Stanford Artificial Intelligence Laboratory
Memo AIM-2 9 2

November 19 7 6

Computer Science Department
Report No. STAN-CS-76-580

TOWARDS A PROCEDURAL UNDERSTANDING OF SEMANTICS

by

Terry Winograd

Research sponsored by

National Science Foundation

COMPUTER SCIENCE DEPARTMENT
Stanford University

Stanford Artificial Intelligence Laboratory **November 19 7 6**
Memo AIM-292

Computer Science Department
Report No. STAN-CS-76-580

TOWARDS A PROCEDURAL UNDERSTANDING OF SEMANTICS

by

Terry Winograd

ABSTRACT

The term "procedural semantics" has been used in a variety of ways, not all compatible, and not all comprehensible. In this paper, I have chosen to apply the term to a broad paradigm for studying semantics (and in fact, all of linguistics). This paradigm has developed in a context of writing computer programs which use natural language, but it is not a theory of computer programs or programming techniques. It is "procedural" because it looks at the underlying structure of language as fundamentally shaped by the nature of processes for language production and comprehension. It is based on the belief that there is a level of explanation at which there are significant similarities between the psychological processes of human language use and the computational processes in computer programs we can construct and study. Its goal is to develop a body of theory at this level. This approach necessitates abandoning or modifying several currently accepted doctrines, including the way in which distinctions have been drawn between "semantics" and "pragmatics" and between "performance" and '*competence".

The paper has three major sections. It first lays out the paradigm assumptions which guide the enterprise, and elaborates a model of cognitive processing and language use. It then illustrates how some specific semantic problems might be approached from a procedural perspective, and contrasts the procedural approach with formal structural and truth conditional approaches. Finally, it discusses the goals of linguistic theory and the nature of the linguistic explanation.

Much of what is presented here is a speculation about the nature of a paradigm yet to be developed. This paper is an attempt to be evocative rather than definitive; to convey intuitions rather than to formulate crucial arguments which justify this approach over others. It will be successful if it suggests some ways of looking at language which lead to further understanding.

This research was supported by the National 'Science Foundation under Contract NSF MC S 75-23252. This report will also appear in Revue lnternationale de Philosophie in 1977.

The views and conclusions contained in this document are those of the author(s) and should not be interpreted as necessarily representing the official policies, either expressed or implied, of Stanford University, NSF, or the U. S. Government.

Table of Contents

Part I. General framework for a procedural approach to semantics

1. Fundamental attitudes and assumptions

Every science begins with a framework of questions to be posed about a set of phenomena, and assumptions about the nature of the explanations to be sought. The following assertions are the basis for a procedural approach to linguistics, and introduce a number of concepts and terms which will be further developed throughout the paper.

1.1 In the study of human language, the primary focus is on the mechanisms underlying the production and understanding of utterances in a linguistic and pragmatic context.

1.2 The essential properties of language reflect the cognitive structure of the human language user, including the detailed properties of its memory structures, processing algorithms, and inherent computational limitations.

1.3 The production and understanding of utterances takes place within a structure of ongoing thought processes which include both linguistic and non-linguistic elements. Individual utterances can only be understood in terms of the role they play within this larger con text. A formalized understanding of language processes is a part of a broader formalization of thought processes.

1.4 Each utterance is constructed to serve a combination of communicative goals, using the set of elements and choices provided by the language. "Meaning" is multi-dimensional, formalizable only in terms of the entire complex of goals and knowledge being applied by both the producer and understander.

1.5 The formalisms most appropriate for building theories of thought and language are those that deal explicitly with the structure of knowledge as stored in a cognitive processing system, and of the processes which operate using this knowledge. Some of the fundamental properties of these processes derive from the allocation of finite processing resources.

1.6 Symbol structures and processes which operate on them play a major role in a formalization of cognitive processing. There is no single static structure which can represent the "meaning" of an utterance. Rather, there is a set of structures built and modified by speaker and hearer in the course of communication, and it is necessary to deal with the succession of structures and the nature of the changes.

2. Cognitive symbol systems

The principles of §1 have developed within a framework of "cognitive science" which deals with a broader range of cognitive structures and processes, including those of perception, learning, and problem solving, as well as language. It is not possible within the scope of this paper to present a full picture of the current state of cognitive science. It is in rapid development, and draws on work in a number of disciplines, including cognitive psychology, the theory of computation, and artificial intelligence. The following paragraphs lay out some of the points which most strongly affect the views of language described in this paper. From a procedural perspective, "semantics" is the study of the relationship between linguistic objects and the mental states and processes involved in their production

and comprehension. The "setting" and "meaning" of an utterance are couched in terms of hypothesized cognitive structures, which are not peculiar to linguistics.

2.1 Mind as a physical symbol system

Cognitive science is based on two assumptions:

2.1.1 The human mind can be usefully studied as a *physical symbol system*

> ***A physical symbol system consists of a set of entities, called symbols, which are physical patterns that can occur as components of another type of entity called an expression (or symbol structure). Thus, a symbol structure is composed of a number of instances (or tokens) of symbols related in some physical way (such as one token being next to another). At any instant of time the system will contain a collection of these symbol structures. Besides these structure-s, the system also contains a collection of processes that operate on expressions to produce other expressions: processes of creation, modification, reproduction and destruction. A physical symbol system is a machine that produces through time an evolving collection of symbol structures. Such a system exists in a world of objects wider than just these symbolic expressions themselves.*** -- Newell and Simon 1976, p. 116.

2.1.2 It is both possible and revealing to study the properties of physical symbol systems at a level of analysis abstracted from the physical details of how individual symbols and structures are embodied, and the physical mechanisms by which the processes operate on them.

There are a number of important philosophical and empirical questions raised by these assumptions. Some critics (e.g. Dreyfus, 1972, Weizenbaum 1976) argue that there are aspects of human experience which are not amenable to analysis in terms of any formal system, and others argue that analogic (as opposed to symbolic) representations of the world play a critical role in intelligence. The enterprise of cognitive science (and in particular, the procedural approach to semantics) does not rest on an assumption that the analysis of mind as a physical symbol system provides a complete understanding of human thought, or on the assumption that all physical symbol systems must have identical properties. For the paradigm to be of value, it is only necessary that there be some significant aspects of thought and language which can be profitably understood through analogy with other symbol systems we know how to construct.

Within linguistics, it is often asserted that a general cognitive theory is impossible. Chomsky (1975, p. 17, and pp. 138 ff.) has expressed strong doubts as to whether we can learn very much, as scientists at least, about the mechanisms which relate stimulus conditions to behavior, even given a model of the cognitive state. He labels these phenomena "mysteries" in contradistinction to the "problems" posed by studies of the formal structure of linguistic objects. In laying out the domain of semantic theory, Fodor

and Katz have argued against a taking into account the "setting" (in the socio-physical world) of a sentence, since:

> *. ..a complete theory of this kind is not possible in principle because to satisfy the above necessary condition [that it represent all possible relevant aspects of the setting] it would be required that the theory represent all the knowledge speakers have about the world . . . such a theory would blur the distinction between the speaker's knowledge of his language (his linguistic ability) and the speaker's knowledge of the world (his beliefs about matters of fact).* Katz and Fodor, 1964

This view exhibits the error of confusing the theory with the data. It is indeed impossible to lay down all of the knowledge that a speaker of a language might bring to bear on producing or understanding a sentence, or to determine all of the factors which influence an individual piece of behavior. But it is equally impossible to specify the form and location of all of the particles in any physical segment of the universe or to determine all of the forces acting on them. This does not preclude a science of physics. The goal of cognitive science is to determine the nature of the mechanisms which are available to a person in reasoning, perceiving and understanding. The goal of procedural semantics to to understand the special mechanisms associated with language use and the ways in which they interact with the rest of cognitive functioning. Neither calls for a complete map of human knowledge or of the entire cognitive structure of a specific individual.

. 2.2 The language of thought

In the literature describing current research in cognitive science, there are a number of terms used to describe the set of symbol structures which exist at a given time in a physical symbol system (a computer or nervous system). These include words such as "representation", "knowledge", and "model", which have precise meanings in various areas of science and philosophy and which take on a somewhat different meaning in this context. There are a number of assumptions about the human symbol system which form the basis for much current research.

