American Journal of Computational Linguistics *Microfiche 78*

LANGUAGE REPRESENTATION

DAVID L. WALTZ, EDITOR

Coordinated Science Laboratory
University of Illinois
Urbana 61801

Papers presented in two sessions of TINLAP-2, the 1978 Meeting of the Association for Computational Linguistics, held with joint sponsorship by the Association for Computing Machinery and its Special Interest Group in Artificial Intelligence.

TABLE OF CONTENTS

TABLE OF CONTENTS

Testing The Psychological Reality of a Representational Model

Dedre Gentner
Bolt Beranek and Newman Inc

ABSTRACT

A research program is described in which a particular representational format for meaning is tested as broadly as possible. In this format, developed by the LNR research group at The University of California at San Diego, verbs are represented as interconnected sets of subpredicates. These subpredicates may be thought of as the almost inevitable inferences that a listener makes when a verb is used in a sentence. They confer a meaning structure on the sentence in which the verb is used. To be psychologically valid, these representations should capture (at least)

1 Similarity of meaning
The more similar two verbs seem in meaning to people, the more their representations should overlap.

2 Confusability
The more confusable two verb meanings are, the more their representations should overlap.

3. Memory for sentences containing the verb
The sentence structures set up by the verb's meaning should in part determine the way in which sentences are remembered.

4. Semantic integration
The representations should allow for the integration of information from different sentences into discourse structure

5 Acquisition patterns
The structural partitions in the representations should correspond to the structures children acquire when they are learning the meanings of the verbs

6. Patterns of extension
The representations should be extendible so as to reflect the ways in which people interpret verb meanings when the verbs are used outside their normal context.

7. Reaction times
The time taken to comprehend a sentence using a given verb should reflect the structural complexity of the verb meaning.

Experiments concerned with predictions 1-5 are described here. The results are promising for a general approach of representation of meaning in terms of interrelated subpredicates, but do not clearly distinguish between several similar representations. For example, to test prediction (2), I read people sentences containing verbs with similar meanings, and asked them to recall the sentences. The degree of overlap in the semantic structures was a good predictor of the number of confusions between sentences. In another sentence-memory experiment (prediction (3)), semantically complex verbs that provided more underlying interconnections between the nouns in a sentence led to better memory for the nouns in the sentence than simple general verbs, or than other complex verbs that did not provide such extra interconnections. To test prediction (5), I tested children's comprehension of a set of possession verbs. Both the order of acquisition among the verbs and the kinds of errors fitted well with an account of the acquisition of verb meaning in terms of interconnected subpredicates.

This research illustrates a breadth-first approach to testing a representation. In the breadth-first approach, many different psychological predictions are made. Each different area of prediction requires a set of process assumptions, and in each case the process assumptions used are those that seem most plausible given previous research in the field. If one representational format can make correct predictions about a number of different kinds of psychological phenomena, then that representation stands a greater chance of being generally useful than one which was tested in only one depth-first way.

This paper describes a program of research that tests a representational format for verb meaning. This research grew out of the LNR (Footnote 1) attempt to the represent the meanings of words in a psychologically satisfying way. Verb meaning seemed a natural place to start for two reasons: (1) verbs are important: it is arguable that they provide the central organizing semantic structures in sentence meanings; and (2) verbs are tractable: their meanings are more easily analyzed than those of, for example, common nouns.

Since different disciplines look at meaning in different ways, it may be worthwhile to describe the stance we took. What we wanted was a system of representation in which we could capture our intuitions about what a word typically conveys; or more specifically about the inferences a person normally makes (or believes should be made) when a word is used. The assumption is that the same representations operate when a person uses the word in speech as when the person comprehends it; however the methodology of experimental psychology makes it natural to spend more time pondering the input process than the output process. This approach differs from thinking of meaning in terms of necessary and sufficient truth-conditions, as many philosophers have done, or from thinking about meaning in generation rather than in comprehension, as many linguists have done. Each of those stances leads to useful intuitions. Overall, there has been a reassuring degree of convergence between the representations proposed.

Representation of Verb Meaning

There are many notational systems for representation of verb meaning (e.g., Abrahamson, 1975, Chafe, 1970; Fillmore, 1971, Gentner, 1975, Lakoff, 1970; McCawley, 1968, Rumelhart & Levin, 1975; Schank, 1972, 1975, Talmy, 1975). These models of verb meaning differ from one another in detail, but there is widespread agreement on the idea that verb meanings can be represented in terms of interrelated sets of subpredicates, such as CAUSE or CHANGE. These subpredicates are not merely concatenated within a word's representation. Rather, they are interrelated, in specific ways. Representations of verb meaning include notation for specifying the relationships among the subpredicates that make up a word's meaning. The notation developed by the LNR Group is a network format. In this system of representation, verb meanings are expressed in terms of subpredicates that stand for states, changes of state, actionals, etc.

The Elements of Verb Meaning. Verbs provide a system in which people can talk about happenings in the world, implicitly distinguishing several types of conceptual possibilities. The simplest of these is the state. A stative predicate conveys a relationship that endures for a period of time between two arguments, normally an object (or person) and an object or value within the conceptual field specified by the stative. For example, consider the sentence shown in Figure 1.

Ida owned a Cadillac from 1970 to 1977.

The verb own conveys that a relationship of possession existed between Ida and the Cadillac for some duration. Besides statives for possession, there are a large number of other statives, including location (to be at, to remain at, etc.) and emotion (to hate, to love, etc.).

In addition to simple stative relationships, verbs can be used to convey changes of state. Following Chafe (1970) I will refer to a change of state as a process. For example, the sentence

Ida receives $10.00.

tells us
(1) that Ida now has $10.00
(2) that someone else had the $10.00 before,
(3) that a change has taken place from this previous state of possession to the present state.

More commonly, verbs express not simple changes of state but causal changes of state. We seem to be very interested in processes that are volitionally caused by humans and other sentient beings. Figure 2 shows the representation of the sentence

Ida gives Sam a rose.

An agent may cause a change of state that relates to another object. Or the same person may act on both agent and experiencer of the change of state. The locational verb move can be used in either way, as in the following examples

a. Ida moved the car.
b. Ida moved to the front seat.

In both these cases the action taken by Ida is unspecified. We often don't care exactly what someone did to cause some process to occur. However, there are also verbs in which the causal action is partially or wholly specified e.g., walk, saunter, meander, stride, run, sprint, race, trot, jog. (See Miller (1972) and Miller & Jonnson-Laird (1976) for a more extensive discussion of the verbs of location.)

Thus, this system allows for the representation of verbs as states, changes of state, causal changes of state. simple actions, and complex cases in which specific actions cause changes of state. Further discussion of the LNR system of verb semantics can be found in the articles by Abrahamson, Gentner, Munro, Rumelhart & Levin, and Rumelhart & Norman in the Norman & Rumelhart (1975) volume.

There are certainly gaps in the system, and aspects of verb meaning that are not expressible in this simple vocabulary. Some unresolved issues are discussed later in the paper. However, the system seems plausible at the first level, and allows a fair range of verb meanings to be captured at least roughly.

At this point in the research it seemed appropriate to begin testing the psychological rightness of the system as so far stated before going on to refine it.

Psychological Tests of the Model

One advantage of psychological experimentation (or of computer implementation) is that it forces one to make explicit the assumptions underlying representation and process. At least some of the choices made can then be tested as hypotheses. Some important assumptions are

(1) a verb's representation captures the set of immediate inferences that people normally make when they hear or read a sentence containing the verb;
(2) in general, one verb leads to many inferences
(3) these networks of meaning components are accessed during comprehension, by an immediate and largely automatic process
(4) the set of components associated with a given word is reasonably stable across tasks and contexts
(5) surface memory for exact words fades quite rapidly, so that after a short time, only the representational network remains.

In testing these representations, I took a very literal interpretation of the notion of representation -- namely that the nodes and arrows in a representation correspond to the concepts and relationships that are stored when a person comprehends a sentence containing a verb. The more ferociously literal the interpretation, the better the chances of discovering counter-evidence.

Semantic overlap. One psychological criterion is that the representations should agree with people's intuitive notions of synonymity and similarity in meaning. One straightforward measure of this overlap is the degree to which people rate verbs as similar in meaning. In a study of about 60 selected verbs, I found that people's average rating of the semantic similarity between two verbs agreed very closely with the degree of semantic overlap between their representations.

A more subtle measure of psychological similarity is the degree to which people unconsciously confuse things in memory. People in a sentence-memory experiment probably try to keep their sentence traces clear. But, suppose that within a short time after hearing a verb in a sentence, a person has only the representational network of concepts and relationships, and not the surface verb Assume further that some pieces of the memory representation may be lost or unaccessible at any time (the "fallibility of human memory" assumption). Then the more two verb representations overlap, the more likely it is that sentences containing the two verbs will be confused in memory, despite people's attempts to keep them straight. In an experiment in sentence memory, using verbs of varying semantic overlap, I found that subjects did indeed confuse the verbs in exactly the way predicted by the theory (Gentner, 1974). The correlation between the number of confusions subjects made between two verbs and the semantic overlap between the verbs, as predicated from the representations, was quite high. In fact, the correlation between representational overlap and number of confusions was slightly higher (though not significantly so) than the correlation between the number of confusions and the rated similarity between the verbs. (The similiarity ratings were taken from the first-mentioned study, with a different set of subjects).

Semantic complexity. Semantic complexity refers to the number of underlying subpredicates and interconnections that make up the basic meaning of a verb. More complex meanings correspond to more specific actions or events. For example, stride is more specific than go. Its meaning contains more subpredicates. We know more having heard sentence (a) than sentence (b).

(a) Ida strode across the field.

(b) Ida went across the field.

Various researchers have looked for evidence that semantic complexity may affect comprehensibility, generally on the assumption that more complex semantic structures are harder to process (Kintsch 1974; Thorndyke, 1977). However, the results have been negative. There is no evidence that more complex words lead either to longer reaction-times or to greater processing loads than do simpler words. I believe that it's incorrect to assume across the board that complexity is psychologically hard. Some research of mine suggests that the effects of semantic complexity in memory are more particular.

Semantic Complexity and Connectivity. Although the view that semantic complexity leads to difficulty has not been supported, there is another side to the complexity-issue. The additional semantic components in a complex verb may set up additional connections among the nouns in the sentence. In this case, more complex verbs should lead to a richer and more highly interwoven sentence representation, and thus to better memory for the nouns in the sentence.

Notice that this prediction derives from a fanatically literal interpretation of the verb representations: more paths in the representation means more conceptual paths in memory. This prediction is quite specific. It is not simply a question of certain complex versus simple verbs having some overall effect, but rather of complex verbs providing extra connections between the particular nouns in question. This is clearly true for Ida and her tenants in the case of sell versus give, as can be seen in Fig 3a and 3b.

I tested for this kind of improvement in connectivity in a series of experiments in sentence memory (Gentner, 1977). I read people sentences that differed in the semantic

connectivity of their verbs, such as the following pair of sentences·

Ida gave her tenants a clock. (simple)

Ida sold her tenants a clock. (complex connective)

Then I gave the people the names of the characters and asked them to recall the sentences. As predicted, they were better able to recall the noun tenants when the complex connective verb sell was used then when the simple verb give was used. More semantic connections between the two nouns led to stronger memory connections.

To see the specificity of the prediction, consider a complex verb that merely amplifies the simple verb and does not add connections between the key nouns. For example, the verb mail (Fig 3c) adds the information that the method of transfer was by mailing or some such long-distance transfer. Using mail leads to more inferences (a more specific event description) than using give. However, the knowledge that the object was mailed leads to few, if any, additional connections between the agent, Ida, and the recipient, tenants. Therefore, the prediction was that use of such non-connecting specific verbs would lead to no improvement over use of general verbs in memory between the nouns.

The results were exactly as predicted The object nouns of complex connective verbs were recalled better than those of general verbs and non-connecting complex verbs. These differences were not traceable to differences in imagery or word-frequency. Thus connectivity is beneficial to sentence memory in a very specific way.

Acquisition. There may be a more direct relationship between complexity and difficulty in children than in adults. Young children often fail to comprehend the full meanings of semantically complex terms (e.g., Bowerman, 1975, Clark, 1973, Gentner, 1975, in press). Working with the verbs of possession, I have observed that children act out the simple verbs give and take correctly before they act out the more complex verbs buy and trade. Still later they learn the yet more complex verbs buy, sell and spend. The order in which the verbs are learned is exactly the order of increasing semantic complexity. This complexity ordering can be made quite precise, since the verbs are closely related in meaning. The representation of a verb at the nth level of simplicity is properly nested within the representation of a verb at the (n+1)th level. Further, when children around 4-6 years are asked to act out sell (as in "Make Ernie sell Bert a boat.") they act out give instead (A boat is transferred from Ernie to Bert). Similarly, buy is acted out as take. They systematically act out complex verbs like simple verbs; and more surprisingly, they choose the appropriate simple verb. My interpretation, consistent with Clark's (1973) semantic features analysis, is that children learn these complex verb meanings gradually, by adding components to their partially correct representations. At any given time, the child comprehends language in terms of the components that he has so far acquired.

Semantic Integration. Another important psychological requirement is combinability. The basic notions of state, change of state, cause, and so on must be combinable into networks larger than the individual sentence. When two verbs share parts of their underlying structure. this redundancy should be utilized to combine the two representations into one discourse structure. How can we test whether this happens? One way is to arrange things so that collapsing the redundancies between two verbs should create the representation of a third verb. Then the prediction is that people should use this third verb in recall.

In a study of semantic integration, I read people short passages and tested their memory by having them fill in blanks (Gentner, 1978). Every passage contained a general verb, such as give. Half the passages also contained additional semantic information, such as the fact that the giver actually owed the money he was giving. According to the representational model, the integration of the representation of give with that of owing should have created the structure of pay. If what people have in their minds after hearing the verbs is the network representations, and if these representations are integrated during discourse comprehension, then people who heard give and owe should end up with the representation of pay. As predicted, subjects hearing the extra material falsely recalled the verb which best fit the composite structure (e.g. pay) rather than the verb actually presented.

