or the

provision of explicit tools for controlling memory search and inference (as in KRL).

4. Association of inference rules with primitive elements. In a system which is expected to expand meanings into canonical form (either at input time or in the process of reasoning), inference rules can be associated with the most general primitives (e.g. GO, used in a sense which covers all sorts of change, as in Jackendoff (1976)). In a system which does not expand to a common base, the same inference might have to be repeated in a number of places. The disadvantage arises in the case where an inference is associated with a higher-level meaning (such as "flee" having implications not shared by other instances of going). In a fully canonical system, it is necessary to recognize the particular combination of primitives which triggers the inference. In systems like that of Rieger (1975), there are discrimination nets, used to sort out the appropriate inferences from the expanded forms. This again leads to a combinatorial problem which becomes untenable in all but the smallest systems. Like the other issues, this one is complicated by the ability to build systems which partake of canonical expansion to some degree, either by expanding only along certain dimensions, or by operating with a mixture of expanded forms and non-primitive-based forms from which they were derived.

Possibilities for dealing with extended meaning and metaphor. A recurring theme in discussions of semantics is that of *metaphor.* Any realistic view of language must take into account the fact that words are used in ways which defy simple analytic characterization of their meaning. There are explicitly poetic metaphors, conventional metaphors ("His ideas were *beyond* me", "Carter named three main *targets* in his *war* on inflation"), and a wide range of cases in which meanings are extended beyond their prototypical application. For example, if we define "spend" in terms of a commercial transaction, then it must be extended to deal with "I spent a week in Boston." In general, formal semantic theories have not gone very far in dealing with these problems. Those who base systems on PROTOTYPE or EXEMPLAR reasoning argue that this is an important step towards dealing with the fuzzier aspects of language. However, the computational details needed to make the power of such systems clear have not been filled in. They either stick to trivial cases (as in Moore and Newell, 1973), or operate in ways which do not depend on going beyond standard logical meaning. This area remains one of the most tantalizing and difficult for future research.

REFERENCES

Berlin, B. and P. Kay, *Basic color terms: their universality and evolution,* Berkeley: Univ. of California Press, 1969.

Bobrow, D.G. and T. Winograd, An overview of KRL, a Knowledge Representation Language, *Cognitive Science* 1:1 (January, 1977), 3-46

Bobrow, D.G., T. Winograd, and the KRL Research Group, Experience with KRL-0: One cycle of a knowledge representation language, *Proceedings of the Fifth International Joint Conference on Artificial Intelligence* (August, 1977), 213-222.

Clark, H.H., and E.V. Clark, *Psychology of Language: An Introduction to Psycholinguistics,* New York: Harcourt Brace, 1977.

Dreyfus, H. L., *What computers can't do: a critique of artificial reason,* New York: Harper & Row, 1972.

Fillmore, C., The case for case, In Bach and Harms (Eds.), *Universals in Linguistic Theory,* Chicago: Holt, 1968, 1-90.

Fillmore, C., The future of Semantics, *Berkeley Studies in Syntax and Semantics* I, Dept. of Linguistics, Univ. of California Berkeley, 1974.

Fillmore, C., An Alternative to Checklist Theories of Meaning, *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society,* Cogen et al. (Eds.), University of California, Berkeley, 1975.

Fodor, J.A., *The Language of Thought,* New York: Cromwell, 1975.

Fodor, J.A., Methodological solipsism as a research strategy in psychology, unpublished draft, 1978.

Jackendoff, R., Toward an explanatory semantic representation, *Linguistic Inquiry* 7:1 (Winter, 1976) 89-150.

Katz, J.J., *Semantic Theory,* New York: Harper and Row, 1972.

Katz, J.J., and J.A. Fodor, The Structure of a Semantic Theory, in J. Fodor and J. Katz, (eds.) *The Structure of Language,* Prentice Hall, 1964.

Labov, W., The boundaries of words and their meanings, in C-J. N. Bailey and Roger Shuy (eds.), *New Ways of Analyzing Variation in English,* Georgetown Univ., 1973.

Lakoff, G., Linguistic Gestalts, *Proceedings of the Chicago Linguistic Society (CLS 13).* 1977, 236-287.

Leech, G., *Towards a semantic description of English,* London: Longman, 1969.

Lewis, D., General semantics, in Davidson and Harman (eds.), *Semantics of Natural Language,* Dordrecht: Reidel, 1972.

Martin, W.A., A theory of English grammar, unpublished notes, MIT, 1976.

Miller, G.A., Some psychological studies of grammar, *American Psychologist* 17 (1962), 748-762.

Miller, G.A., and P.N. Johnson-Laird, *Language and Perception,* Cambridge: Harvard University Press, 1976.

Moore J., and Newell. A., How can MERLIN understand?, In Gregg (Ed.), *Knowledge and Cognition,* Baltimore, Md.: Lawrence Erlbaum Associates, 1973.

Rieger, C., Conceptual memory and inference, in R.C. Schank, *Conceptual Information Processing,* Amsterdam: North Holland, 1975, 157-288.

Schank, R. C., Conceptual dependency: A theory of natural language understanding, *Cognitive Psychology,* 1972, 552-631.

Schank, R.C. and R.P. Abelson,, *Scripts Plans Goals and Understanding,* Hillsdale: Lawrence Erlbaum Associates, 1977.

Szolovits, P., L.B. Hawkinson, and W.A. Martin, An Overview of OWL, an language for knowledge representation, M.I.T. LCS-TM-86, 1977.

Wilks, Y., Good and bad arguments about semantic primitives, D.A.I. Research Report No. 42, University of Essex, May 1977.

Winograd, T., Towards a Procedural Understanding of Semantics, *Revue Internationale de Philosophie,* 1976 fasc. 3-4 (117-118).