2.2.1 The rules of composition for a language user's internal symbolic structures are not necessarily the same as those for a natural language. The symbol structures form a "language" in the same sense as a mathematical language or a computer language -- they are built up of symbols according to specified rules of composition. The processes which operate on them are not identical to the processes of communication which use natural language. Their detailed form is not assumed to directly reflect language, although there may be interesting correspondences to be discovered.

2.2.2 The entities represented by symbol structures do not all stand for physical or "real" objects, and need not reflect a consistent ontology. They include abstractions which correspond to:

physical and institutional objects

perceptual properties

events

abstract categorizations of objects, properties, and events

complex conceptual objects, properties, and events built up out of descriptions couched in terms of other mental entities.

linguistic objects such as words and expressions

hypothesized versions of the entities in the symbol systems of other people

The procedural approach to semantics does not propose any solutions to the age-old problems of ontology and epistemology. The initial assumption is only that there is a well-defined set of processes which operate on the symbols, and a means of building symbol structures on the basis of experience. It is consistent with the procedural approach to believe that many conceptual entities exist prior (both logically and temporally) to any use of language, while others are initially created to correspond to a word, with their other properties gradually being added on the basis of experience. Natural language can provide part of the material for the language of thought, in a way which is not directly reducible to pre-existing entities.

2.2.3 Many of the symbol structures correspond to propositions about the entities represented by their constituent symbols. Many of the operations on these structures produce other structures which can be loosely described as the results of "thinking" or *'reasoning" about those entities and propositions. The entire collection of structures is often referred to as a "model of the world", even though it is only a partial description. The word "model" is used in a loose sense which implies only the intention of correspondence to the real world, and the word "world" is used in a loose sense to have the breadth of the word "entity" as described in the previous paragraph. A person's model corresponds to the world only to the degree that the processes for creating symbol structures from experience and the process for deriving them from other symbol structures are accurate in their correspondence to perceptions and lead to valid predictions. The operations on symbol structures in a procedural semantics need not correspond to valid logical inferences about the entities they represent, but only to the kinds of reasoning done by people. The procedural approach begins from a standpoint very different from the more traditional logical approach. The symbol manipulating processes themselves are primary, and the rules of logic and mathematics are seen as an abstraction from a limited subset of them.

2.2.4 Although the set of symbol structures available to the processes in a language user are often referred to as that person's "knowledge", this does not imply that they are available to conscious self-examination. The world model can include a model of the internal knowledge structures themselves, but the correspondence between this model and the actual structures is as contingent as the correspondence between the rest of the model and the actual state of the world.

3. The abstract procedural model of language use

The procedural approach to language follows the generative view of language in dealing primarily with cognitive capacities, rather than the utterances themselves. It diverges in using a model of the processes of producing and comprehending sentences as an organizing framework for theories and detailed descriptions of linguistic phenomena. This focus on cognitive structures, rather than on abstract mappings between utterance forms and formal

semantic structures or truth conditions, emphasizes the effects of context (both linguistic and pragmatic) on the interpretation of utterances.

This section presents an abstracted model of ***language use,*** based on a prototypical verbal interaction, in which a speaker is addressing a single hearer face to face. The same concepts can be naturally extended to other forms of language, such as writing and silent verbalized thought, as discussed in §4.3.

3.1 There are two participants, a "speaker" and "hearer".

3.2 Each participant is engaged in using language as a part of a broader cognitive process, which includes the perception and interpretation of an environment, the formulation of goals, and the design and execution of plans to achieve those goals.

3.3 Each participant comes to the interaction with a set of capacities including:

3.3.1 A general cognitive processing capacity, which includes (among other things) a symbol system and a means of storing symbolic knowledge as described in $2.

3.3.2 A variety of specific representational forms, processing modes, and strategies used in different tasks of perception, action, reasoning and understanding. Those for language use include representations for linguistic objects at a variety of levels of analysis, such as phonological sequences and syntactic structures.

3.3.3 A body of specific beliefs (expressed as symbol structures based on the specific representational forms) making up the person's "model of the world". For the purpose of analyzing language use, some relevant parts of this model include:

a body of linguistic knowledge including the syntactic structuring, lexical meanings, speech act conventions, etc. of the language.

the language user's analysis (at multiple levels) of the events of the conversation.

a model of the other person, including his or her knowledge, current goals and processing state.

a model of the language user's own goals and knowledge.

3.4 Each utterance is the end result of a design process, in which the speaker devises a linguistic structure which attempts to achieve a combination of communicative goals using the resources available in the language.

3.4.1 There is no single formal object which is the "meaning" or "primary intention" of an utterance. Overall meaning is an abstraction covering all of the goals. Many of the goals are at a meta-communicative level, dealing with the personal interaction between speaker and hearer rather than the putative content of the utterances. The goals of any one utterance can include:

Causing an action, either verbal or non-verbal, on the part of the hearer.

Causing the hearer to go through an intended set of inferences or emotional reactions, either about the subject matter or about the interaction between speaker and hearer.

Conveying information about an entity al ready assumed to be in the hearer's world model.

Getting the hearer to create a new conceptual entity corresponding to some entity in the speaker's, world model.

Directing the hearer's attention to some entity or to selected properties of a known entity (possibly to establish context for a subsequent utterance).

Some of these goals are subgoals of others (e.g. establishing reference to an object in order to state a fact about it), while some are at least partially independent (e.g. stating a fact, drawing the hearer's attention to some entity, conveying an overall posture towards the hearer).

3.4.2 Like any complex design process, the design of an utterance does not proceed in a simple sequence of separable stages. It can involve a feedback process in which decisions are made, and then changed on the basis of exploring their consequences. It is largely unconscious, and involves the use of all of the speaker's knowledge, including a current model of the hearer, knowledge about the entities being described, and knowledge of the language. Each utterance is part of a larger sequence, and part of the design process deals with considerations of the sequence as a whole. Since language operates within a larger system of cognition and communication, there are other aspects of the total system (e.g. mental states of the speaker, and feedback from the hearer) affecting the shape of the utterance in a way which cannot be attributed directly to explicit communicative goals.

3.4.3 In the course of design, structures are built at different levels of abstraction -- some representing linguistic objects such as words and phrases, others dealing with propositions and descriptions, and others representing plans for communication. There is no single structure which can be interpreted as a static "snapshot" capturing the entire meaning, since the structures undergo change during the production process.

3.5 In comprehending an utterance, the hearer is not simply "decoding" the speaker's linguistic structures, but is carrying out a broader process of analysis and inference having elements in common with the process applied to understanding other perceptual inputs.

3.5.1 The linguistic forms provide a starting point for an analysis process (again largely unconscious) which builds up a set of conceptual entities at several levels. The results of this analysis include:

Establishing the points of correspondence between the hearer's and speaker's world models by finding or creating conceptual entities which can be presumed to correspond to those existing for the speaker.

Drawing inferences which the speaker intended the hearer to draw.

Drawing inferences based on the content of the utterance, which may not have been specifically intended by the speaker. This includes a search for larger scale patterns which are consistent with the detailed propositional content.

Drawing inferences about the current state of the speaker, including the goals of the utterance, his focus of attention, and his state of knowledge (including his knowledge of the hearer).

3.5.2 The results of the hearer's analysis are the joint product of:

The utterance (including aspects such as intonation and tone of voice).

The knowledge possessed by the hearer, including his model of the relevant "world" and of the speaker.

The current state of the hearer, including immediate attention focus and his model of the conversation previous to the utterance.

The limitations of the hearer's processing capacity, including the degree of processing resources available for understanding the utterance. This can be limited both by other simultaneous tasks and by the arrival of a subsequent utterance.

3.5.3 The analysis takes into account both the "syntagmatic context" in which the utterance is set, and the "paradigmatic context" -- the alternatives to all of the choices which the speaker made in designing it. In many cases, the most important part of a communication is conveyed by the fact that a certain alternative was ***not*** chosen. One of the central aspects of the structure of a language is its imposition of finite systems of choice.

4. The scope of semantic theory

There are a number of distinctions conventionally drawn in linguistic theory which are not valid within a procedural paradigm. It is impossible to debate them at length here, but they can provide a focus for discussing the boundaries of what should be included within a semantic or linguistic theory.