Further Issues

I have made the assumption that a verb carries with it a set of inferences that are normally made during comprehension, as well as several supporting assumptions. This view has been fairly well supported by the research presented here, but nevertheless it seems to me an oversimplification. There remain a great many questions, some large and some small.

(1) Where should the line be drawn around a word's meaning? As Clark and Clark (1977) have put it, is word meaning more like a dictionary or an encyclopedia? The extreme of the dictionary approach would be to take a minimal contrast approach, storing with a word only enough to distinguish it from all other words. The extreme of the encyclopedia approach would be to access the entire long-term memory whenever any word is used. The question is, how to define a reasonable middle ground.

(2) What is the process of expansion into a semantic representation during comprehension?

a) Are there invariable inferences? When an incoming word is processed, is there a set of inferences (such as the set I have called the "almost-inevitable inferences" that is always made during comprehension, or is there variation in which inferences get made?

b) If there is variation, is it quantitative or qualitative? Do context and the person's interests and attention determine which inferences get made, so that there are qualitative differences in what inferences get made? Or is the difference merely quantitative, with the radius of expansion varying with the amount of attention (or energy, or interest) that the person brings to bear?

The notion of at least quantitative variation a seems hard to avoid. It is a fairly strong intuition that we process word meanings with varying degrees of energy Further, the phenomenon of instantiation (Anderson, R.C., Stevens, K.C., Shifrin, Z., & Osborn, J., 1977) makes it clear that a model of sentence comprehension must allow for qualitative differences in the final set of inferences stored. For example, compare the sentences

Rover ate his dinner.

Mr. Pritchard ate his dinner.

The verb eat conveys vastly different action sequences when used with different agents, though its causal change-of-state structure remains more-or-less constant. It is possible that this qualitative variation can be accounted for by simple underlying quantitative processes spreading activation. We may have to settle for a more complex model, in which some parts of a verb's meaning are almost always accessed while other inferences develop out of the interaction of the verb with its context, including its pragmatic context. In Hewitt's (1976) terms, there may be both if-added inferences and if-needed inferences. Where in this model (and whether) we want to draw a line between meaning and knowledge-of-the-world is not at all clear to me. (3) Carrying the notion of variable verb meaning still further, how does metaphorical extension work? Most common verbs can be used in several related ways. For example, consider the range of meanings that give can convey depending on the nouns it is used with

	a rose
	a job.
	an heir.
Ida gave Sam	an excuse
	a talking to.
	all his best ideas.
	the time of his life.

Clearly the subpredicate structure varies across these sentences, so much so that some might want to describe this as a collection of entirely different senses of the same word. This misses the structural similarities. Some kind of metaphorical extension of meaning seems a necessary part of a theory of verb meaning, since it is generally the verb that does most of the adjusting. A series of studies by Albert Stevens and me suggests that people faced with an odd sentence assume that some of the subpredicates normally conveyed by the verb are not meant to apply in the sentence at hand A current project is to model the rules for which subpredicates apply in different contexts.

(4) I have so far treated nouns as nodes in the semantic representation. Clearly in order to analyze sentence interactions it is necessary to have a representation of noun meaning. Some progress been made with abstract nouns, such as kinship terms. But the truly nounlike nouns ---basic-level nouns--- resist analysis. I believe that these differences in amendability to analysis reflect differences in the kind of meaning that verbs and nouns have, and that a useful representation of concrete noun meaning may be quite different from that used for verbs, prepositions and even abstract nouns.

(5) There are several aspects of the representational scheme that need further thought. To single out one issue, consider the notion of change of state. The LNR representation represents a verb like get as conveying a change from an initial state of possession to a final state of possession. Schank's Conceptual Dependency theory would represent the entire sequence as a primitive act. Many generative semanticists have represented only the inchoative part of the chain (the change to the final state) as belonging to the assertion of the verb, considering the initial state to be more in the nature of a presupposition (e.g. Fillmore, 1966). All these positions seem to me to have merit. The LNR use of change from initial to final state allows a change-of-state verb to hook automatically with relevant state information. The use of acts as primitives captures the psychological wholeness of change. The use of the inchoative captures the intuition that people seem more interested in the results of an event --i.e. in the final state-- than in the setting state. The explicit change-of-state formats (LNR format and inchoative format) have a natural way of capturing some kinds of metaphorical extension by substituting a different stative while preserving the rest of the verb's structure.

Summary

This work is just beginning. Neither the representations nor the processes that are assumed to operate on them come very close to capturing the subtlety of human language use. Still, the results of the experimental investigation are promising some kind of decompositional model along these lines.

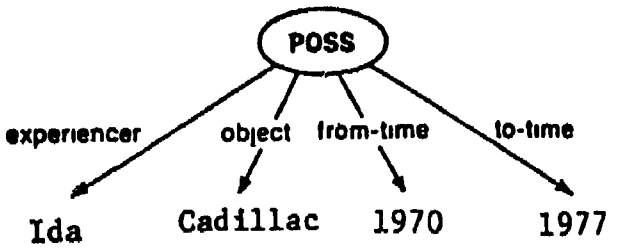


Figure 1. Ida owned a Cadillac from 1970-1977.

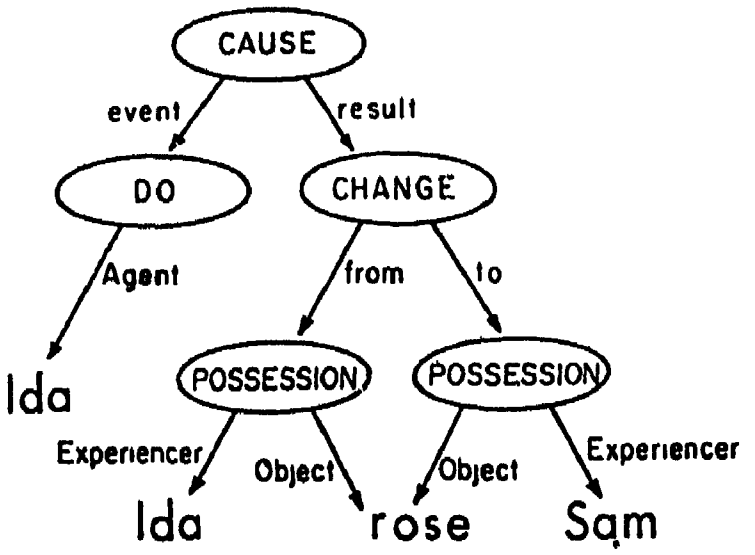
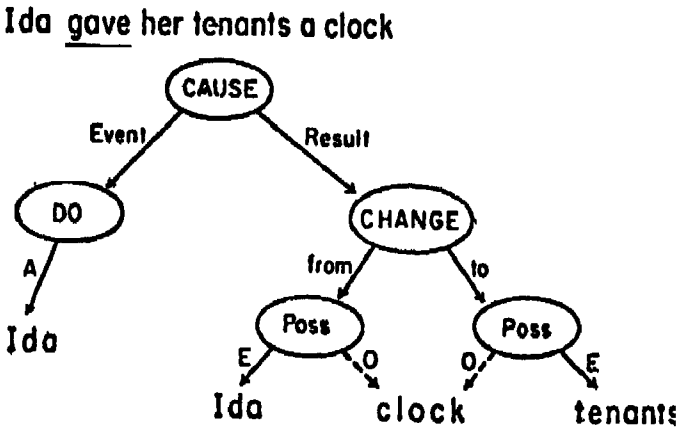
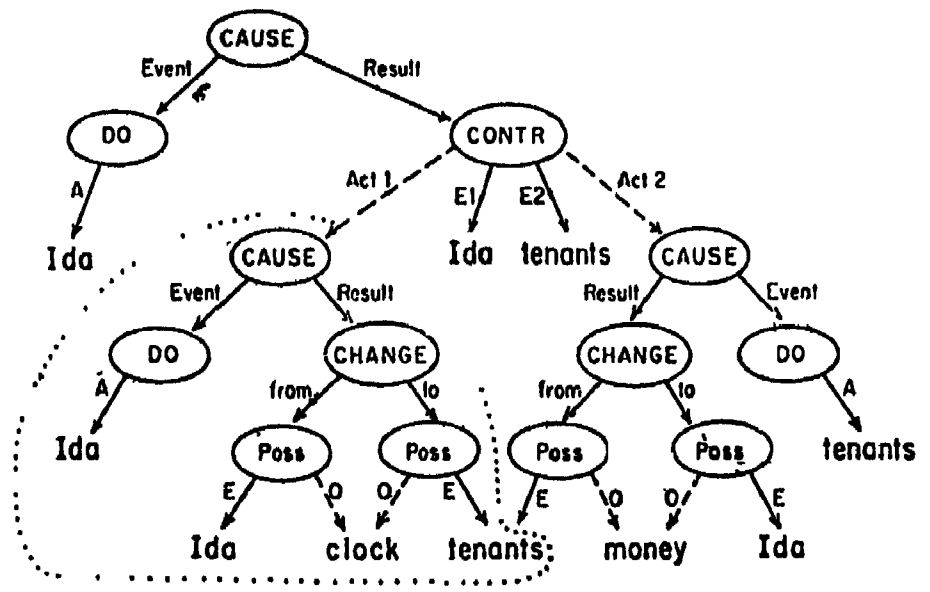


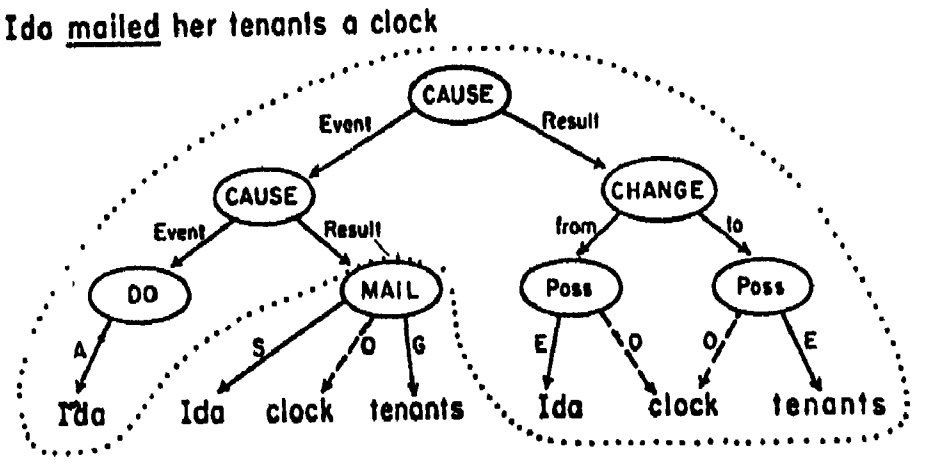
Figure 2. Ida gives Sam a rose.



GENERAL VERB (FEW CONNECTING PATHS)
Figure 3a.



SPECIFIC VERB (MANY CONNECTING PATHS)
Figure 3b.



SPECIFIC VERB (FEW CONNECTING PATHS)
Figure 3c.

Footnote

1. The representational format shown here was developed by a group of researchers at the University of California at San Diego: Adele A. Abrahamson, Dedre Gentner, James A. Levin, Stephen E. Palmer, and David E. Rumelhart. The system is explained in detail in Norman & Rumelhart, 1975.

References

Abrahamson, A.A. Experimental analysis of the semantics of movement. In D.A. Norman, & D.E. Rumelhart, (Eds.), Explorations in Cognition. San Francisco: W.H. Freeman & Co., 1975.

Anderson, R.C., Stevens, K.C., Shifrin, Z., & Osborn, J. Instantation of Word Meanings in Children, May 1977.

Bendix, E.H. Componential analysis of general vocabulary: The semantic structure of a set of verbs in English, Hindi, and Japanese. The Hague; Mouton, 1966.

Bowerman, M. The acquisition of word meaning: An investigation of some current conflicts. Paper presented at the Third International Child Language Symposium, London, September 1975.

Chafe, W.L. Meaning and the structure of language. Chicago: University of Chicago Press, 1970.

Clark, E.V. What's in a word: On the child's acquisition of semantics in his first language. In T.E. Moore (Ed.), Cognitive development and the acquisition of language, New York: Academic Press, 1973.

Clark, H.H. & Clark, E.V. Psychology and Language. New York: Harcourt Brace Jovanovich, Inc., 1977.

Fillmore, C.J. Review of Bendix's Componential analysis of general vocabulary: The semantic structure of a set of verbs in English, Hindi, and Japanese. International Journal of American Linguistics, 1966, 32, Part II, No. 2. Publication 41.

Gentner, D. Towards a psychological theory of the meaning of the possession verbs. Unpublished doctoral dissertation, University of California, San Diego, 1974.

Gentner, D. Evidence for the psychological reality of semantic components: The verbs of possession. In D.A. Norman and D.E. Rumelhart, Explorations in Cognition, San Francisco: W.H. Freeman & Co., 1975.

Gentner, D. On relational meaning: The acquisition of verb meaning. Child Development, in press.

Gentner, D. Semantic integration of word meanings Bolt Beranek and Newman Inc. Report No. 3826, May 1978. Also to appear as a Center for the Study of Reading Technical Report.

Hewitt, C. Viewing control structures as patterns of passing messages. M.I.T. AI Working Paper 92, 1976.

Lakoff, G. Adverbs and modal operators. Indiana University Linguistics Club Reprint. Bloomington: Indiana University Linguistics Club, 1970(a).

Lakoff, G. Irregularity and syntax. New York: Holt, Rinehart and Winston, 1970(c).

McCawley, J.D. The role of semantics in a grammar. In E. Bach and R.T. Harms (Eds.), Universals in linguistic theory. New York: Holt, Rinehart and Winston. 1968(b).