4.1 Semantics vs. pragmatics

The previous discussion has largely ignored the traditional distinction between "semantics" and "pragmatics". This distinction is a central feature of the approach which has dominated the linguistic literature on semantics. This "formal structural" approach to semantics is based on looking only at structural relationships between words and utterances which hold independently of the context of their use, and can be formalized using mathematical logic. A typical statement of the central goals of narrow semantic theory is:

> ***It is fairly uncontroversial to say that an adequate semantic description must enable us to state, for each of the infinite number of sentences in a language, whether it is analytically true, whether it is contradictory or anomalous, with which sentences it enters into full or partial paraphrase relations, and with which sentences it enters into entailment relations.*** -- Wilson(1975) p.3.

> ***. ..the [larger] problem is to define the non-truth-conditional aspects of appropriateness. These seem to me to be clearly non-homogeneous, including reference to social conventions, discourse-convention, psychological considerations and contextual factors of many different types. Moreover, they seem to me in most, if not all cases, to be clearly non-linguistic, and certainly not matters of speaker-hearer's competence. For these reasons I would want to exclude them on principle from semantic description....*** -- Wilson(1975). p.14.

Those aspects of meaning which interact with context (social, textual, or psychological) are relegated to the realm-of "pragmatics". It is difficult, if not impossible, to state within this framework the concerns raised in section II below. Even if one were to accept the reduced goals implied by ignoring context, there- is no *a priori* reason to assume that a theory so defined is possible. There have been many criticisms (see for example Bolinger, 1968) of the attempt to reduce lexical meaning to formal structures combining primitive markers.

The procedural approach does not draw a sharp distinction between those symbol structures which represent "knowledge of the language", '*knowledge of language use" and "knowledge of the world". It begins with the hypothesis that there is much to be gained by studying what the representation and processing of these different areas of knowledge have in common.

4.2 Competence vs. performance

Procedural approaches to language have been eschewed in current linguistic theory, largely due to the dogma that theories based on considerations of language use or cognitive processing are "performance" theories, and therefore by nature secondary to more fundamental "competence" theories. Many linguists accept this notion rather uncritically. It is based on the observation that there are many factors affecting the details of language use which are not related to the essential structure of language. Theories should try to abstract away from the coughs, stumbles, and lapses of attention which punctuate normal human language use, just as a simple theory of mechanics postulates an imaginary friction-free world.

In avoiding processing models altogether, however, they are also adopting the rather questionable assumption that it is possible to formulate a characterization of language structure which is independent of the processes of language use, and that the resulting characterization will be simpler and logically prior to any characterization which is based explicitly on the processing. The truth of this hypothesis is an empirical question -- the analogous statement will be true for some sciences and false for others. There is some evidence that it is true for those details of syntax which lead to grammaticality judgments (although this is open to question for those areas where such judgments are fuzzy), but very little evidence that is true for language as a whole, particularly when meaning is taken into account.

Within the procedural approach, processing concepts are used to explain the structure of language at a level of abstraction which is still far removed from the minute "performance" issues which are not relevant to a first level understanding of the basic processes. At every stage of developing the theory it is necessary to decide which aspects of processing will be relevant, and which should be ignored in order to simplify the problems. But at each stage, the focus is on the set of "capacities" which form human competence to use a language. In this sense it can be called a "competence" theory. Many of these capacities are procedural -- they involve algorithms for processing perceptual inputs and remembered information, strategies for drawing inferences, and representations for storing knowledge in symbol structures. In this sense it involves "performance". The competence-performance distinction is meaningless since it is based on the assumption that language competence can be characterized in formalisms which do not deal with language use, and this assumption may very welt be false.

4.3 Language vs. communication

There is an extensive philosophical literature on the relationship between language and communication. It is sometimes argued (for example in Chomsky, 1975 p. 57 ff.) that communication is not a useful starting point for understanding language, since there are uses in which there is no single intended hearer, and others in which the linguistic forms exist only in the mind of their originator.

The model described in § 3 is applicable to a more general sense of "hearer" and "speaker" than the simple face to face exemplar used there. First of all, the speaker can have in mind an audience of which he or she is not directly cognizant (such as the set of readers of a paper or listeners to a radio broadcast). The detailed nature of the utterances will change, since the model of the hearer will be less precise and more abstracted. Assumptions about shared context must be limited to things assumed general to the culture or presented directly within the text. There is no direct linguistic or paralinguistic feedback about the current state of the receiver, and therefore the design of utterances must be based on inferences about how the expected receiver might be processing the utterances. However, the general nature of the knowledge involved and of the design process is essentially the same.

In the special case-of linguistic structures (vocalized or silent) designed only for the producer's use, the model is applicable in a closely related way. First, the act of producing internal utterances can be a step in distilling and storing information available to the language user. The attempt to encode thoughts into the limited set of categories and relational forms provided by a natural language will modify the content. In some cases, this process may filter out those aspects of the thoughts which are not amenable to linguistic encoding, changing the focus. In others, it may clarify and expand, as a result of the further reasoning and decisions which must be carried out to select lexical items and syntactic structures.

Further, many of the interpersonal goals discussed in §3 are applicable as internal goals within an information processing system. For example, specific linguistic forms can serve to trigger further inferences, emotional reactions, and associations which would not have been caused by the pre-formulated thought. By thinking in language, a person can produce useful effects on other areas of his or her own thought.

4.4 Speech acts vs. meanings

Part of the controversy about "language" and "communication'* has been formulated as a distinction between the "meaning" of an utterance, and the "speech act" effected in its use. In placing primary emphasis on the processes going on for speaker and hearer, the procedural approach is strongly oriented towards a speech act analysis. There is a secondary place (as one of a wide variety of symbol structures processed by speaker or hearer) for the more limited notion of "meaning" as the formal propositional content inherent in the utterance, independent of its use.

Procedural semantics differs from some speech act theories (see for example Searle 1970, Sadock 1974) in focussing on the entire complex of goals for the speaker rather than on a single primary "illocutionary act". An utterance is analyzed as simultaneously satisfying a set of communicative and interpersonal goals, signalled in different ways using the mechanisms of the language and of paralinguistic interaction. It is important to distinguish this multi-dimensionality of meaning (all of the different levels of meaning can be valid and simultaneously intended) from ambiguity, in which one of a possible set of meanings is to be selected.

The goals of a speech act need not map in a simple way onto the content of the utterance. They often demand an intended inference (see §3.4) by the hearer. Such inferences can be based on knowledge which ranges over a spectrum from linguistic convention (such as the standard interpretation of "How do you do?") through those which are conventional but have some pragmatic basis (such as "Can you pass the salt?" and others discussed by Gordon and Lakoff 1971, Searle 1970, etc.) to those demanding special knowledge of the world (as in the case of the parent who says to a misbehaving child "Do you remember that the circus is coming to town next week?"). In many cases, the speech act conveys information at a totally different level than its apparent form (as in saying "Oh, you're here" to someone who clearly knows his own location).

Part 11. A procedural approach to some specific problems in semantics

5. The nature of lexical meanings

5.1 The problem of delimiting meaning

The word "bachelor" has been used in many discussions of semantics, since (save for obscure meanings involving aquatic mammals and medieval chivalry) it seems to have a formally tractable meaning which can be paraphrased "an adult human male who has never been married". Traditional theories of semantics deal with tasks such as determining whether the sentence "my bachelor uncle is unmarried" is analytic. In the realistic use of the word, there are-many problems which are not as simply stated and formalized. Consider the following exchange:

> Host: I'm having a big party next weekend. Do you know any nice bachelors I could invite?
>
> Friend: Yes, I know this fellow X...

The problem is to decide, given the facts below, for which values of *X* the response would be a reasonable answer in light of the normal meaning of the word "bachelor". A simple test is to ask for which ones the host might fairly complain "You lied. You said *X* was a bachelor.*':

A: Arthur has been living happily with Alice for the last five years. They have a two year old daughter, and have never officially married.

B: Bruce was going to be drafted, so he arranged with his friend Barbara to have a justice of the peace marry them so he would be exempt. They have never lived together. He dates a number of women, and plans to have the marriage annulled as soon as he finds someone he wants to marry.

C: Charlie is 17 years old. He lives at home with his parents and is in high school.

D: David is 17 years old. He left home at 13, started a small business, and is now a successful young entrepreneur leading a playboy's life style in his penthouse apartment.

E: Eli and Edgar are homosexual lovers who have been living together for many years.

F: Faisal is allowed by the law of his native Abu Dhabi to have three wives. He currently has two and is interested in meeting another potential fiancee.

G: Father Gregory is the bishop of the Catholic cathedral at Groton upon Thames.

5.2 Words as symbols for abstract exemplars

The cast of characters in $5.1 could be extended indefinitely, and in each case there are problems in deciding whether the word "bachelor" could appropriately be applied. In normal use, a word does not convey a clearly definable combination of primitive propositions, but evokes an *exemplar* which posesses a number of properties. This exemplar is not a specific individual in the experience of the language user, but is more abstract, representing a conflation of typical properties. A prototypical bachelor can be described as:

1. a person
2. a male
3. an adult
4. not currently officially married
5. not in a marriage-like living situation
6. potentially marriageable
7. leading a bachelor-like life style
8. not having been married previously
9. having an intention, at least temporarily, not to marry
10. .....

Each of the men described above fits some but not all of these characterizations. Except for narrow legalistic contexts, there is no significant sense in which a subset of the characteristics can be singled out as the "central meaning" of the word. In fact, among native English speakers there is little agreement about whether someone who has been previously married can properly be called a "bachelor" and fairly good agreement that it should not apply someone who is not potentially marriagable (e.g. has taken a vow of celibacy).

Not only is this list open-ended, but the individual terms are themselves not definable in terms of primitive notions. In reducing the meaning of "bachelor" to a formula involving "adult" or *'potentially marriageable", one is led into describing these in terms of exemplars as well. "Adult" cannot be defined in terms of years of age for any but technical legal purposes and in fact even in this restricted sense, it is defined differently for different aspects of the law. Phrases such as "marriage-like living situation" and "bachelor-like life style" reflect directly in their syntactic form the intention to convey stereotyped exemplars rather than formal definitions. There have been attempts to use quantificational and statistical methods to provide a more precise formalization of "fuzzy" concepts. Labov (1974), for example has attempted to map out the characteristics of a population of speakers in distinguishing between exemplars such as "cup" and "glass" as the characteristics of the object being described are varied along a number of dimensions. However, these still provide an overly uniform mathematical structure on the different characteristics of the exemplar.

An exemplar brings together different kinds of characteristics which have quite distinct properties in selecting lexical items. There is an intuitive impression, for example, that the meaning of "bachelor" includes some absolutely necessary conditions, and is being metaphorically stretched when applied to individuals violating them, (as in talking about "my bachelor aunt" or responding to the question "Has your dog ever sired puppies?*' with "No, he's a bachelor'*). But even for seemingly straightforward predicates such as "human" and "male", it is easy to generate examples where notions of primitive predication are inadequate (for example in situations involving eunuchs and near-human monsters).

There is also a tendency (as G. Lakoff, 1972 has pointed out) to use "hedges" when applying a term to cases where characteristics which are non-essential but typical are being connoted (**Anthony's wife is away on business trips so often that he's a *regular* bachelor", or "That little episode with Sarah confirmed my belief that Carl is a *true* bachelor."). The ways in which different hedges emphasize different characteristics is highly complex, and it is impossible to draw a sharp line between "essential" and "secondary" or "defining" and "characteristic'* properties.

The complexity of all of these problems should have an impact on the nature of semantic theory. A theory which ignores them cannot claim to be a full explication of the meaning of words. This case is not at all pathological: in fact, "bachelor" has been used as an example because its definition seems so precise compared to words like "friend" and "game".

5.3 Informal prototype semantics

The idea of a semantics based on exemplars has been developed by Fillmore (1975) and other linguists. They have generally avoided rigorous formalization of the theory while trying to gain maximal breadth in the kind of data explored. Research in this style is primarily anecdotal, describing examples and pointing out interesting phenomena in the ways words are used. There is a good deal of work on prototypes in psychology (see, for example Rosch1975), and in the philosophical literature similar issues have discussed for many years, beginning with Wittgenstein's discussion of the word "game" (Wittgenstein, 1953).

5.4 A procedural framework for a prototype semantics based on description matching

There is not yet a fully developed procedural approach to lexical meaning, but there are beginnings of a theory, using a number of computational ideas applicable to situations involving the "matching" of exemplars to individuals. The notion of matching is related to that of testing the truth of a set of propositions applied to an entity, but differs from standard truth-theoretic notions by dealing explicitly with *processing resources* and *differential accessibility* of information associated with stored entities and with the language user's current state. Due to the explicit concern with amount of computation, matching cannot be reduced to truth conditional concepts, except by axiomatizing the detailed behavior of the complete cognitive processing system.

Much of the current research in cognitive science (see for example Bobrow and Winograd 1977, Norman, Rumelhart, and the LNR group 1975, Minsky 1975, Newell and Simon 1972) deals with these issues. The discussion here only touches on those aspects most relevant to lexical choice. The same notions of matching can be applied to other parts of the language process, including determining the intended reference of phrases, as discussed in §6.

5.4.1 A word can have associated with it in the language user's store of linguistic knowledge an entity which is an exemplar, described in the same representation used for other conceptual entities (e.g. real-world objects). Not all words have a simple exemplar (e.g. function words such as "the", logical connectives such as "although") and some words have more than one. In selecting a lexical item, the speaker looks for the word whose associated exemplar has the best fit with the information to be conveyed. The hearer is choosing (or creating) the conceptual entity which best corresponds to the resulting choice. On both sides, the notion of "best fit" is relative to the current context. In making a lexical choice, the speaker uses his or her model of the hearer to select an item which will lead to desired inferences at all levels, not only in selecting the right entity. (e.g. in choosing to use the word "thug" rather than "accused assailant" in describing a robbery or choosing "shut up" vs. "be quiet" in formulating a request).

5.4.2 There is a class of cognitive "matching" processes which operate on the descriptions (symbol structures) available for two entities, looking for correspondences and differences. These processes are used in many different aspects of reasoning, and are not specific to language use. At the simplest level, a match corresponds to checking whether a description (made up of symbol structures) is identical to one associated

with an entity. There can be a concept of "bachelor", and it is possible that the speaker or hearer has a stored structure which predicates it directly of an individual. At the next level, subsidiary processes can be triggered. For example, two descriptions can be matched successfully if one of them contains information which is directly inferrable from general knowledge and information explicitly contained in the other. One common form of subsidiary processing is the attempt to match two entities which play analogous roles in the two original descriptions. In comparing descriptions of two acts, for example, the matching process might include comparing previously known properties of the two actors.

5.4.3. The result of a matching process is not a simple true/false answer. It can be stated in its most general form as: "Given the set of alternatives which I am currently considering (due to the context, including my model of the other participant), and given the amount of processing which can be devoted to selecting one (in light of all the other mental processes going on, both linguistic and non-linguistic), and looking in order at those stored structures which are most accessible in the current context, here is the best match, here is the degree to which it seems to hold, and here are the specific detailed places where match was not found."

5.4.4 The process of matching is *resource limited.* (See Norman and Bobrow, 1975 for a discussion of this concept). In any specific instance, the process will go on for a finite amount of time accumulating evidence for whether the match is applicable. In some cases, there will be strong positive or negative evidence, while in others, there will be only a loose weighting. The same match inputs could result in different (even strongly contradictory) results given different amounts of processing resources. For example, a strong mismatch might exist, but in a part of the symbol structure which would not be checked until far along in the processing. A high degree of match would be found by any process which did nor have enough resources to get to that part.

5.4.5 The set of structures associated with a word or entity are not uniformly accessible. There are differences which are not reflections of the truth or degree of belief of the associated propositions, but of a wide variety of other factors having to do with motivations, history of learning and use, emotions, etc.

5.4.6 The selection of the order in which sub&structures of the description will be compared is a function of their current accessibility, which depends both on the form in which they are stored and the current context. The relative accessibility of two structures can be arbitrarily different for different states of the processor (e.g. current goals, focus of attention, set of words recently heard, etc.).

5.4.7 In performing a match, it is possible for the context to override the normal default decisions as to which properties of the exemplar should be dealt with, and how contradictions should be treated. For example, in recognizing metaphorical uses and hedges, the hearer is in a state where obvious mismatches are taken not as a signal that the match process as a whole should fail, but that it should be carried out more assiduously (looking for interesting inferences) on some of the other properties.

6. Problems of reference

This section contrasts the procedural approach to semantics with the more common truth conditional or referential notions. As with the examples of $5, the issue at stake is not a

better solution to the problems as they have been posed, but an attempt to redefine the nature and structure of the relevant problems. Morgan (1975) discusses the problems of reference in a way closely related to the procedural approach, and several of the examples are due to him. Within the overall framework of a procedural approach, it is possible to provide formal content for many of the informal concepts he proposes.