Norman, D. A., Rumelhart, D.E. & the LNR Research Group. Explorations in cognition. San Francisco: W.H. Freeman & Company, 1975.

Rumelhart, D.R. & Levin, J.A. A language comprehension system. In D.A. Norman & D.E. Rumelhart, Explorations in cognition, San Francisco: W.H. Freeman & Co., 1975.

Schank, R. Conceptual Dependency: A Theory of Natural Language Understanding, Cognitive Psychology, 1972, 3, 552-631.

Schank, R.C. The structure of episodes in memory. In D. Bobrow & A. Collins (Eds.), Representation and understanding. New York: Academic Press, 1975.

Talmy, L. Semantic structures in English and Atsugewi. Unpublished doctoral dissertation, University of California, Berkeley, 1972.

Thorndyke, P.W. Cognitive structures in comprehension and memory of narrative discourse. Cognitive Psychology, 1977, 9, 77-110.

What Makes Something "Ad Hoc"

by Roger C. Schank
Yale University
Department of Computer Science
New Haven, Conn. 06520

Only one of the questions posed before this session really inspires me to take pen in hand. "How general are various formalisms? Are they really ad hoc solutions to relatively narrow domains?"

That is not exactly my favorite question. I find the thought of having to address it palatable only if I can delude myself into believing that this is the last time I shall have to deal with it So, proceeding on the basis of that delusional belief, I shall begin.

Ad Hocness, I have come to believe, is a disease that all new theories in the three fields in which I consider myself well-versed, namely linguistics, psychology and Artificial Intelligence, contract at conception, sort of like original sin This would not be so bad if it were a disease for which there were a cure, but alas there is none.

We are all familiar with the phrase "beauty is in the eye of the beholder." In this case we have an instance of "the disease is in the eye of the beholder" which of course explains why the cure is so elusive. The beholder rarely wants to do anything about it. To discuss this more subjectively, let's take a neutral case. Before doing so, we shall have to point out what a case can be expected to look like. A case of "ad hocness" usually fits the form (or should I say the "ad hoc" form)

Theory X is called "ad hoc" by group with rival theory Y

The research described in this paper was supported by the Advanced Research Projects Agency of the Department of Defense and monitored by the Office of Naval Research under contract N00014-75-C-1111.

To get to our neutral case, we shall start our discussion where X is Conceptual Dependency and Y is Transformational Generative Grammar. Before I begin, I should note that there are conditions on X and Y relative to each other, namely that X must be a theory that has been conceived at a date later than Y was conceived. Furthermore Y should have been dominating some academic field which X is seeking to invade.

What makes a theory X assailable by Y as ad hoc? There are a number of criteria:

1 - X must explain a phenomenon that Y chose to ignore and that Y would rather go on ignoring since Y could not possibly explain it.

2 - X must be fundamentally at variance with Y, so that if X were right Y would be necessarily wrong.

3 - X must use different criteria of judgment of how a phenomena should be explained than Y does.

The following rules are used for the strategy to be followed in labelling an X as ad hoc:

1 - Since X will undoubtedly show how its theory explains a given particular phenomenon, accuse X s theory of only working in that case. This will put the burden of proof for generality on X rather than Y and also has the desirable effect of putting X in the position of not being able to prove anything with out proving everything.

2 - Choose a phenomenon to explain in which it is virtually impossible to explain everything, thus giving game and set to Y.

Consider our hypothetical case where Conceptual Dependency is X and Tranformational Grammar is Y. An examination of the literature will show that criteria 1 through 3 as well as the two available strategies have been used by the Transformationalists. In various articles and public performances charges of "ad hocness" have

been raised against Conceptual Dependency. We are told that our structures only work for the examples we discuss, that we have "no principled way of going from a sentence to a conceptualization" (Dresher and Hornstein (1976)) or that "Schank provides no demonstration that his scheme is more than a collection of heuristics that happen to work on a specific class of examples" (Weizenbaum (1976)). (If the reader is wondering how Weizenbaum got to be a transformationalist in my view, he need only read Weizenbaum s further remarks extolling Chomsky as having met the criteria that he claims I have not met.)

To what extent are these charges valid? To not knowing if one can extract a conceptualization from any sentence (and its corroborating charge of not proving that there exists a right CD diagram for any sentence) I plead guilty. But of course, I would be less than completely honest if I did not also note that there does not exist any theory or theorist who would not also have to plead guilty. Have the transformationalists shown us that they have some principled way of extracting conceptualizations from sentences or determining the correct representation for any sentence? Unless they are keeping their solution as a secret plan not to be revealed until after the election, I would have to imagine that the answer to this is that they do not have a solution to the problem. So clearly, they are no more or less ad hoc than we are. (Of course I might note here that we do have programs that suggest that we can do a large class of examples and show that our parsers are at least the beginning of some set of principles that work, but I won t).

What about Weizenbaum s attack? Perhaps it is all heuristics. To this charge I plead no contest. It might be that, in the end, we will have built a working program that solves the entire natural language problem and it will be easily labelled as a grand set of heuristics. Won't that be terrible! To quote Dresher and Hornstein again, "Not only has work in AI not yet made any contribution to a scientific theory of language, there is no reason to believe that (AI)...will ever lead to such theories".

And what will they say after success has been achieved and the ultimate natural language system has been designed? The same thing of course. Chomsky himself (personal communication) has claimed that such an achievement would be no more interesting than the achievement of the 16th century clockmakers.

I mention all this in the hope of pointing out that it is not just me and my theories that are damned by criticisms of ad hocness. We are all damned by them. Our ultimate success would not be even recognized much less applauded by those who criticize our solutions as ad hoc. Suppose every domain we worked on required yet another ad hoc solution. This might well be the case after all. What would we lose if this happened? Nothing at all. That s what artificial intelligence is all about. AI is the designing and testing of theories about human understanding capabilities. There is, at the moment, no reason to believe that people solve puzzles the way they read newspapers or that they play chess the way they answer questions. Of course, we all hope that there exist some general mechanisms that solve all these problems in some neat way. We hope this in large part because we are lazy. We would not like to have to work on each problem individually. We also hope this because we believe our intuitions when they tell us how reading a newspaper is a lot like watching a soap opera. A word of caution is necessary here. Beware of your intuitions. As a child you learned how to do each of these things separately and were pained to deal with each one of them. Of course, we do expect there to be some general principles that apply across domains.. But if these principles are affix - hopping or trace - deletion we are all in trouble.

Part II

Having said all this, now let me tell you what I actually believe. I do not believe that any of our theories are ad hoc. Just because CD needed to be modified by causal chaining rules, and those by scripts and those by plans and goals and themes, and those by triangles, does not mean that what we are doing is ad hoc. We are no more ad hoc in hypothesizing our primitive elements than chemists were in hypothesizing theirs. I do not know what the ultimate result will be. How many elements make up the correct number, or what other kinds of formalisms will need to be added to those listed above is still unknown.

I do know how AI does its research however. We build a program to do a small class of examples and when we are finished we rip it apart and build a bigger and better program to do larger examples. In so doing, ad hoc entities (oftimes called kludges) cannot survive. If a formalism does not keep handling more data it is either abandoned or moved down to a special purpose role within a larger program.

Well, in ten years of research by my research group what has survived? After ten years and probably a hundred different kinds of programs, Conceptual Dependency is still with us. It still works for us. I challenge any other theory that has been programmed to say the same! Is it ad hoc? I leave that as an exercise for the reader.

PART III

Just to give the reader a feel for the nature of ad hoc thinking in AI that I believe to be worth espousing, I will now consider a problem that I have recently been working on. We have had a problem in representing certain kinds of political concepts in our old representation. Since we have been very concerned with the problem of newspaper story understanding it is very important that we be able to handle such concepts in a clean representation that will facilitate computer understanding.

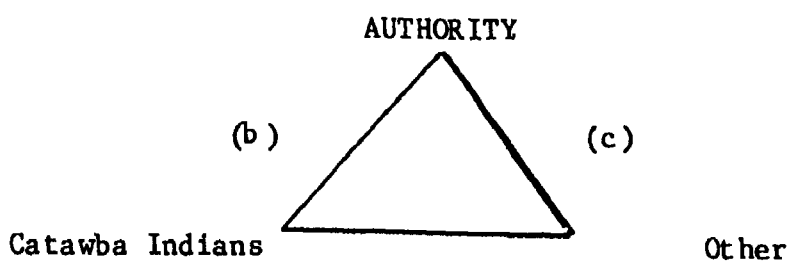
The problem we are attempting to solve can be illustrated by looking at a recent New York Times headline "Catawba Indians land claim supported." The problem here is to be able to represent what "land claim" and "supported" mean. We know that a land claim is more than what we might use to

represent it in Conceptual Dependency.

Something like "Indians MTRANS land be possessed by Indians" is possibly true, but it misses the point. A "land claim" is in a sense a petition to a higher authority to resolve a dispute between two parties. That is, the Indians are saying to the U.S. Government, "this land is ours". It may not be possible to infer the particulars of this land claim. Indians have been known to take the land by force, to file documents in government offices, to complain to newsmen and so on. The important point here is that we really need not know, and in most cases a reader would not bother to worry about, exactly which method has been selected. Rather, a reader feels that he understands such a sentence when he has been able to identify the relationships and aims of the parties involved.

A program must recognize that a "land claim" is a type of petition to a higher authority to resolve a dispute about land ownership. We do not know who presently owns the land, but we know enough about ownership of property to infer that there is probably a counter petition of some sort. We also know about petitions to authority. They usually get resolved by the authority. In this case then, "supported" refers to the decision of the authority in the case.

This information can be represented graphically by a kind of triangle (example 1)

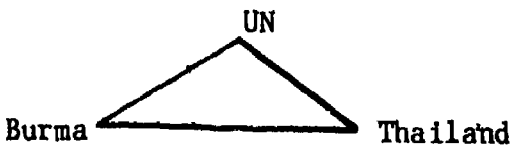


(a)

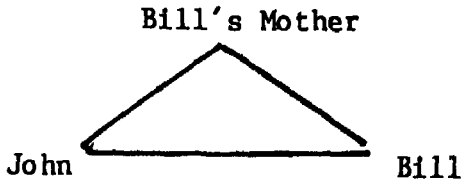
In this triangle (a) represents the dispute between the Indians and the owners of the land, (b) represents the appeal to authority to resolve the dispute made by the Indians, and (c) represents the authority's decision.

Triangles of this sort have use in representing any type of dispute. For example, in (2) and (3) such triangles can also be constructed

(2) Burma appeals to UN to settle border dispute with Thailand.



(3) John complained to Bill's mother that Bill hit him.



Of course, these triangles just suggest the basic relationships involved. In order to add substance to the bare bones of the triangles we shall have to deal with some representational issues that are being glossed over here. The important point at this juncture is that there is an essential similarity across (1), (2) and (3), that the similarity must be represented in some way, and that that similarity can be exploited for use in an understanding system.

The first representational problem we encounter in trying to make explicit much of what is implicit in the triangle representation is that we will need to design a new set of ACTs to take care of the various relationships.

In the primitive ACTs of Conceptual Dependency we have a system that represents physical actions by using a small set of basic actions that can combine in various ways to describe detailed or complex actions that underlie seemingly simple verbs and nouns. The primitive ACTs do not account for intentionality and goals underlying physical action. To account for such things we devised a complex apparatus discussed in Schank and Abelson (1977). If we wish to account for social events, we will need a system of basic social ACTs to represent the social actions that comprise the events. I term these "basic social ACTs" rather than primitive ACTs because in the end most social ACTs have some physical manifestation. Often their physical manifestation is uninteresting however. For example a government decision may be MTRANS-ed in a variety of ways. The manner of the MTRANS (written, announced in a speech, etc) is often not significant with respect to the overall social effect of the action. Furthermore the MTRANS itself is only slightly interesting. The standard inferences from MTRANS apply, but there are some highly significant inferences that need to be made that are not obviously available.

For example, the most significant inference to be made from an authority's decision is that simply by virtue of that decision something has actually happened. That is, a government authorization is a truly performative ACT. Thus, if the government says some property is mine, or that a man is a criminal, then it is so by virtue of their saying it. Similarly other authority figures have the same power. A professor can say a thesis is finished and a student has a Ph.D. and these things are the case by virtue of his saying it.

Not all authority's decisions are like this to be sure. Sometimes an authority gives an order and that order must be carried out for the decision to have effect. Frequently these orders come about as a result of a governmental decision or authorization. If the government says the land belongs to the Catawba Indians, then it does, but they may have to send in the National Guard to get the original owner off the property.

What I am proposing then is two basic social ACTs - AUTHORIZE (abbreviated AUTH) and ORDER. AUTH is something only an authority can do. (This is a bit circular actually since if you actually can AUTH then that defines you as an authority.)

In a sense then, an authority is one who when he acts like he is doing an AUTH (that is he does the physical ACTs that ordinarily correspond to an AUTH) in fact causes some things to happen as a result of the AUTH that were supposed to be the results of the AUTH. In other words, you cannot really tell if an AUTH has taken place until it becomes clear that the person doing the AUTH can back up his AUTH in some way.) The object of the AUTH is the authorization or new state of the world. AUTH takes a recipient, namely the relevant parties in the dispute.

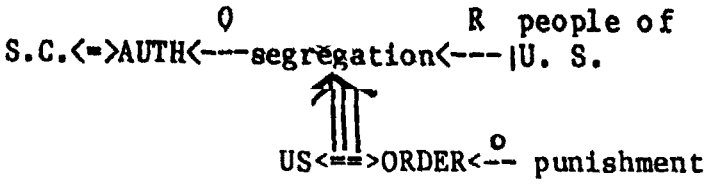
ORDER is a frequent inference of AUTH. The government can AUTH the army to fight a war, but that doesn't, simply by virtue of the statement, imply that they are fighting it. A subsequent ORDER is required that carries with it the implicit punishments that are relevant in carrying out an order.

Why can t we do these things with CD primitives we now have? What is the advantage of these new ACTs? To answer these questions, we need to look at the purpose of a primitive ACT. It is possible to represent ORDER in CD for example. The verb 'order' means to MTRANS to someone that they must do a particular action or face some (usually implicit) consequence. Thus, implicit in the verb 'order' but explicit in the CD representation for 'order', is the idea that if the required ACT is not performed then someone will possibly do something to harm the recipient of the order in some way. This implied punishment is a part of the concept 'order' but is it necessary that we think of it each time that we understand an order to have taken place?

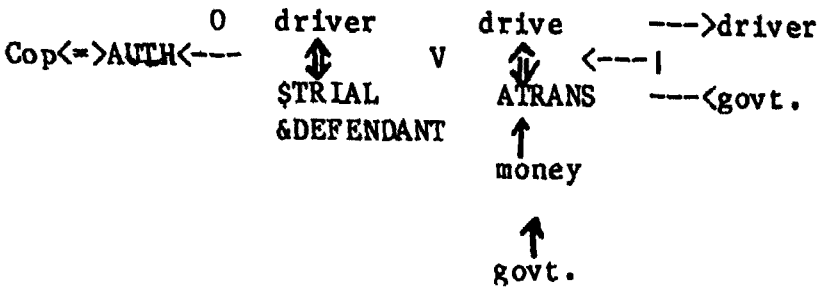
The same question can be asked with respect to authorize . We understand what authorization or governmental decision is, but we need not access all that information each time we understand the word. Consider the problem of explaining the meaning of these words to a child for example. It is very difficult to explain them precisely because they are so complicated at the level of physical primitive ACTs. Yet these ideas are really not complicated at all at a social level of ACTs. Such simple concepts such as ORDER and AUTHORIZE form the basis of the organization of societies. What is complex at one level is simple at another. This idea of nested levels of complexity, each with their own set of primitives, is a very important one for the representation of information in artificial intelligence. By choosing a good set of primitives we can effectively organize what we need to know. Thus, ORDER and AUTHORIZE have inferences that come from them just as the physical primitive ACTs do. The main difference is that these basic social ACTs are not primitive in the same sense. They can be broken down but we would rarely choose to do so.

The use of these new basic ACTs is much like the use of the original primitive ACTs. We can predict what will fill slots reasonably in a conceptualization and make inferences about slot fillers and consequent inferences as we would any conceptualization. Thus we represent sentences such as the following using AUTH

(4) The Supreme Court decided segregation is illegal.



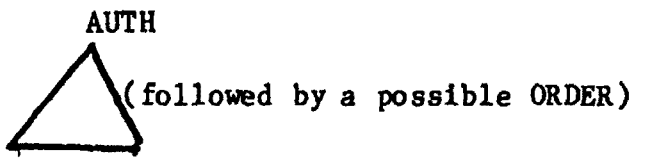
(5) The cop gave the speeder a ticket.



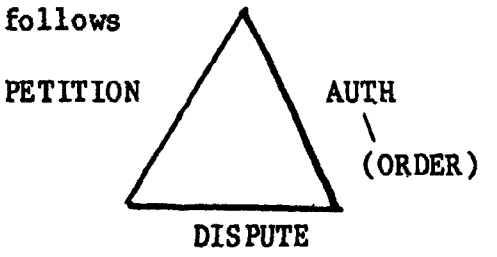
In (4) we have chosen to ignore representing segregation for the moment, since it is obviously complex. Supreme Court decisions are AUTHs. They also carry with them (as do most AUTHs) an implicit ORDER for 'punishment' if certain circumstances are not met The straightforward inference from (4) then is that someone practicing segregation can expect to be punished.

Policemen are authorities also. In (5) the ticket is a written manifestation of an AUTH that either puts the driver in a DEFENDANT role in a $TRIAL script or forces him to pay a fine. The instrument of the AUTH is the actual PTRANS of the ticket (left out here). The important point here is that we could represent (5) using PTRANS only. However, what we would be describing is the physical ACT itself when it is the social ACT that is significant here. (When I was young there was much talk of bad kids getting "JD cards". I never understood what was so horrible about that. Couldn't they just throw them away?) The social significance of an ACT must be represented if it is understood.

Now that we have presented these two ACTs let's return to our triangle



We have named one side of the triangle. The other sides represent ACTs as well. The complete triangle is as follows

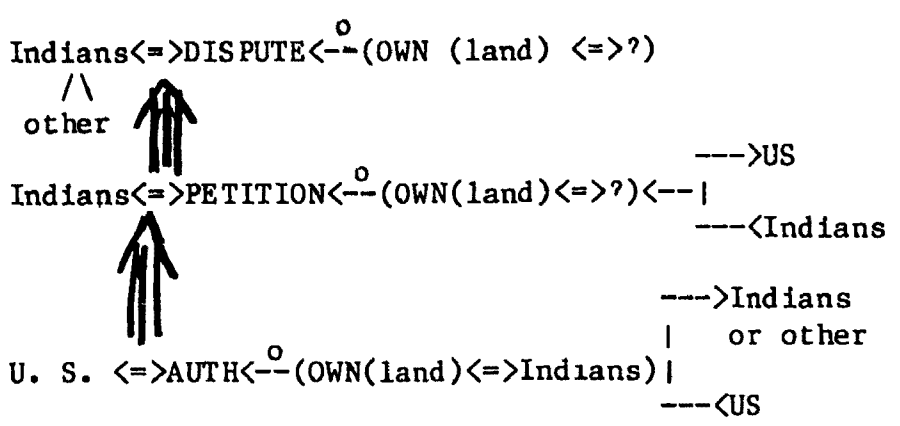


The ACT PETITION represents an individual or group's act of requesting AUTH's from an authority. Thus a "civil suit" is a PETITION to the courts using some legal scripts. A protest demonstration is a PETITION to unstated

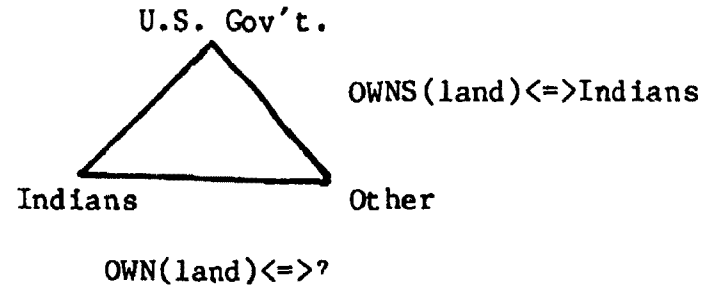
authorities using some demonstration script. The point here is that we cannot do away with the scripts that describe the actual physical manifestations of these events. However, the scripts are instruments of the social ACT involved - PETITION. The most important inference from PETITION is, of course, that an AUTH is expected that will resolve the issue that is the object of the PETITION.

The issue that is the object of the PETITION is the DISPUTE itself. DISPUTE takes two actors (one of whom may be quite passive). The object of the DISPUTE is the issue involved. DISPUTE takes no recipient as it is not an inherently directed ACT. It is the ACT of PETITION that directs it to a particular authority who can AUTH something that will resolve it.

We are now ready to deal with sentence (1) (Catawba Indians Land Claim Supported). The representation using the new social ACTs is



Since this representation is not as easy to write as the triangular one, we shall continue to use triangles in the remainder of the paper. Thus (1) is



We will leave out the arrows and the ACTS for diagrammatic purposes, but the above triangle should be understood as containing all the information given in the CD diagram for (1). (Actually the triangles contain more information.)

Triangles provide us with a method for representing the social significance of actions. As with any other representation scheme, the advantage of the symbols we create can only be in the new symbols or actions that they spawn. That is, it is the inferences that come from the triangles that are of key importance. When we created the original primitive ACTs we said that PROPEL was no more than the set of inferences that it fired off. The same is true here, so we must ask what these inferences are.

The first thing we can recognize about potential inferences here is that they will come in two varieties. The first are the inferences that are fired off from the new social ACTs that we have created. The second kind are those that come from the triangles themselves. That is, there should be patterns of triangles that are recognizable for the triangles they spawn as well as a set of inferences that come from the fact that certain triangles exist.

As examples of this let us consider again sentence (2)

(2) Burma appeals to UN to settle border dispute with Thailand.

Since the representation of (2) involves a PETITION we can employ the inference rules that are fired by PETITION. Some of these are

a. For every PETITION we can expect a corresponding AUTH.

b. For every PETITION there was probably a DISPUTE that gave rise to it.

These rules lead us to the inferences available from AUTH and DISPUTE. Of course, inferences from inferences have a lower probability of truth, so for (2) the inferences below would be somewhat less certain.

c. An AUTH can cause a DISPUTE to end.

d. An AUTH can cause a PETITION to a higher authority from the party unfavorably affected by the AUTH.

e. An unfavorable AUTH can cause a rebellion, or lack of acceptance of the validity of the AUTH. This can give rise to ORDERs to effect the AUTH in the case of individuals versus governments or wars in the case of governmental conflicts

f. An AUTH causes a new state of the world to exist, often ending an old state in conflict with the new state.

g. A DISPUTE can cause one party to PETITION.

h. A DISPUTE can cause a PROPEL to cause damage to occur for individuals, or a WAR triangle to be initiated for countries.

There are, of course, a great many more of these kinds of inferences than we are listing here. The above list is mostly intended to give the flavor of basic social ACT inferences. It is important to note that the social ACTs give rise to inferences at both of the other levels of representation besides those at the same level of representation. That is, given a social ACT we may be able to infer another social ACT, a new primitive ACT, or a new triangular representation.

Thus, for (2) we have two representations to start with one is at the standard CD level and uses MTRANS, the other is at the social level and uses PETITION. Both of these representations would be available as output from the parser.

The MTRANS representation would fire off inferences about the methods of communication possibly used - that the UN now knows about the problem and so on.

The PETITION representation would fire off inferences about the expected AUTH from the UN. Since we know how the UN does its AUTHs, this would fire off a UN script of some kind that dealt with voting and debate. PETITION would also cause DISPUTE to be inferred which would cause inferences about the kind of methods possibly employed by the quarreling countries, both in creating the DISPUTE and escalating it.

The existence of the PETITION-AUTH-DISPUTE triangle would fire off an inference that the country kind of triangle existed. Thus, a new triangle that was lopsided showing possible aggression from Thailand towards Burma would be created. This triangle would in turn fire off inferences about attempts to RESOLVE the DISPUTE (one of which was (3) itself) and would predict an escalation towards the WAR triangle with its normal inferences if a RESOLVE did not take place.

Although the above is rather sketchy, the point should be clear. We need additional representational mechanisms to handle the many levels at which statements can be interpreted. Triangles provide us with a new set of inference rules providing more power to the understanding system. Are they ad hoc? Of course they are. My point is simply that such ad hoc mechanisms will either solve the problem or help us create a more general solution that will solve the problem. The program that we are writing that uses triangles is also ad hoc. Is is a kludge? No. If it were it wouldn't be worth a thing. But, here again, if the program we write can handle many examples as we rewrite it because of what we have learned from it, then it will hav been worthwhile.

The program below reads newspaper headlines in English and generates, by use of triangles and the inferences available from triangles, a paraphrase of the input. This English paraphrase is generated by the program.

```
*****************************************

TRIANGLE analyzer loaded.

INPUT SENTENCE
(CATAWBA INDIAN LAND CLAIMS SUPPORTED)

 (PARSE I1)   CON4

 Expanding token  CON4 =
((CON ((ACTOR (*PP* CLASS (#GROUP)
 CFEATURE (*AMERINDIAN*) TYPE
(*ETHNIC*) NAME (CATAWBA) TOK NP1)
<=> (*PETITION*) OBJECT ((ACTOR
(*PP* CLASS (#REGION) TOK NP2 REL CON1)
IS (*OWN* VAL NP1)) TOK CON1)
FROM NP1 TO (*PP* CLASS (#INSTITUTION)
MEM *COURT* TOK NP3)) TOK CON2)
IR ((ACTOR NP3 <=> (*AUTH*) OBJECT CON1
RECIP1 NP1 RECIP2 GAP1 FROM
GAP2) TOK CON3)) TOK CON4)
```

The Catawba Indians asked a Federal Court to rule that they own the land.

The Catawba Indians requested a Federal Court to rule that the land is owned by them.

The Catawba Indians appealed to a Federal Court.

The Catawba Indians asked a Federal Court to rule that they own the land and it decreeed that the land is owned by them.

[Generating inferences from CON4]

>(TELL-STORY)

The Catawba Indians and the other parties disagreed over the ownership of the land.

The Catawba Indians requested a Federal Court to rule that they own the land.

A Federal Court decided that the land is owned by the Catawba Indians.

The other parties will probably appeal the decision.

The other parties might use force against the Catawba Indians to assert that they own the land.

*****

This program was written by Jaime Carbonell and Stephen Slade.

References

Dresher, B.E. and Hornstein, N. [1976] On some supposed contributions of artificial intelligence to the scientific study of language, Cognition, 4(1976) 321-398.

Schank, R.C. and Abelson, R.P. [1977], Scripts, Plans, Goals and Understanding An Inquiry into Human nowledge Structures, Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Weizenbaum, J., [1976], Computer Power and Human Reasoning, W.H. Freeman and Company, San Francisco.

The Relation of Grammar to Cognition--a Synopsis

Leonard Talmy
Program in Cognitive Science / Center for Human Information Processing

Abstract

A sentence (or other portion of discourse) is taken to evoke in the listener a meaning complex, here called a "cognitive representation". The lexical elements of the sentence, to simplify, by and large specify the content of the cognitive representation, while the grammatical elements specify its structure. Thus, looking systematically at the actual notions specified by grammatical elements can give us a handle for ascertaining the very makeup of (linguistic-) cognitive structuring. We accordingly examine a number of grammatically specified notions, observe the categories and systems in which they pattern, and speculate on broader cognitive connections.