6.1 Definite and indefinite reference

(6-1) *When we got home, the sherbet was gone, and the empty carton was in the sink.*

(6-2) The reason I didn't do it was that I got a phone call.

(6-3) Examining the cabinet, we noticed that a door was marred.

(6-4) When the presidential plane arrived at Dulles airport, the reporters were greeted by (a / the) sullen and snappish Henry Kissinger.

(6-5) The unicorn is a mythical beast.

In a simple truth-conditional view of semantics the appropriate questions deal with the truth conditions (on the model, not the state of the language user) under which the use of a definite determiner is correct. Russell (1919) hypothesized that "...propositions about 'the so-and-so' always imply the corresponding propositions about 'a so-and-so' with the addendum that there is not more than one so-and-so." (in Rosenberg and Travis, p. 172.) This might apply to phrases such as "the presidential plane" in sentence 6-4, but as has been pointed out by Strawson(l950) and Donnellan (1966), it is quite inadequate for dealing with all of the other cases in these examples. In most of the occurrences of the definite determiner "the" and the indefinite "a", the meaning would be significantly changed, or the sentence would become unacceptable if the other were used (using the appropriate indefinite form for plural and partitives).

The problem is to find a formal way of talking about the conditions (including current attention focus and goals) in the minds of speaker and hearer which affect the selection and interpretation of determiners. Most philosophical discussions dealing with definite referring phrases have not focussed on these more subtle issues of definite reference, on the assumption that they represent second order uses of a device with a simpler basic structure, and that the underlying device could be explained in terms of truth conditions, without explicit reference to the communicative desires of the speaker or knowledge state of the hearer.

6.2 The use of definite referring phrases

The sentences in §6.1 use definite determiners to convey several distinct messages:

6.2.1 Reference to objects implied by standard knowledge: One of the most common uses of definite referring phrases is exemplified by the phrases *"the empty carton"* (6-1), *"the sink"* (6-1), and *"the reporters" (6-4).* In each case, there is some piece of knowledge about the world which enables the hearer to pick out an entity (which may be a set) which can be reasonably inferred to exist, and to which the speaker is referring. The hearer does not need to have previous knowledge of the object (as with the carton), nor does it need to be uniquely specifiable -- most houses have more than one sink, but it can be inferred that the reference is to a sink in the kitchen.

6.2.2 Reference to previously established context: Many uses of definite phrases assume that there is a previously established context which the speaker expects the hearer to know about. The phrases *"the* ***sherbet"*** (6-l) and *"the* ***cabinet" (6-3)*** are used in a context where their intended reference is unclear without further information about where the sentence was uttered. This context might include prior direct knowledge (e.g. the hearer was there when the sherbet was purchased) or the linguistic context (e.g. an earlier sentence ***"Last week just before our trip, we bought a gallon of strawberry sherbet..."***). It is not, however, part of what would be expected of a native speaker's knowledge independent of the situation.

6.2.3 Implied uniqueness: In using the phrase *"The* ***reason I didn't do it," (6-2)*** the speaker conveys a new piece of information -- that there was only one reason. In ***"a door was marred,"*** the choice not to use the definite "the" invites the inference that the particular cabinet had more than one door. This notion of uniqueness is extended in selecting a determiner in the phrases *"the* ***sullen and snappish Henry Kissinger,"*** and *"a* ***sullen and snappish Henry Kissinger" (6-4).*** The choice of an indefinite "a" implies that the referent is not specified uniquely -- i.e. that there is more than one Henry Kissinger (or more than one persona which he adopts), and that the sentence refers to a specific one of them. The result is to focus attention on the distinctive characteristics of this particular persona, an effect intended by the speaker.

6.3 The basis for a procedural theory of definite reference

6.3.1 The speaker and hearer each have a set of previously established mental entities. These can correspond to actual objects and events in the world, to imaginary or fictional objects and events, to abstractions, etc. (see §2.2)

6.3.2 For each such entity a person has a set of symbol structures representing descriptions which he believes to apply to it. These are the structures used in the matching process described in §5.4.

6.3.3 Using linguistic knowledge and general knowledge about the world, the hearer can infer that a description conveyed in a phrase applies to an entity already existing in the model, or that a new entity should be added to it.

6.3.4 Speaker and hearer each have a model (partial and not necessarily correct) of the set of entities, descriptions, and reasoning processes available to the other. This model changes during the course of interacting, as a result both of what is said and the shared situational context.

6.3.5 In choosing to use a definite referring phrase, the speaker is conveying to the hearer:

> ***There is a unique entity in my mental world about which I want to say something.***
>
> ***The descriptive phrase is sufficient to allow you to find or create a unique entity within your world model, using all the context and general knowledge (about the world, the rules of conversation, and my own state) I can presume you to have on the basis of my current model of you.***

This definition depends on a procedural approach in that it calls on the "current context" of speaker and hearer. The inferences which a speaker can presume the hearer to make depend not only on logical propositions, but on issues of attention, memory access, and goals as discussed in the previous section: The definition also involves more explicit formal levels

than the simple truth-theoretic notions. First, it describes definite reference in terms of a message conveyed from hearer to speaker about the entities and descriptions in their respective mental worlds, rather than propositions about real world objects. Second, it refers explicitly to the fact that there is an interpretive process going on in understanding -- the speaker can base a choice of determiners on inferences about the processes which the hearer will use in interpreting the phrase, and the hearer can base an interpretation on inferences about the state and processes of the speaker. This is particularly important in cases such as the use of "a door" in sentence 6-3, where the speaker's choice not to use a definite reference can be interpreted as conveying the fact that there was more than one door.

In the simplest situation, the definition given in this section is equivalent to a referential definition. There is an entity actually existing in the world. Speaker and hearer each have a mental entity corresponding to it, and descriptions which correspond to its actual properties. The description appearing in the definite phrase applies to that entity and to no other in the world or in the mental world of hearer or speaker.

In analyzing other uses of definite referring phrases, the complexities of the procedural model become much more relevant and necessary. In the great majority of uses of definite phrases, it is not the case that "the descriptive phrase applies to a unique object", but it is the case that it "is sufficient to allow you to establish a unique referent...*using all* ***the context and general knowledge...".*** Several of the above examples illustrate this point. Much of the classical discussion of ***sense*** and ***reference*** has a straightforward reinterpretation in this light. The sentences "The morning star is the evening star," or "Scott is the author of Waverly" are produced by a speaker who has a single conceptual entity with two descriptions, and a model that the hearer has two distinct entities associated with these descriptions. The sentence "Unicorns have never existed" is spoken by someone who has a conceptual entity of "unicorn" with an associated description that such beasts do not actually exist, and a model of a hearer who has a corresponding entity which is lacking this description.

In cases of reported thoughts (or modals which refer to other possible worlds) ambiguities can arise due to the multiplicity of models. In the sentence "John thought the murderer was Bill's father." there as an ambiguity resting on whether the description "Bill's father" is associated with the entity in the world model of the speaker or in the speaker's model of John's model of the world.

The ***referential*** and ***ascriptive*** use of referring phrases discussed by Strawson (1950) differ in the correspondence between the entities in the speaker's conceptual world and those in the real world. A phrase such as "the king of France" refers to a unique entity in the speaker's conceptual world, independently of whether there is a real king of France, and whether the speaker believes there to actually be one. It is easy to invent circumstances in wh-ich the phrase might be used in which neither or only one of the these circumstances hojds. For example, there are cases such as "Who does deGaulle think he is, the king of France?" in which both speaker and hearer recognize there is no such entity, but can infer properties from its description.

Donellan's distinction between ***referential and attributive*** uses is more oriented towards the resulting conceptual world of the hearer. He states that (in Rosenberg and Travis, p.211) "...in the referential use as opposed to the attributive, there is a right thing to be picked out by the audience and its being the right thing is not simply a function of its fitting the description." An attributive reference succeeds if the hearer creates an appropriate conceptual entity corresponding to the speaker's, while a referential use of the same phrase succeeds only if the conceptual entity for the hearer corresponds to the same real world object as does the conceptual entity being referred to by the speaker.