Some provisional findings have already emerged. Grammatical specifications for structure are preponderantly relativistic or topological, and exclude the fixed or metrically Euclidean. The categories in which grammatical notions pattern include:

plexity	perspectival mode
state of boundedness	level of synthesis
state of dividedness	level of exemplarity
degree of extensionality	axial characteristics
pattern of distribution	scene-breakup "

Grammatical specification of structuring appears to be the same, in certain abstract characteristics, as the structuring of visual perception.

0. Introduction

A sentence (or other portion of discourse) is taken to evoke in the listener a particular kind of experiential complex--here to be termed a "cognitive representation" or "CR".[1] There appears to be a significant way in which different portions of the language input specify, or code for, different portions of the CR. The major finding is that--for a first approximation--the lexical fraction of a sentence codes mainly for the content, or substance, of a CR, while the grammatical fraction of a sentence codes mainly for the structure of a CR. Determining the structure within a realm of phenomena has been a central concern for analytic science, including linguistics and psychology. With grammar seen in the above light, it can be used in determining the structure, of the language-related portion of human cognition, with possible connections to further portions. In particular, looking systematically at the actual notions specified by grammatical elements can give us a handle for ascertaining the very nakeup of (linguistic-) cognitive structuring.[2] The beginnings of such an endeavor are the aims of this paper

Several ideas here require some immediate elaboration. The distinction between lexical and grammatical is made entirely formally--i.e., without any reference to meaning--on the basis of the distinction between open-class and closed-class.[3] All open-class elements--i.e., the stems of nouns, verbs, and adjectives[4]--are considered lexical. Everything else is considered grammatical. Included here are all closed-class morphemes and words--inflections, particles, adpositons, conjunctions, demonstratives, etc.--as well as syntactic constructions, grammatical relations, categorial identities, word order, and intonation. Terminologically here, "grammatical element" will be used to refer to any of these.

The nature of content and of structure, and the distinction between them, are not understood well enough to be addressed analytically in this paper and must be left to our intuitive sense of the matter.[5] Taking them for granted, however, we can now more finely characterize the linguistic-cognitive cross-relationships noted earlier. While most of a CR's content is specified by the lexical fraction of a sentence, the lexical items do usually specify some structural notions along with the contentful ones. The grammatical elements of a sentence more unalloyedly specify only structural notions and specify them more determinately in the case of conflict with a lexical item, establishing perhaps the majority of a CR's structure.[6]

In other work in the present direction--notably Fillmore's (e.g., 1975, 1976)--concern has also been with ascertaining structre, but the sentence elements used as starting-points have generally been lexical items with prominently inmixed structural specifications (like buy and sell). The present work, in part a complement to the other, takes advantage of grammar's greater directness and completeness in specifying structure.

This paper is divided into three sections. In the first, a sampling of grammatical elements is examined for the notions that they specify, both as an introduction to out method and for the aim of noticing properties common to such notions as well as properties excluded from them. In the second, we present a number of the categories in which grammatically specified notions have been observed to pattern. In the third, we speculate on broader cognitve connections.

1. The Nature of Grammatically Specified Notions

In this section we examine a small sampling of grammatical elements for the particular component notions that they specify. The sample will give a heuristic indication of the kinds of notions that get grammatically specified as well as of kinds of notions that possibly never do. The excluded kinds will be seen as readily specifiable by lexical elements. A further comparison between the characteristics of grammatically specified notions and of lexically specified ones is then made. To indicate the major finding at the outset, it seems that grammatical specifications for structure are preponderantly relativistic or topological, and exclude the fixed or metrically Euclidean.

For a first simple case, many languages have inflections for the noun (English has -Ø and -s) that specify the uniplex or the multiplex instantiation of the object specified by the noun. By contrast, no languages appear to have inflections that specify the redness or blueness, etc.--i.e., the particular color--of the object specified by a noun. In the preceding, the underlined are instances of "notions". The first set are grammatically specified and can be readily seen to play a structuring role in a CR.[7] The second set are perhaps never found specified by grammatical elements, though they are everywhere found specified by lexical elements (such as (red and blue).

For another case we consider a deictic like the English this or that as in This chair is broken. A grammatical element of this type specifies the location of an indicated object as being, in effect, on the speaker-side or the non-speaker-side of a conceptual partition drawn through space (or time or other qualitative dimension). This integral specification can be analyzed as containing the following component notions (enclosed by quotes):

(1)
- a-b. a 'partition' that divides a space into 'regions'/'sides'
- c-e. the 'locatedness's' (a particular relation) of a 'point' (or object idealizable as a point) 'within' a region
- f-g. (a side that is the) 'same' as or 'different' from
- h-i. a 'currently indicated' object and a 'currently communicating' entity

Notions that might at first be ascribed to such deictics, such as of distance or perhaps size, prove not to be, on the evidence of sentence-pairs like (2):

(2) a. This speck is smaller than that speck.
    b. This planet is smaller than that planet.

The CRs evoked by (2a) and (b) differ greatly, involving tiny objects millimeters apart or huge objects parsecs apart. Yet the sentences differ only lexically, not grammatically. Hence, the CRs' notions as to the magnitude of size or distance cannot be traced to the deictics (or to other grammatical elements) in the sentences. Thus, the notional specifications of a this or a that appear, in part, to be genuinely topological: the establishment of a partition remains a constant, but its position can vary unlimitedly (or, using topology's characterizability as "rubber-sheet geometry", the partition's distance away can be stretched indefinitely) without any constraints imposed by the deictics' specifications per se. This finding about the deictics alerts us to noticing whether any grammatical elements make specifications about magnitude. A spot check through English and various other languages suggests that--while there are apparently grammatical specifications for relative magnitude[8]--there are possibly never any for absolute or quantified magnitude, whether of size, distance, or other parameters.

For a third case, we consider the type of adposition that specifies, for a moving object, certain characteristics of path and of point- or frame-of-reference! An example of this type is English through as used, e.g., in:

(3) a. I walked through the water.
    b. I walked through the timeber (i.e., woods).

In this usage, through specifies, broadly, 'motion along a line that is within a medium'. The component notions contained here include:

4)
- a-e. 'motion'--i.e., 'one-to-one correspondences' between 'adjacent' points of 'space' and adjacent points of 'time'
- f. motion that describes a 'line'
- g. the locatedness of a line within a 'medium'
- h-i. a medium, i.e., a region of three-dimensional space set apart by the locatedness within it of 'material' that is in a pattern of distribution' of a certain range of character (still to be determined)

Again, with (3a) and (b) differing only lexically, any notional differences in their CRs cannot be attributed to through. Thus, not within the specificational purvue of that element are: the 'kind of substance' comprising the medium and the 'sensorimotor characteristics' attendant on executing the motion--as, here, those attendant on wading vs. weaving amidst obstacles. With other sentence pairs like

(5) a/b. I crawled/ran through the timber.
(6) a/b. I zig-zagged/arced throught the timber.

it can be further determined that 'rate of motion' and 'shape/contour of linear path' are also not specified by the grammatical element.

As one step in a program to ascertain any properties common to grammatically specified notions, the notions just found are gathered together in Table 1. For heuristic purposes, the notions are very provisionally divided into three groups on the basis of their relation to topology. In group (a) are the notions that properly belong, or are readily definable, in the actual mathematical system of topology. In group (b), the notions might not be part of topology proper but intuitively seem like those that are--and might be includable in a related mathematical system that could be constructed. In group (c) are the notions that fall outside of any usual conception of a mathematical system. The number of notions in the first two groups combined is 13 while the third has 6--an indication of a preponderant propensity for grammatical elements to specify quasi-topological notions. The ratio in this direction is in fact improved if we consider that even several notions in group (c)--the bottom three--resemble topological ones in the sense of involving relativistic relationships between quantities rather than absolutely fixed quantities.

(7) Table 1: Some notions found to be specified by grammatical elements

a. topological	b. topology-like
partition	same
region/side	different
point	pattern of distribution
line	"adjacency" of points (monotonicity)
locatedness	
within	
uniplexity	c. non-topological
multiplexity	
one-to-one correspondences	matter
	space
	time
	motion
	medium
	currently indicated/ communicating entity

For a complementary program of ascertaining any properties excluded from grammatical specification, the notions found above not to be specified by the elements investigated are listed in Table 2. Rather than topological, topology-like, or relativistic, these notions involve Euclidean-geometric concepts (e.g., set distance, size, contour), quantified measure, and various particularities of a quantity--in sum, characteristics that are absolute or fixed.

(8) Table 2: Some notions seemingly never specified grammatically

absolute/quantified magnitude (of distance, size, etc.)	kind of substance
	speed
shape/contour of line	color
sensorimotor characteristics	

The provisional conclusion to be drawn from these findings is that, if grammatical specifications largely correspond to (linguistic-) cognitive structuring, then the nature of that structuring is largely relativistic or topological rather than fixed or absolute.

In a search for contrasts between grammatical and lexical specification, a difference that presents itself at this point is that the relativism vs. absolutism restrictions do not apply to the latter. Lexical items can specify topological and relativistic concepts, as the very words listed in Table 1 attest to. And they can also specify Euclidean or absolute concepts. Thus, for the notion of color in Table 2, there are such lexical items as red, blue; for contour, there are circle, straight; for quantified magnitude, there are inch, mile; for sensorimotor characteristics, there are wade, nimble, effort.

For a further contrast between the grammatical and the lexical type of specification, we consider the full complement of both element-types in a single whole sentence, viz., that selected in (9):

(9) A rustler lassoed the steers.

We first list the grammatical elements present in the sentence and the notions that they specify:

(10)

a.	-ed:	'occurring at a time before that of the present communication'
b.	the:	'has ready identifiability for the addressee'
c.	a:	'not before in discussion or otherwise readily indentifiable for addressee'
d.	-s:	'multiplex object'
e.	a...-Ø:	'uniplex object'
f.	the grammatical category of "verb" for lasso:	'eventhood'
g/h.	the gram. category of "noun" for rustler/steer:	'objecthood' (one possible spec. of "N")
i/j.	the grammatical relations of "subject"/"object" for rustler/steer:	'agent'/'patient' (among possible specs.)
k.	active voice:	'point-of-view at the agent'
l.	intonation, word-order, state of auxiliaries:	'the speaker "knows" the situation to be true and asserts it'

The lexical items in the sentence can have their specifications characterized as follows:

(11) A complex of concepts involving:

a. rustler:	property ownership, illegality, mode of activity
b. steer:	appearance, physical makeup, relation to animal kingdom
	institution of breeding for intended purposes, esp. human consumption
c. lasso:	certain materials (a body and a lasso) in certain configurations
	movement sequences of materials' parts
	concomitant mental intentions, directings, monitorings, etc.

In surveying the lists, we can see these differences emerge: The grammatical elements are more numerous and their specifications seem simpler and more structural. Together, their specifications seem to determaine the main organizational and communicational delineations of the CR evoked by the sentence. The lexical elements are fewer in number, but their specifications are more complex and seem to comprise most of the content of the CR. The lexical specifications are complex in three ways: compared to a grammatical specification, each has a) more total information, b) greater intricacy of information, and c) more different types of information together.

These grammatical-lexical differences can be set into further relief by in turn varying one element-type while keeping the other constant. Thus, varying only the grammatical elements of (9), as is done in (12), seems to alter the organizational and communicational characteristics of the scene but to leave its basic contents intact:

(12) Will the rustlers lasso a steer?

Varying only (9)'s lexical elements, as in (13), shifts us to a new scene altogether, and yet the essential breakup of the scene and of the communicative setting seem to remain the same:

(13) A machine cancelled the stamps.

2. Categories of Grammatically Specified Notions

The preceding sampling of grammatical elements has yielded a set of notions helpful toward discovering common properties. But the set has been small and haphazardly arrived at. With a broader and more systematic investigation, patterns of organization become evident. Grammatically specified notions can be seen to pattern in categories, and the categories, in turn,

in integrated systems. In this section we look at some of these categories and systems.

The grammatical elements here will not be treated in isolation, but in association with lexical items. That is, the grammatically specified structural notions will be considered in interaction with that portion of lexical specification that is also structural. This interaction entails cognitive processing, and different cases of such processing will be considered along the way.

The note on methodology should be made that our direction of analysis has been from grammatical specification to category, not the reverse. That is, the categories considered below were discovered to be relevant to the specifications of various grammatical elements. They were not part of some a priori conceptual schema which then sought corroborative examples.

2.1 Dimension / Kind of Quantity

The category of "dimension" has two member notions, 'space' and 'time'. The kind of "quantity" that exists in space is--in respectively continuous or discrete form--'matter' or 'objects'. The kind of quantity existing in time is 'action' or 'events' ("action" is meant to refer to any obtaining circumstance not just (willed) motion). In tabular form, these notions relate thus:

(13) space: matter/objects
time: action/events

A number of grammatical and lexical referents are specific with regard to one or the other pole of this category. But since the category cross-cuts the ones treated next, we will not exemplify it here but will endeavor in the following to present both space and time examples side by side.

2.2 Plexity

The category here to be termed "plexity" is a quantity's state of articulation into equivalent elements. Where the quantity consists of only one such element, it is "uniplex", and where it consists of more than one, it is "multiplex". When the quantity involved is matter, plexity is, of course, equivalent to the traditional category of "number" with its component notions "singular" and "plural". But the present notions are intended to capture the generalization from matter over to action, which the traditional ones do not.[9]

Specifications as to plexity are made by both lexical items and grammatical elements, and the interplay between the two when they are both in association must be noted. Example English lexical items that basically specify a uniplex referent are--for matter and action, respectively--bird and (to) sigh. They can occur with grammatical elements that themselves specify a uniplexity, like those underlined in (14a) (many languages have here a more regular, overt system of markers than English). But they can also occur with grammatical elements that specify a multiplexity, as in (14b). In this association, such elements can be thought to trigger a particular cognitive operation--in this case, one of "multiplexing". By this operation, an original solo referent is, in effect, copied onto various points of space or time.