In an extreme case, neither speaker nor hearer need believe the applicability of the description. Donellan points to an example where there is an imposter on the throne, and he is known to be an imposter both by the palace guard and by the visitor who asks "Is the king in his countinghouse?". This example points out that the issue in understanding what is happening in definite reference is not really one of truth conditions applied to the world, but one of establishing a correspondence between conceptual entities of the speaker and hearer. The procedural approach begins from this foundation.

One might imagine applying truth theoretic ideas to cognitive models of speaker and hearer, allowing predicates over cognitive entities and building up a set of axioms formalizing the nature of these entities, and their relationships to objects and predicates on the world. The procedural approach formalizes notions such as "current context", "could be expected to infer", and *'attention focus" in terms of the processes by which cognitive state changes as a person comprehends or produces utterances. These processes interact in complex ways, subject to influences of resource limitation and differential accessibility as discussed in §5.4 An axiomatization which produced equivalent results in determining the meaning of referential phrases would have to contain within it a theory of the human language processor.

6.4 The problem of--. identifying conceptual entities

An analysis of definite reference in procedural terms does not solve the fundamental epistemological and ontological problems implied in the notion of "conceptual entity". Some obvious considerations include:

6.4.1 Conceptual entities need not correspond to real world entities, e.g., fictional and imaginary characters.

6.4.2 In the case where a conceptual entity corresponds to a real-world entrty, the descriptions a person attributes to it may not be consistently applicable to the real-world object to which it corresponds.

6.4.3 Two people can have conceptual entities which they believe to correspond to the same object (i.e. each has a mental model of the other which includes an appropriate conceptual entity), even if there is no corresponding real object. For example, two people can agree that they are talking about the same "Moses" or "Santa Claus" without having identical or even compatible models. In cases of miscommunication, they can have entities corresponding to distinct objects without being aware of that fact.

6.4.4 A person can have a conceptual entity which he believes to correspond to a unique individual, while being aware that he does not know which real-world individual it is, as conveyed by the opaque reading of the definite reference in ***"The man that I marry will like taking care of children."***

6.4.5 A conceptual entity can correspond to an abstract prototype rather than an individual (as in "the unicorn" in sentence 6-S).

6.4.6 A specific manifestation or persona of an individual can be a separate conceptual entity, as in sentence 6-4.

These problems, however, are not problems of definite reference or even of language use. They are representation problems which have been addressed in many different forms in building cognitive models, and are at the heart of much research in cognitive science. There are further specifically linguistic problems involved in sorting out instances of definite references from other phrases which use some of the same signalling words, such as

"the" in "Peaches are selling for three dollars the bushel".

7. Explicit procedural modifiers and connectives

This section discusses some problematic sentential connectives and adverbs as a final brief example of semantic problems which may be amenable to a procedural approach. Sentences 7-l to 7-5 contain words whose meaning can be explicated only via explicit reference to the models the speaker and hearer have of each other.

(7-l) He's ***a politician,*** but ***he's honest.***

(7-2) Even ***Gerald understood what the memo implied.***

(7-3) ***He*** even ***made his bed.***

(7-4) ***If*** ***you encourage him*** some, ***he'll explain his whole theory.***

(7-5) ***If you encourage him*** any, ***he'll explain his whole theory.***

In each of these sentences, a part of the conveyed meaning is a statement by the speaker about the hearer. "But" can be glossed as "The following statement contradicts an inference I would expect you to make." "Even" conveys "There are other entities inferrable from the context which could be substituted for the one modified by "even". You are more likely to expect the substituted statement to be true than one actually given." In applying this to sentences like 7-3 it is necessary to analyze what is being asserted in ways which may be orthogonal to the standard logical decomposition (such as the modality "He did soniething" and the details of what the event was) 4s with the examples of §5 and §6 this is not an isolated problem related to language use, but ties in with much more general issues having to do with they way that facts and events are represented in the human symbol system. In 7-4 and 7-5, the speaker's choice of "some" versus "any" conveys a message indicating whether the speaker is advising the hearer to take the hypothesized action. This implied message, rather than the statement of a logical condition, is in fact primary to the utterance.

Providing informal glosses for words and constructions such as those above is useful only if phrases such as "I expect you to infer..." can be given a precise meaning in terms of processes and symbol structures. This formalization of cognitive context is one of the main goals of procedural semantics.

Part III. Goals for a procedural understanding of language

8. The focus of linguistic theory

The approach suggested in this paper goes against many of the canons of accepted meta-theory in contemporary linguistics. The major departure is a focus on the cognitive process, rather than on the linguistic objects produced and interpreted. The form and meaning of an utterance is analyzed in terms of the goals and mental state of the speaker and hearer. Some linguists (see, for example Morgan 1975) have also suggested this shift of emphasis away from formal descriptions of linguistic objects towards formal means for representing the cognitive processing of language users, but it is not part of the current standard theory. This section discusses the ways in which a procedural perspective on language can provide a basis for explanation.

8.1 The basis for a procedural theory

The major points of Parts I and II which have implications for the nature of acceptable linguistic theories can be summarized:

8.1.1 The appropriate focus for a theory of language is on the cognitive process, rather than on the linguistic objects produced and interpreted. The specific structure and content of the observable utterances is important, but in a secondary role as one part of the data generated and used in the ongoing processes.

8.1.2 Context is of primary importance. and is best formulated in terms of cognitive structures, rather than the linguistic text or facts about the situation in which an utterance is produced. It indirectly (through the speaker's and hearer's models) includes those aspects of context traditionally called linguistic, social, and pragmatic.

8.1.3 It is possible to scientifically study the processes involved in cognition, and in particular of language use. $2 discusses the general shape of this theory and the justification for believing it is possible.

8.1.4 There are several distinct levels of analysis of the mechanisms involved in language. None of them can be reduced to a set of consequences of principles at lower levels. Many current theoretical arguments are based, at least implicitly, on setting out one level as "theoretically interesting", and assuming that the others are secondary.

8.2 Levels of explanation

Imagine a group of alien scientists arriving on earth and setting out to understand motorized vehicles. Due to their lack of appropriate tools, they are unable to open up the hoods and look inside, and therefore must base their theories on external behavior. They do have some vehicles they can experiment on, but these experiments can only involve manipulating externally available variables, such as which pedals and buttons are pushed in what sequence, what is poured into the various openings, etc. They even have a few cases of vehicles which were damaged in accidents and have partially impaired function. What would be the nature of the "automotive theory" they should seek to develop?

At first it might seem that the necessary theories are those of physics. An automobile, like any physical device of similar size, operates according the principles of newtonian mechanics and classical thermodynamics. But this is clearly insufficient to explain how the automobile works. A physicist with a complete grasp of all of the relevant theories may have no idea whatsoever about the behavior of anautomobile. The problem isn't that any one principle is missing, but that the overall performance is the result of many interacting mechanisms.

A physicist is missing the level of explanation which a mechanic has -- just what the pieces are, how they fit together, and how they are supposed to function. The mechanic can explain the automobile at a completely different level, which is not reducible to physics, and is at least as important in explaining the behavior of the specific class of objects being studied. This level itself includes descriptions of both the forms and the functions of the various parts and subsystems. It can include alternative conceptualizations of how the system should be divided. There might be one useful division based on the location of parts in the chassis, another separating hydraulic, mechanical, and electrical subsystems, and another distinguishing the power train, braking system' etc.

In addition to levels of explanation corresponding to physics and mechanical design specifications, there are others relating to larger systems, extending beyond the individual automobile. Many features can only be "explained" in terms of an evolutionary process in which new materials were added' and new concerns caused modifications. There are many things in an automobile which do not make sense except in the context of a history in which concerns such as pollution control had a belated influence on what was already a stable design. More global explanations as to why cars have specific characteristics of size, fuel use, etc. must be couched in terms of the social and economic systems in which they were developed. The nature of roads, of fossil fuel resources, and of human family structure have all had an influence which is visible in the physical design. Of course in this larger perspective the influence is mutual. The nature of the automobile has shaped the system as well as being shaped by it.

In applying this analogy to language, the human language facility is being compared to a complex artifactual object whose behavior is explained in terms of both general underlying principles and specific design of the mechanism. Much of the work in procedural semantics is concerned with the design and structure of the processes which take place within the human language user, assuming that most of the observable regularities are to be explained at that level rather than as consequences deducible from basic principles. The role of general theories more akin to those of physics is discussed below.

8.3. Axionrat ic deductive theories

The model which has dominated much of science is the axiomatic deductive theory, exemplified by classical physics (or at least the popular image of it). The scientist discovers a small set of fundamental axioms, or "laws" operating within a formalism (a "calculus") for describing conditions and events. Given a description of a set of conditions, the laws can be applied to predict how those conditions will change over time, or to specify other conditions which must be co-occurent. The power and elegance of such a theory comes from the fact that the underlying set of laws is small, and the intricacies of behavior come from the complexities of the settings in which they are applied.

This model has had a tremendous influence on linguistics in its preoccupation with parsimony of mechanisms. A physical theory which describes the motions of the planets and the falling of earthly bodies with one set of mechanisms is clearly superior to one which demands separate analyses. Much of the justification for the details of linguistic theories has been in the form "...this theory is better because it is able to account with one single mechanism for the following observations..." or "...this theoretical framework is better

because it has only one type of constraint on the application of rules, instead of..."

Of course, there must be a notion of simplicity, and an attempt to find the most parsimonious theories. But if the phenomena being studied are in fact the result of interactions between complex mechanisms, the scale of expected complexity of the theories must be revised. Consider trying to account for the relation of an automobile's speed to accelerator pressure and time. The actual data are complex, and result from: the response characteristics of the engine, which are themselves non-linear; the type of transmission, and the detailed adjustments which cause it to shift on the basis of variables including accelerator pressure, engine vacuum, and speed; the pollution control mechanisms which adjust engine speed in an attempt reduce emissions.

Any attempt to analyze the data will be unsuccessful if the scientist assumes that the regularities are directly indicative of deep underlying principles. The data results from interactions between the different mechanisms, each of which in turn is a complex mechanism. Theories of mechanics and thermodynamics are relevant only at the very lowest level of structure.

One way to reduce complexity of this sort is to average the data from a large number of different observations, and try to fit mathematical formulas to the results. There have been many "theories" of this form in psychology (for example, the behaviorist "learning theories"), and to a lesser degree in linguistics. This approach has been useful in those fields (e.g. economics) where behavior in the large is of interest, and the detailed mechanisms are beyond the reach of current techniques. In linguistics, it can be applied to studying large-scale phenomena such as vocabulary distribution and language evolution, but it does not even attempt to answer the kinds of questions about the nature of the human language user which form the focus of the procedural approach.

Complexity can be attacked more directly by combining an analysis at the level of axiomatic theory with one at the level of a "blueprint" of the system under study. Other sciences of complex biological systems deal with them in terms of "physiology" and "anatomy", laying out in detail the observed structures and processes. It seems unlikely that cognitive theory can exist without the same sort of detailed mapping of mechanisms. Attempts at understanding the human being as a chemical and physical device have led to the conclusion that it is many times more complex than human-made devices such as automobiles. Any adequate explanation of language and cognition will be based on theories and mechanisms which are in turn orders of magnitude more complex than those of the kidney or the circulatory system. It seems highly implausible that there are separable mechanisms corresponding to a single "rule type" or information corresponding to a small number of independent "components" or "faculties".

The interesting (and parsimonious) general theories will be at the level of computation, communication, and cognitive processing. A specific device such as the human language processor will bear the same relationship to them that the automobile bears to physics. A physicist is not interested in predicting the detailed behavior of an automobile, since it is mereiy a specific idiosyncratic combination of the deeper principles. The case of language study differs in that although the human language capacity is merely one specific device which exhibits behavior based on deeper principles, it is one in which we have a great deal of interest, and is the only one of such a degree of complexity which we have available to study.

8.4. Understanding the design of a complex system

In explaining the automobile, scientists need hypotheses about the different constituents of the system -- a driving engine, a variable power transmission train, a fuel supply system, a braking system, etc. To some degree there are natural divisions implied by the external

differences (e.g. the different pedals to be pushed), and behaviors (e.g. slowing down vs. speeding up). However, these divisions are not isomorphic to clear functional or anatomical divisions in the mechanisms.

The problem of finding the right modularity is central to building a model of the human language user. Our goal is to find divisions of the system into modules whose explanation is simplified by describing their internal structuring and their interactions separately. There may not be separate "semantic" or "syntactic" mechanisms, or even a separate "language faculty" as Chomsky (1975) proposes. These imposed categories can serve as a first cut to simplify analysis, but it is an empirical, and still very open question, as to whether they represent divisions which help provide coherent descriptions of the larger system of language and cognition.

8.5 Verification of hypotheses

In trying to infer the design of a complex system, the conventional notions of theory testing and experimentation need to be extended. It is generally not possible to formulate a separable piece of the theory, find a critical experiment which will potentially falsify it, and then carry out the experiment. Before performing experiments to determine the correct model, the first step is the development of potentially adequate models. At this stage of the science, we are not yet ready to decide between alternatives, but have only begun to come up with ideas for what the mechanisms might be like. The strategy which is being taken within the procedural approach is to build artificial (computer based) systems which carry out limited language comprehension tasks. They provide an experimental adequacy test which demonstrates the strengths and lacks in the hypothesized mechanisms. This test involves a larger system into which the components must fit, since a proposed mechanism for one component (e.g. syntax) must be consistent with its use in the actual language process.

The resulting simulation models are not intended as literal step by step models of activity in a human brain. As mentioned above, the goal is to build models of processing which are the same as those in the human language user at some natural level of description. A successful model can differ in detail, or along some dimensions of description while being similar at a different level of structure, concerned with the major components and their interconnections. The validity of a model is not measurable in terms of a simple notion of fit.

The appropriate level of description may well cut across the distinction between physical and conceptual structure. One of the fundamental insights of computation theory is that there is no sharp line between data and process. In any physical symbol system, there is a level of "hard-wired" built-in processors, which operate on data whose structure is determined by the nature of the machine. However, this system can be used as an interpreter, operating on data structures built up out of the primitives available, and can thereby simulate a machine with a different set of processes and a different organization for its data. Some of the deepest results in computation theory relate to this equivalence of machines (see Minsky 1967 for a survey). It is impossible to discuss the consequences of this duality fully here. For an extended discussion, see Newell, 1972.

The human language processor has inherent properties which constrain the kinds of representation it can use, the types of processing strategy. the limits of attention, the nature of the learning it can achieve, and many other detailed characteristics. A comprehensive cognitive theory will include an explanation of these details. However, it is possible to simulate the detailed nature of a physical symbol system on one with different physical characteristics. In doing so, it is necessary to keep in mind that there may turn out to be fundamental differences which shape the nature of the computations which can be done. However at the stage of development so far, the main differences are those visible in detailed process characteristics such as the relative time and effort needed to carry out

different operations, and the sorts of errors naturally made.

If the procedural approach is successful, it will eventually be possible to describe the mechanisms at such a level of detail that there will be a verifiable fit with many aspects of detailed human performance. There are some attempts now at experiments which test small subparts (see for example Kaplan, 1975) but we are nowhere near having explanations which cover language processing as a whole, including meaning. More detailed testing at this point runs the risk of cutting the analysis along the wrong lines. One could imagine doing detailed experiments (involving careful measurements of time, forces, etc.) on automobiles by turning the ignition key and watching their behavior without putting in any gasoline. This might be seen as a "simpler" experimental situation. However, the details of interaction between the transmission, starter, and engine compression would create a good deal of incomprehensible complexity without the additional clarification of understanding the larger context in which these mechanisms function.

Lacking the ability to carry out precise experimental verifications for hypotheses, it is necessary to take a different criterion for selecting and rejecting alternatives. The major heuristic is that of "independent justification". Within generative linguistics, this notion is one of those discussed under the rather ill-defined concept of "explanatory adequacy". If a single mechanism explains a set of distinct phenomena, it is more highly valued than one which must be generated solely to handle one part of the data. The postulation of molecules in physics gained great strength from its usefulness in explaining both chemistry and thermodynamics.

The procedural approach differs from most of current generative linguistics in the breadth of phenomena over which this heuristic is to be applied. A mechanism which accounts for some syntactic data is more highly valued if it can also be used to account for semantic phenomena. A hypothesis about about language production or comprehension gains plausibility if it can also be used in explaining other cognitive processes, such as perception or problem solving. We are far from having a small set of concepts whose usefulness is as broad as that of molecules or wave motions. The future of cognitive science lies in finding them.