(14)	matter	action
a. uniplex	A bird flew in.	He sighed (once).
b. multiplex	Birds flew in.	He kept sighing.

The reverse of the preceding circumstances is also to be found in language. First, there are lexical items that intrinsically specify a multiplexity. English examples are furniture or timber (i.e., 'standing trees') for matter and breathe for action, as used in (15a). And, too, there are grammatical elements able to appear in association here, as in (15b), that signal an operation the reverse of multiplexing--one that can be called "unit-excerpting". By this operation, a single one of the specified equivalent units is taken and set in the foreground of attention.

(15)	matter	
	action	
a. multiplex	Furniture overturned in the 'quake.	
	She breathed without pain.	
b. uniplex	A piece of furniture overturned...	
	She took a breath/breathed in...	

The grammatical elements that above signaled multiplexing-- -s and keep -ing --have a directly manifested surface form. The ones signaling unit-excerpting are in part abstract in form, as represented in (16):

(16) $\underline{\text{matter}}$ $\underline{\text{action}}$
(a) N_{unit} of + ___ V_{dummy} (a) [___ + X]$_N$

eg: a piece of furniture take a breath

or: ___ + Prtcle (eg: in)

2.3 State of Boundedness

Another category of attributes specified both grammatically and lexically for a quantity is its "state of boundedness" When a quantity is specified as "unbounded", it is conceived as continuing on indefinitely with no necessary characteristic of finiteness intrinsic to it. When a quantity is specified as "bounded", it is conceived as demarcated off as an individuated unit entity.

Among English examples of lexical items, water and (to) sleep seem basically to specify unbounded quantities, whereas sea and (to) dress seem basically to specify bounded ones. These specifications are demonstrated by the words' respectively unacceptable and acceptable occurrence with the grammatical element "in $NP_{\text{extent-of-time}}$", which specifies boundedness:

(17)	matter	
	action	
a. unbounded	*We flew over water in 1 hr.	
	*She slept in 8 hrs.	
b. bounded	We flew over a sea in 1 hr.	
	She dressed in 8 mins.	

Now, there are grammatical elements suitable for co-occurrence with unbounded-type lexical items which therewith, in effect, trigger a cognitive operation of "bounding". By this operation, a portion of the specified unbounded quantity is demarcated and placed in the foreground of attention. Examples of such grammatical elements in English are:

(18) $\underline{\text{matter}}$ (a) $N_{\text{bounded-quantity}}$ of + ___

$\underline{\text{action}}$ for $N_{\text{extent-of-time}}$ + ___

Particular cases of them in use are:

(19) We flew over a body of water in 1 hr.
She slept for 8 hrs.

The question arises whether the reverse of the preceding circumstances is ever to be found in language. Entailed would be the existence of grammatical elements that, when used with lexical items specifying a bounded quantity, Would trigger an operation of "debounding". By this, e.g., the referent of sea would be shifted to 'pelagic water', and that of (a) tear, to take another lexical bounded case, would shift to 'lachrymal fluid'. It seems likely that such grammatical elements exist; the closest candidate known to the author is the French suffix -age, but this has a range of meanings and many occurrence restrictions--and does not, e.g., happen to combine with the French words for "sea" or "tear".[10]

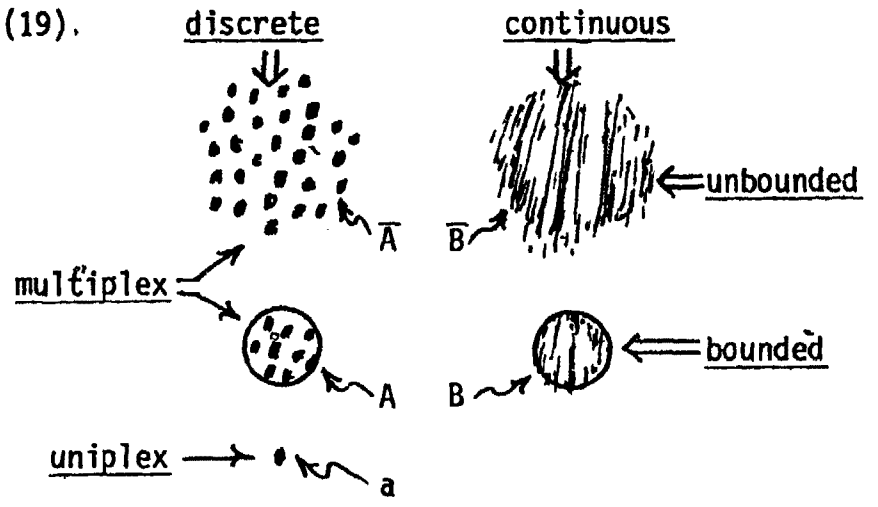
2.4 State of Dividedness

The category of "state of dividedness" refers to a quantity's internal consistency. A quantity is "discrete" (or "particulate") if there are breaks in its continuity. Otherwise, the quantity is "continuous".[11] Both lexical and grammatical elements are sensitive, in their specifications, to the distinctions of this category. But there appear to be no grammatical elements that solely specify discreteness or continuity for a quantity, and also none that signal an operation for reversing a quantity's lexically specified state of dividedness.[12] In consequence, there is difficulty in demonstrating this category explicitly by itself, and so we defer its treatment until the next section, where it can be seen in interaction with the other categories.

2.1 - 2.4 The Disposition of a Quantity

The preceding four categories of attributes all pertain to a quantity simultaneously and, taken together, can be considered to constitute a system of attributes that may be termed a quantity's "disposition". The particular intersections of the several attributes will be the main object of attention here. These, firstly, can be schematized as in (19):

(19).



\+ the distinction between matter and action, which cross-cuts all of the above[13]

Each intersection of attributes indicated here has been found specified by various lexical items. An example or two (most seen earlier) is given for each intersection in (20):[14]

(20)

Ā:	timber/furniture (to) breathe	B̄:	water (to) sleep
A:	(a) family (to) button up	B:	(a) sea/tear (to) zip up
a:	(a) bird (to) sigh		

Now if the particular contentful referent for which one chooses a lexical item happens to be wedded, by that lexical item, to an unwanted set of structural specifications, there generally are grammatical means available for altering this to a desired set. Such means range in directness from specifying the single apt alteration to involving a circuitous sequence of operations. A number of starting- and ending-points for alterations, and the means for accomplishing them, are indicated in (21):

(21)

Ā → A	a stand of timber breathe for 1 hr.	B̄ → B	a body of water sleep for 1 hr.
Ā → a	a piece of furnit. take a breath/ breathe in	-----	
A → a	a member of a fmly go through a step of buttoning up	-----	
A → Ā	members of a fmly (A → a → Ā) button on and on	B → B̄	tears (*tearage) (B → a → Ā → B̄) zip on and on
a → Ā	trees keep sighing	-----	
a → A	a stand of trees (a → Ā → A) sigh for a while	-----	

2.5 Degree of Extensionality

Implicit in the vertical dimension of the schematic arrangement in (19) is a further category[15] that can be called "degree of extensionality". This category has three member notions, terms for which are given in (22) together with schematics of the notions for the linear dimension:

(22)

point	bounded extent	unbounded extent
•	•——•	- - - ——— ..

Lexical items with either a matter or an action referent can make concurrent structural specifications for their referent as to its basic degree of extensionality. Three examples--specifying objects of different linear extensionalities--are the words

(23)

speck	ladder	river

Now a lexical referent that is perhaps most basically to be conceived as of one particular degree of extensionality can, by various grammatical specifications that induce a shift, be idealized as being of some other degree of extensionality. For a first example, consider the event referent of climb a ladder, which seems basically of bounded linear extent (of time), as is in fact manifested in (24) in conjunction with the grammatical element "in + $NP_{extent\text{-}of\text{-}time}$":

(24) She climbed up the fire-ladder in 5 mins.

With a different accompanying grammatical element, like the "at + $NP_{point\text{-}of\text{-}time}$" in (25), (as well as different contextual specifications), the event referent of the preceding can be shifted toward idealiz-

ation as a point of time--i.e., as being point-durational:

(25) Moving along on the training course,
she climbed the fire-ladder at exactly midday.

This shift in the cognized extensionality of the event can be thought to involve a cognitive process of "reduction" or of "taking the long-range view". The shift can also go in the other direction. The event referent can be idealized as an unbounded extent from the effect of grammatical elements like "keep -ing", "-er and -er", and "as + S", as in (26):

(26) She kept climbing higher and higher up the
fire-ladder as we watched.

Here there would seem to have taken place a cognitive process of "magnification" or of "taking the close-up view". In such a process, a perspective is established whereby the existence of any exterior bounds falls outside of view and attention--or, at most, are asymptotically approachable.

The preceding event referent was continuous, but a discrete case can exhibit the same shifts of extensionality. One such case, perhaps to be considered as most basically of bounded extent, is shown with that degree of extensionality in (27a). But the referent can also be idealized as a point, as in (27b) (it is clear that the cows here did not all die at the same moment, and yet the spread of their death times is conceptually collapsed into such a single moment). Or, the referent can be idealized as an unbounded extent, as in (27c):

(27) a. The cows all died in a month.
b. When the cows all died, we sold our farm.
c. The cows kept dying (and dying)
until the serum finally arrived.

The alternative idealizations of extensionality just seen as specifiable for an event referent are generally also available for an object referent. Thus, e.g., the referent of (a) box can be specified for idealization as a point or as a bounded extent (of area or volume). Some grammatical elements making such specifications are illustrated in (28). Also set forth here are the homologies between these and the event-specific elements:

(28)

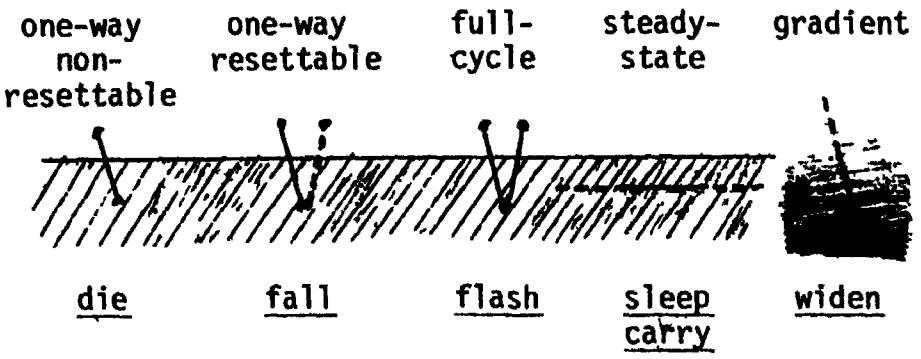
point	The box is 20 ft. away from the wall. I read the book 20 yrs. ago.
bounded extent	The box is 2 ft. across. I read the book in 2 hrs.
(point within) bounded extent	The ball is in the box. She arrived as I was reading the book.

2.6 Pattern of Distribution

The pattern of distribution of matter through space or of action through time is a further category of notions that can be both grammatically and lexically specified.[16] For action through time--the only dimension we will be looking at now--this category together with the preceding one largely constitute the traditional category of "aspect".

Several of the main patterns of distribution for action through time are shown schematically in (29) (the dots here, representing situatedness in complementary states, should really be adjacent, but they are sketched apart with a connecting line to show the crossing of state-interfaces). Shown, too, are example verbs whose basic distributional specifications are as in the corresponding schematic:

(29)



One can determine that these lexical items have the specifications indicated by noting the grammatical elements with which they can and cannot occur (or, to put the latter case in our terms: ...grammatical elements toward whose specifications they will not shift). A full demonstration is not in order here, but a few examples show the principle: The resettable type of a one-way event is distinguished from the non-resettable type by its compatibility in sentences like: He fell 3 times, which the other lacks: *He died 3 times. This same one-way form is distinguished from a full-cycle form by its ability to appear in sentences like: He fell and then got up, which the latter cannot do: *The beacon flashed and then went off.

We can now consider the cirsumstance where a verb of one type appears with grammatical elements of another type and shifts in certain of its specifications of distribution. For an example we again take die, whose basic specifications can be adjudged as point-durational one-way non-resettable--schematizable, now more precisely, as: . This verb is used with its basic specifications in a sentence like (30a).

(30) a. He died as she looked on.
b. He was (slowly) dying as she looked on.

But in a sentence like (30b), the grammatical element "be + -ing" induces a shift. In effect, the infinitesimal interval between the two states involved for die--viz., 'aliveness' and 'deadness'--is spread out, with the creation thereby of an extent-durational gradient. This is the shift in the distribution pattern's structural type. But concomitantly, a shift in the basic contentful referent is engendered. Instead of 'dying', the new gradient refers to 'moribundity'. The distinction becomes clear in noting that one can have been dying without having died, and, correlatively, one can have died without having been dying.[17]

2.7 Perspectival Mode

A specified action (which, in our terms, can as equally be static as involve change) has been seen to have its own, perhaps most basic, pattern of distribution through time. But, as it turns out, there can be independent specification for a mode of attending to the action that has a distinct temporal pattern of distribution, one that is either equal or unequal to the action's. In what we shall now consider, there are two types of such "attentional" or "perspectival mode" viz.:

(31) The assuming of:
a. a steady-state long-range perspective point with synoptic scope of attention
b. a moving close-up perspective point with local scope of attention

To illustrate, we first consider an example with a basically steady-state referent, viz., objects in location. The (31a) type of perspectival mode--the one more congruent with such a referent--holds in (32a), multiply specified/determined there by the set of grammatical elements shown underlined. But by substituting grammatical elements coding for the (31b) perspectival mode, as is done in (32b), the scene evoked can be shifted to one where one's mental gaze or one's own projected location jumps in turn from object to object. In effect, a steady-state multiplexity of objects has been converted to a sequential multiplexity of events, viz., of conceptualized encounters with the objects.

(32) a. There are houses here and there in the valley.
b. There is a house every now and then through the valley.

In a comparable case, the moving-perspective form, shown in (33b), is the only mode that can be specified using everyday language. One must resort to scientific language, as in (33a), in order to establish the synoptic perspective:

(33)

a. The telephone poles' heights form a gradient that correlates with their locations on the road.
b. The telephone poles get taller the further down the road they are.

The reverse of the preceding circumstances is also encountered. An example involving a sequential multiplexity of events is shown in (34a) with the more congruent moving-perspective mode specified. In (34b), the same referent instead becomes the object of synoptic viewing. In metaphorical terms, the effect here is as if the vertical time line is tilted up into present-moment horizontality for integrated or summational assessment.

(34)

a. I took an aspirin time after time during/ in the course of the last hour.
b. I have taken a number of aspirins in the last hour.[18]

2.8 Level of Synthesis

The category to be considered now pertains to bounded quantities, like those schematized in the A/B row in (19). One form of locution already seen to specify such quantities is the particular type of "NP of NP" construction illustrated in (35a). Here the second NP specifies the identity of the quantity involved, itself conceptualized as without intrinsic bounds, while the first NP specifies the bounding (or "portion-taking") per se of the quantity:

(35) a. a set of trees — a body of water
b. a cluster of trees — a puddle/drop of water

Now, beyond the fact alone of bounding off a portion, the first NP can additionally specify the particular configuration or form that the portion takes, as in (35b).[19] Especially with regard to internally discrete quantities--as with a cluster of trees--the two NPs can here be seen as coding for two different "levels of synthesis": The later NP specifies an unsynthesized multiplexity, while the earlier NP specifies a particular geatalt synthesized therefrom.

There is a further cognitive distinction involved here that language usually makes: either level of synthesis can be placed in the foreground of attention while the other level is placed in the background. One grammatical form that specifies this involves placing the foregrounded NP-type first, as shown in (36a). With the use of this grammatical device, moreover, predications can be made that pertain solely to one level of synthesis or the other, as seen in (36b):

(36) a. the cluster of trees / the trees in the cluster
b. That cluster of trees is small.
≠ The trees in that cluster are small.

There are certain surface forms, furthermore, whose referents are keyed to applying to only one or the other level of synthesis. Thus, together (toward each other) tends to correlate with multiple objects, while in (upon itself) tends to correlate with a composite thereof:

(37) The bricks in the pyramid came crashing together/?in.
The pyramid of bricks came crashing in (upon itself)/?together.

The preceding has involved shifting attention from a multiplexity to the gestalt that it constitutes. Also encountered in language are means for specifying the reverse: shifting attention from a gestalt to the components that constitute it. This procedure can take place when the starting lexical item specifies an entity taken to be already at the more synthetic level, as is the case with iceberg in (38a). By grammatical devices like those seen in (38b), such an entity can be broken down from conception as a coherent whole and presented in terms of component parts and their interrelations:

(38) a. The iceberg broke in two.
b. The two halves of the iceberg broke apart (*in two).

Again we encounter a surface form--in two--that correlates with only one level of synthesis and not the other.[20]

2.9 Level of Exemplarity

The specification for a multiplexity of objects can have a further cognitive distinction made pertaining to it. This distinction does not affect the basic reference to all the members of the multiplexity, but addresses how attention is directed therein. Either the full complement of the multiplexity is in the foreground of attention, with perhaps individual items here and there singled out in the background of attention. Or a single exemplar out of the multiplexity is placed in the foreground of attention, with the remaining items more dimly conceived in the background of attention. Perhaps most languages have several grammatical devices for specifying this distinction as to the "level of exemplarity". But English stands out in the extensiveness of its forms:

there are different pairs of grammatical elements that mark the distinction for a number of distinct types of multiplexity. A rather full list of these pairs is illustrated in (39):

(39)

a. Oysters have siphons/a siphon.
   An oyster has siphons/a siphon.[21]

b. All oysters have siphons/a siphon.
   Every oyster has siphons/a siphon.

c. All the members raised their hand(s).
d. Each member raised his hand(s).

d. Many members raised their hand(s).
   Many a member raised his hand(s).

e. Some members here and there raised their hand(s).
   A member here and there raised his hand(s).

f. Members one after another raised their hand(s).
   One member after another raised his hand(s).

g. Hardly any members raised their hand(s).
   Hardly a member raised his hand(s).

h. No members raised their hand(s).
   No member (Not a member) raised his hand(s).

i. She held a gun in both hands.
   She held a gun in either hand.[23]

2.10 Other Categories and Processes

More notional categories and cognitive processes have been worked up than there is opportunity to present here. Some of this other material is treated in an earlier work, Talmy (1977) (which itself lacks some of the material presented here). But we will briefly indicate some of the concepts involved.

The adjectives in a pair like sick/well behave differently in association with grammatical elements specifying vectoral degree, as shown in (40). In this they parallel the behavior of certain spatial expressions like at the border/past the border:

(40)
He's slightly { sick/past the border. / *well/*at the border. }
He's almost { well/at the border. / ?sick/?past the border. }

This behavior can be accounted for by positing that such adjectives are not simply "opposites", but, rather, imply for some semantic notion, e.g., that of 'health', a particular abstract topological axis of which each adjective labels a certain portion. The forms here seem in particular to imply a line bounded at one end; well refers to the end-point while sick refers to the remainder of the line. These are the lexical items' "axial characteristics", i.e., the particular (topological) relations each has to a particular semantic axis and to other items along the same axis. Certain grammatical elements, like those underlined in (40), also specify axial characteristics. Used incompatibly, they can cause a shift in an associated adjective's specifications. Thus, in (41), sick seems to label an end-point, and of a different axis as well, that of 'feeling bad':

(41) (After eating the shrimp, he felt worse and worse and) he was almost sick at one point/ he finally got sick in 5 hrs.

Lexical expressions like cottage and hotel room may be taken to have "associated characteristics"--here, respectively, those of 'permanent residence' and 'temporary lodging'. These attributes may mesh or conflict with the specifications of another element in the same sentence, e.g., with the directional adverb home, which specifies a permanent residence. In the case of conflict, as in (42b), the lexical item is operated on by a cognitive process that leaves its essential characteristics intact but replaces its incidental characteristics:

(42) a. He drove home to his cottage in the suburbs.
     b. He drove home to his hotel room.

The "scene-breakup characteristics" of a lexical item like serve refer to its basic specification of a dyadic event, in particular, a social event involving the two roles of 'host' and 'guest', as is manifested in (43a). But in a sentence like (43b), such a lexical item shifts to specifying a monadic event comparable to a basically monadic lexical expression like that in (43c). This shift in (42b) takes place in accommodation of the subject-plus-reflexive's single-role specification. (Though this grammatical element is determinative in setting the role-number as monadic, the verb's influence remains: blended in here is the metaphoric suggestion of a dyad, as if both 'host' and 'guest' are to be found in the "I"):

(43) a. The host served me some dessert from the kitchen.
     b. I served myself some dessert from the kitchen.
     c. I went and got some dessert from the kitchen.

A major aim in cognitive linguistics must be to investigate the interactions between lexical and grammatical specifications arising in a single sentence. Included here are the cognitive accommodations that take place where there are conflicting specifications. A number of interactions have been provisionally identified, and four seem definitely established: operations, shifts, blends (of two kinds: superimposed and introjected), and juxtapositions. The last three of these are treated at length in Talmy (1977).

2.11 Nesting

The operations and shifts seen in 2.1 - 2.6 need not take place singly. The output of one can serve as the input to another, up to as many as five hierarchical levels of "nesting". While there are a number of interesting examples of this for different types of matter and action, we will go directly to illustrating one of the longest cases:

(44)

a. The beacon flashed (as I glanced over).
b. The beacon kept flashing.
c. The beacon flashed 5 times in a row.
d. The beacon kept flashing 5 times at a stretch.
e. The beacon flashed 5 times at a stretch for 3 hrs.

In (44a), the lexical verb flash appears with its basic structural specification as a point-durational full-cycle uniplex event. This undergoes the process of multiplexing, to yield the unbounded multiplexity in (44b). This then undergoes bounding in (44c). This bounded multiplexity is then first put through

the process of reduction to become idealized as a point. and this is in turn multiplexed, yielding (44d). This new unbounded multiplexity is finally then bounded in (44e). The nesting of structural specifications in this last stage can be represented schematically as in (45):

(45) [('''''') - ('''''').....(''''''') - ('''''')]

3. Further Cognitive Connections

Grammatically specified structuring appears to be similar, in certain of its characteristics and functions, to the structuring in other cognitive domains, notably that of visual perception. In particular, the characteristic of being quasi-topological can be pointed to, and three major functions can be identified: classification, synoptics, and continuity. The thinking here is not equally far along on all these matters, but something of its directions can be indicated.

Grammatical specifications can be seen to constitute a classification with regard to the vast variety of learned, conceived, and perceived material. They gather different portions of the material together into subdivisions distinct from each other. By this, any particular currently cognized element is associated with its implicit "subdivision-mates". An illustrative case here are the twenty-odd motion-related prepositions in English, such as through and into, which together subdivide the domain of paths considered with respect to reference-objects'. This domain covers a great and varied range, but any particular "path" falls within the purvue of one or another preposition, associated there with other "paths" The associations are often language-specific and sometimes seem arbitrary or idiosynchratic. Thus, as seen earlier, classed together by through are such dissimilar cases as a straightforward liquid-parting course (walking through water) and a zig-zag obstacle-avoiding course (walking through timber). The question arises why such distinctions should be effaced by the grammatical system, while they are observed by the lexical and other cognitive systems. Why are grammatical elements--say, such prepostions--not a large and open class marking indefinitely many distinctions? One may speculate that the cognitive function of such classification lies in rendering contentful material manipulable--i.e., amenable to transmission, storage, and processing--and that its lack would render content an ineffective agglomeration.

The original assumption made in this paper about grammatical specification involved the synoptic function. That is, the grammatical elements of any particular sentence together specify the structure of the cognitive representation evoked by that sentence. Their specifications act as a scaffolding or framework across which contentful material can be splayed or draped. It can be speculated that such structure is necessary for a disparate quantity of contentful material to cohere in any sensible way or to be simultaneously cognized as a gestalt.

In the course of discourse, a great welter of notions pass in rapid succession. But there are several ways in which a cognitive continuity is maintained through this flux and a coherent gestalt is summated over time. For one, there are cognitive processes whereby the successive notions generally can be sensibly connected together or fit into a conceptual matrix. For another, rhetorical specifications --all the yes, buts, on the other hands, and a number of subtler elements not generally recognized for this--direct the illocutionary flow and make up the "logical" tissue of the discourse. Through this, grammatical elements appear to play a determinative role. Their specifications establish a structural level with greater temporal constancy amidst more fleeting aspects of content.

These forms of grammatically specified structuring seem to parallel forms discernable in the operation of visual perception.[24] First, the perception of any particular object is mediated by its association with related objects in a classificatory schema.

Secondly, the welter of visual sensations cognized at any given moment for some whole scene is rendered coherent by the perception of structural delineations running through it. One specialized form of this is discernable when one intends to move through a space, say, from one to the opposite corner of a restaurant. The sensations of tables, chairs,etc. are, in effect, perceived in simplified spatial arrangements as if from an aerial view, and the plot of a course one could follow through that is sensed.

Thirdly, in the course of motion through space over time, there is a great flux of visual sensations rushing past, but sense of continuity is maintained by the perception of structure running through the successive scenes. Two levels of "scene-structure constancy" are maintained. In the first, the perceived delineations afford greater permanence than the sensory flux, but do slowly shift. This is the level where, say, in walking past a table, its perceived outline is maintained but shifts gradually from a quadrilateral to a trapezoid and back to a quadrilateral. A deeper level of greater constancy is also maintained, from which the table continues to be perceived as a rectangle no matter where one is in relation to it. For a final parallel with grammatical specification, the topology-like nature of visual perception is evident here. For certain abstract characteristics of a scene and its contents are maintained constant while other, more metrical and Euclidean characteristics are free to vary without relevance thereto.

4. Notes

1. The word "evoke" is used because the relationship is not direct. The CR is an emergent, compounded by various cognitive processes out of the sentence elements' referential meanings, understanding of the present situation, general knowledge, etc.

Our term "cognitive representation" is similar in purport to Fillmore's (1975) "scene" but is chosen over that more specifically visual term. The linguistically evoked complex can have much from other sense modalities (notably som/kinesthetic and auditory) as well as meta-modal aspects.

2. Comprehension, rather than production, is the direction we limit ourselves to in the initial endeavor. This direction would seem to yield more immediately reliable findings, since its starting point is with more overtly manifest, hence handleable, forms like grammatical elements rather than with meanings and experiential complexes, which rely more on introspection and reports of introspection. Nevertheless, each direction does involve both the manifest and the experiential sides of language.

3. This is a classical linguistic distinction. A class in which morphemes are formally gathered is considered open if it is quite large and easily augment-

able relative to other classes. A class is considered closed if it is relatively small and fixed in membership.

4. Also includable here are "lexical complexes" like lodge a complaint or zero in on. Excluded are adverbs, which seem in all languages to derive from the other three open classes rather than from any open class of specifically adverbial stems.

5. Since the term "structure" has broad usage, we can help focus in on the intended sense with alternative terms: "principles of organization", "pattern of delineations", "schematic framework".

6. The fact of dual lexical specifications that can lead to conflict is a mojor issue that will be treated below under shifts. Some grammatical elements also cross the line and make contentful specifications along with structural ones. This is a more tangential issue that can be touched on here. The crossing ranges from the incorporation of a single contentful notion to the orderly interweaving of contentful and sturctural notions. Thus, upon in We rode/sailed/rushed upon the enemy incorporates the notion of 'attack', seemingly equivalent to the paraphrase 'into attack upon'. The closed-class adverb tomorrow is equivalent to the phrase 'during the day that occurs next after the day during which I am now speaking', an example of an organized interlacing.