8.6 Larger frameworks within which to explain language

Our explanation of language would not be complete even with a full analysis (both a blueprint and an underlying cognitive theory) of the processes which go on in a single adult language user. There are larger contexts into which this process must be set, each with its own type of explanation.

8.6.1 Theories of inherent limitations

Following Chomsky (1965), many current theoretical linguists have viewed the ultimate goal of their study as the discovery of the "formal linguistic universals" which characterize the abstract conditions that must be satisfied by any generative grammar. Work of this type often makes explicit reference to results in the theory of formal grammars and computation theory. Formal results on the generative power of grammars have been sought as guides towards finding the best characterization language competence.

The experience in computer science (as distinct from abstract automata theory) has been that these abstract characterizations of computational "power" are not fruitful sources of explanation for the properties of complex computational systems. Any system of interest is theoretically universal, and its interesting properties are determined by its specific structure, not its capabilities in the abstract limit.

Returning to the automotive allegory, one could imagine theories which set abstract limitations on the capacities of vehicles. No car could ever travel faster than the speed of light, or accelerate in the absence of outside forces without using up some of its mass. These results are of great scientific importance, as are results such as the Godel theorem on the universality of Turing machines. But they are not a basis for explaining automobiles, and mathematical results on formal systems are not a general basis for explaining language.

8.6.2 Functional explanations for the design

In studying a system which is a conscious technological artifact, one attributes its design to a set of explicit decisions based on the functions it is intended to serve. One "explains" a specific mechanism by knowing why it was included, and why it was done the way it was, instead of in alternative ways. In studying natural systems, one often attributes to nature (and the evolutionary process) the same kind of functional motivation. A mechanism is the way it is in order to serve some useful purpose in the overall functioning of the organism. It is impossible to "explain" all the details of a system on a functional basis. There is no simple reason why a person has precisely ten fingers, or why the words "dog" and "day" begin with the same phoneme. On the other hand, there are many ways in which the detailed mechanisms are shaped by the needs they must serve.

The procedural approach is inherently closer to functional explanation than generative linguistics since it deals explicitly with processes. However. there are independent levels of explanation concerning the range of possible alternative systems and the pressures which shape the exact nature of the human language system. Many linguists (e.g. the Czech school, and many of those represented in Grossman, San and Vance 1975) are concerned with this sort of analysis, and it is complementary to a procedural approach. The traditional concerns of linguistics with cross-linguistic comparisons and universal grammar are also attempts to explain specific features of language use within a broader context.

8.6.3 Evolution and learning

Many of the most interesting facts about language have to do with the way it changes over time, both for an individual person, and for a society as a whole. Many of the details of the human language processor must be understood in the context of language learning and change, just as many of the physiological and anatomical details of an organism can be understood in terms of its ontogeny and phylogeny. So far, there has been little research on the connections between these macro-processes and the procedural details of language use. As with functional analysis, it seems that the basic orientation towards cognitive processes will make such connections more natural and profitable than with current formalisms. The interaction is in both directions -- a detailed model of language use can serve as a framework in which to see what changes happen over time, while observations of language change and child development can serve as good heuristics for developing appropriate models.

8.6.4 Intuitive overall grasp

The most common use of a phrase like "I don't understand how that thing works" reflects a sense that there is a global coherence which the person has not grasped. It is often said of large systems (for example, computer time sharing systems and space vehicles) that they are so complex that nobody can understand them. This is more than saying that nobody can remember the entire list of components. It reflects the impossibility of having simultaneously in mind the many basic design issues and interactions which form a gestalt of the system *qua* system.

At the level of explanation dealing with the design of a cognitive system, it is possible that no one person could have a sense of understanding the entire design. Critics have argued that even if one had a computer program which fully duplicated the language abilities of a human, one would not have a theory of language. It would be necessary to develop a theory of the program in order to understand it. This observation may be true, and points to the need for better conceptual tools for dealing with complex systems of all sorts, both natural and artifactual. System theorists (see for example Bateson, 1972) have pointed out that our current scientific methodologies are useful primarily in those areas where small closely related sets of phenomena, and isolated mechanisms can be studied without considering the larger systems in which it takes part. Language use and cognition certainly fall within the range of biological systems for which a more system oriented view may suggest explanations which bring semantics out of the realm of Chomsky's "mysteries".

References

Bateson, Gregory, *Steps Toward an Ecology of Mind,* Ballantine, 1972.

Bobrow, Daniel *G.* and Allen M. Collins, (Eds.) *Representation and Understanding: Studies in Cognitive Science,* New York: Academic Press, 197%

Bobrow, Daniel G., and Terry Winograd, An overview of KRL, a knowledge representation language, *Cognitive Science* 1:1 (January 1977).

Bolinger, Dwight, The atomization of meaning, *Language* 41, 555-573.

Chomsky, Noam, *Aspects of a Theory of Language,* Cambridge: MIT, 1965

Chomsky, Noam, *Reflections on Language,* New York: Pantheon, 1975

Donnellan, Keith S., Reference and definite descriptions, *The Philosophical Review,* LXXV, No. 3 (1966), 281-304. Reprinted in Rosenberg and Travis (eds.), *Readings in the Philosophy of Language,* Prentice-Hall, 1971.

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

Fillmore, C., Against a checklist theory of meaning, Proceedings of the First Annual Meeting of the Berkeley Linguistics Society, Institute of Human Learning, Berkeley, 1975.

Gordon D., & Lakoff, G., Conversational postulates, Papers from Seventh Regional Meeting, Chicago Linguistic Society, Chicago: University of Chicago Linguistics Department, 197 1.

Grossman, R.E., San, L.J, and Vance, T.J. (eds.), *Papers from the Parasession on Functionalism,* Chicago Linguistic Society, 1975.

Kaplan, R. On process models for sentence analysis. In Norman, D. A., et. al. *Explorations in cognition,* San Francisco: Freeman, 1975.

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, William, "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., Hedges: A study in meaning criteria and the logic of fuzzy concepts, Papers from the Eighth Regional Meeting, Chicago Linguistics Society, Chicago: University of Chicago Linguistics Department, 1972.

Minsky, Marvin, *Computation: Finite and infinite machines,* MIT, 1967.

Minsky, M., A framework for representing knowledge, In Winston, P. (Ed.), *The psychology of computer vision,* McGraw-Hill, 1975.

Morgan, Jerry L., Some remarks on the nature of sentences, in Grossman, San, and Vance (eds.), *Papers from the parassession on functionalism,* Chicago Linguistic Society, 1975.

Newell, Allen, A note on process-structure distinctions in developmental psychology, in *Information Processing in Children,* Sylvia Farnham-Diggory (ed.), Academic, 1972.

Newell, Allen., & Simon, Herbert A., *Human Problem Solving,* Prentice Hall, 1972.

Newell, Allen., & Simon, Herbert A., Computer Science as an empirical inquiry: symbols and search, *Communications of the ACM* 19:3 (March 1976), 113-126.

Norman, D. A., & Bobrow, D. G. On data&limited and resource-limited processes. *Cognitive Psychology, 1975, 7, 44-64.*

Norman, D. A., Rumelhart, D. E. and the LNR Research Group, *Explorations in cognition,* San Francisco: Freeman, 1975.

Rosch, Elanor, Cognitive representations of semantic categories, *Journal of Experimental Psychology: General,* 1975, 104, 192-233.

Russell, Bertrand, *Introduction to Mathematical Philosophy,* London: Allen and Unwin, 1919. Chapter XVI on Descriptions reprinted in Rosenberg and Travis (eds.), *Readings in the Philosophy of Language,* Prentice-Hall, 1972.

Sadock, Jerrold M., *Toward a Linguistic Theory of Speech Acts,* Academic, 1975.

Searle, J.R., *Speech Acts,* Cambridge, 1970.

Strawson, P.F., On Referring, *Mind,* LIX No. 235 (1950), 320-44. Reprinted in Rosenberg and Travis (eds.), *Readings in the Philosophy of Language,* Prentice-Hall, 1971.

Weizenbaum Joseph, *Computer power and human reason,* San Francisco: Freeman, 1976.

. Wilson, Deirdre, *Presuppositions and Non-Truth-Conditional Semantics,* Academic 1975.

Winograd, Terry. *Understanding Natural Language. New* York: Academic Press, 1972.

Wittgenstein, Ludwig, *Philosophical Investigations,* Oxford: Blackwell, 1953.