7. One can note, for example, the effect on one's internal cognitive representation in considering first the sentence I looked at the dog and then I looked at the dogs. The addition of the grammatical element -s has a major effect on the delineational breakup of--to put it visually--the scene before the mind's eye.

8. For example, augmentative and diminutive inflections, insofar as they refer to actual size, seem to specify size relatively greater or lesser than the norm for an object. And grammatical elements specifying distance (like English way and just appearing, e.g., before up there) appear to specify notions of 'far' and 'near' that are relative to the current situation.

9. It is true that there are the traditional terms "semelfactive" and "iterative" referring, respectively, to one and more than one instantiation of an event. But there is no real equivalent to number: "aspect" includes too much else about the temporal structure of action. And in any case, none of the traditional terms refer generally to both the dimensions.

10. The mechanism actually resorted to by both English and French in many such cases, including that of tear, is the use of the plural, as in:

(i) Tears flowed through that channel in Hades.

There seems to be a sequence of cognitive operations here in getting from a bounded to an unbounded quantity. Speculatively, the bounded quantity is first treated as a uniplex entity, it is then multiplexed, the resultant entities are conceived as spatially juxtaposed, and their boundaries are lastly effaced.

11. The present category may be prone to confusion with the preceding one. Contributory here is the normal meaning range of continuous, which as easily covers 'boundlessness' as it does 'internal seamlessness'. However, the two categories can vary independently. Thus, in the preceding section, the lexical examples given for unboundedness, water and sleep, happened also to be internally continuous; but the same demonstration of unboundedness could have been made with internally discrete examples like timber and breathe

12. There do exist certain mechanisms for such reversal. Thus, taking an unbounded case, the continuity-specifying word water can be shifted toward being cognized as discrete by the locution particles of water, as in:

(i) Water/Particles of water filled the vessel.

However, the grammatical complex used here does not directly specify the shift but, like the one in Note 10, seems to involve a several-atage route of cognitive operations.

13. For schematizing action along the one-dimensional time axis, an adaptation of the two-dimensional A, B, A, and B diagrams would be necessary--and can be readily visualized.

14. The lexical types for several of these intersections, it should be noted, do have traditional terms. Thus, nominal forms of the a, A, and B types, respectively, have been called count nouns, collective nouns, and mass nouns. And verbal forms of the a and B types, respectively, have been called punctual and durative verbs. The matrix presented here augments, systematizes, and generalizes the traditional notions.

15. It may be considered an extension of the category of state-of-boundedness via the incorporation of the notion of uniplexity.

16. This category might be considered an extension or generalization of the "disposition of a quantity". Clearly, this category and the preceding five all belong together in treating the greater disposition of a quantity, but the relationships have not yet all been worked out.

17. Our main purpose here is to note the shift in structure type. The shift in content, which will doubtless prove to have some regularity is not clearly understood at this point.

18. A major function of perfect forms in language indeed appears to be the one involved here. More particularly, the perfect seems able to specify the temporal counterpart of matter located within a bounded extent of space, as in (i). That is, a sentence containing the perfect, as in (ii), suggests a paraphrase like that in (iii), which is homologous with (i):

(i) There were 5 aspirins on the table.
(ii) I have taken 5 aspirins in the last hour.
(iii) There were 5 aspirin-takings in the last hour.

(In support of this interpretation, as pointed out to me by Peyton Todd, the per[illegible] can be noted always to involve a temporal span bounded at both ends.)

19. All three notion--identity of a quantity, portion-taking of a quantity, configuration of the portion--are generally specified simultaneously (or, "conflatedly" --see Talmy (1975)) by lexical items that would fit in the A/B row of (20). For example, (a) tear specifies not only a certain shape of quantum, but also the

material involved: lachrymal fluid. Such words generally do not participate in an "NP of NP" construction --like *a tear of milk--unless they in fact accede to a shift toward the type of word represented in drop.

20. There is a foursome of apt terms that can be applied to the two levels of synthesis in the two directions of shift, as indicated in (i). Employed here is the term "Figure" as it is used in my other work (Talmy 1978, 1976):

(i) cluster: "composite Figure" iceberg: "meta-Figure"
trees: "multiple Figures" 2 halves: "component Figures"

21. For the plural form oysters, the plural form siphons is ambiguous as to whether there are one or more siphons per oyster. All the other combinations unambiguously indicate the number of siphons per oyster. Thus, the exemplar form is always unambiguous in this reagard--one of its advantages over the full-complement form. This same arrangement holds through the list.

22. I have long wondered what the differences between each and every might be. One apparent difference shows up here. Each seems to be the exemplar counterpart of all the but not of all without the (*Each oyster has a siphon makes a poor generic assertion). Every is not constrained in this way, though it does strike me as more comfortably the counterpart of all without the.

23. One more pair can be added to this list by adjoining two complementary unpaired forms from two different languages. The English form some, as in some friends of mine, requires the plural and has no singular counterpart. The Italian form qualque, as in qualque amico mio, requires the singular and lacks a plural.

24. It seems likely that the language-related portions of the brain could have evolved to their present functions only in the presence of these already existing cognitive mechanisms and have incorporated their operation.

5. References

Fillmore, C. 1975. An alternative to checklist theories of meaning. In: Berkeley studies in syntax and semantics, vol 1. University of California, Berkeley.

-----. 1976. The need for a frame semantics within linguistics. In: Statistical methods in linguistics. Stockholm: Skriptor.

Talmy, L. 1975. Semantics and syntax of motion. In: Syntax and semantics, vol. 4. J. Kimball, ed. Academic Press.

-----. 1976. Semantic causative types. In: Syntax and semantics, vol. 6. M. Shibatani, ed. Academic Press.

----. 1977. Rubber-sheet cognition in language. In: Papers from the 13th regional meeting, Chicago linguistic society. University of Chicago.

-----. 1978. Figure and Ground in complex sentences. In: Universals of human language. Greenberg, Ferguson, Moravcsik, eds. Stanford University.

27

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 expectations 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 'proto-typical 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:

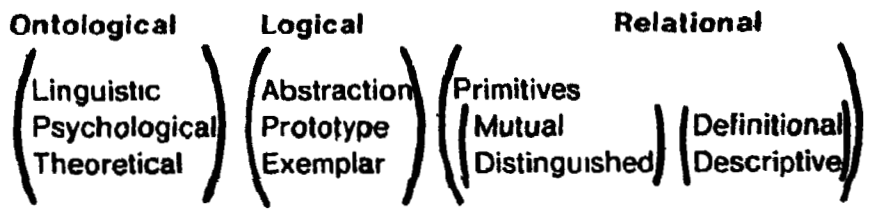
PRIMITIVES. The most straightforward use of semantic primitives would be a system in which the full meaning of any word or phrase could be expressed as a structure whose components are chosen from a small set of primitives, combined according to a well-defined set of rules. No existing system is pure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

Taxonomic Lattice Structures for Situation Recognition

William A. Woods
Bolt Beranek and Newman Inc.
50 Moulton Street
Cambridge, MA 02138

1. The Role of a Knowledge Network for an Intelligent Machine

The kinds of intelligent computer assistants that we would like to be able to construct are very much like intelligent organisms in their own right. Imagine for a moment an intelligent organism trying to get along in the world (find enough food, stay out of trouble, satisfy basic needs, etc.). The most valuable service played by an internal knowledge base for such an organism is to repeatedly answer questions like "what's going on out there?", "can it harm me?", "how can I avoid/placate it?", "Is it good to eat?", "Is there any special thing I should do about it?", etc. To support this kind of activity, a substantial part of the knowledge base must be organized as a recognition device for classifying and identifying situations in the world. The major purpose of this situation recognition is to locate internal procedures which are applicable (appropriate, permitted, mandatory, etc.) to the current situation.

In constructing an intelligent computer assistant, the roles of knowledge are very similar. The basic goals of food getting and danger avoidance are replaced by goals of doing what the user wants and avoiding things that the machine has been instructed to avoid. However, the fundamental problem of analyzing a situation (one established either linguistically or physically or by some combination of the two) in order to determine whether it is one for which there are procedures to be executed, or one which was to be avoided (or one which might lead to one that is to be avoided), etc. is basically the same. For example, one might want to instruct such a system to remind the user in advance of any upcoming scheduled meetings, to inform him if he tries to assign a resource that has already been committed, to always print out messages in reverse chronological order (when requested), to assume that "the first" refers to the first day of the upcoming month in a future scheduling context and the first day of the current month in a past context, etc.

The principal role of the knowledge network for such a system is essentially to serve as a "coat rack" upon which to hang various pieces of advice for the system to execute. Thus the notion of procedural attachment becomes not just an efficiency technique, but the main purpose for the existence of the network. This does not necessarily imply, however, that the procedures involved consist of low-level machine code. They may instead, and probably usually will, be high level specifications of things to be done or goals to be achieved. The principal structure that organizes all of these procedures is a conceptual taxonomy of situations about which the machine knows something.

To support the above uses of knowledge, an important characteristic required of an efficient knowledge representation seems to be a mechanism of inheritance that will permit information to be stored in its most general form and yet still be triggered by any more specific situation or instance to which it applies. Moreover, the nodes in the network (or at least a major class of nodes) should be interpretable as situation descriptions. One of the most fundamental kinds of information to be stored in the knowledge base will be rules of the form "if <situation description> is satisfied then do <action description>", or "if <situation description> then expect <situation description>". Situation descriptions are in general characterizations of classes of situations that the machine could be in. They are not complete descriptions of world states, but only partial descriptions that apply to classes of world states. (The machine should never be assumed or required to have a complete description of a world state if it is to deal with the real world.) A situation in this partial sense is defined by the results of certain measurements, computations, or recognition procedures applied to the system's input. Examples of situations might be "You have a goal to achieve which is an example of situation Y" "You are perceiving an object of class Z", "The user has asked you to perform a task of type W", etc.

More specific situations might be: "trying to schedule a meeting for three people, two of which have busy schedules", "about to print a message from a user to himself", "about to refer to a date in a recent previous year in a context where precision but conciseness is required".

The major references to this conceptual taxonomy by the intelligent machine will be attempts to identify and activate those situation descriptions that apply to its current situation or some hypothesized situation in order to consider any advice that may be stored there. Note that "considering advice of type X" is itself an example of a situation, so that this process can easily become recursive and potentially unmanageable without appropriate care.

Conceptually, one might think of the process of activating all of the descriptions that are satisfied by the current situation as one of taking a description of the current situation and matching it against descriptions stored in the system. However, there are in general many different ways in which the current situation might be described, and it is not clear how one should construct such a description.

Moreover, until it is so recognized, a situation consists of a collection of unrelated events and conditions. The process of recognizing the elements currently being perceived as an instance of a situation about which some information is known consists of discovering that those elements can be interpreted as filling roles in a situation description known to the system. In fact, the process of creating a description of the current situation is very much like the process of parsing a sentence, and inherently uses the knowledge structure of the system like a parser uses a grammar in order to construct the appropriate description. Consequently, by the time a description of the situation has been constructed, it has already been effectively matched against the descriptions in the knowledge base.

2. Parsing Situations

As suggested above, the process of recognizing that a current situation is an instance of an internal situation description is similar to the process of parsing a sentence; although considerably more difficult due to a more open ended set of possible relationships among the "constituents" of a situation. That is, whereas the principal relationship between constituents in sentences is merely adjacency in the input string, the relationships among constituents of a situation may be arbitrary (e.g. events preceding one another in time, people, places, or physical objects in various spatial relationships with each other, objects in physical or legal possession of people, people in relationships of authority to other people, etc.) However, the basic characteristic of parsers, that the objects recognized are characterized as structured objects assembled out of recognizable parts according to known rules of assembly, is shared by this task of situation recognition.

Note that it is not sufficient merely to characterize a situation as a member of one of a finite number of known classes. That is, where it is not sufficient for a parser to simply say that its input is an example of a declarative sentence (one wants to be able to ask what the subject is, what the verb is, whether the sentence has past, present or future tense, etc.), in a similar way it is insufficient to merely say that an input situation is an example of someone doing something. One must generate a detailed description of who is doing what to whom, etc.

It is also not sufficient to characterize a situation as a single instance of an existing concept with values filled in for empty slots. In general, a situation description must be a composite structured object, various subparts of which will be instances of other concepts assembled together in ways that are formally permitted, in much the same way that the description of a sentence is put together from instances of noun phrases, clauses, and prepositional phrases. The specific instance built up must keep track of which constituents of the specific situation fill which roles of the concepts being recognized. Moreover, it cannot do so by simply filling in the slots of those general concepts, since a general concept may have multiple instantiations in many situations. Rather, new structures representing instances of those concepts must be constructed and pairings of constituent roles from the concept and role fillers from the current situation must be associated with each new instance.

3. The Process of Situation Recognition

The process of situation recognition consists of detecting that a set of participants of certain kinds stand in some specified relationship to each other. In general, when some set of participants is present at the sensory interface of the system (immediate input plus past memory), the task of determining whether there is some situation description in memory that will account for the relationships of those inputs is not trivial. If the total number of situation descriptions in the system is sufficiently small, all of them can be individually tested against the input to see if any are satisfied. If the

number of such descriptions is sufficiently large, however, this is not feasible.

Alternatively, if there is some particular participant that by virtue of its type strongly suggests what situation descriptions it might participate in, then an index from this participant might select a more manageable set of situation descriptions to test. Even in this case, however, the number of situations in which the constituent could participate may still be too large to test efficiently. In the most difficult situation, no single participant in the input is sufficiently suggestive by itself to constrain the set of possible patterns to a reasonable number. However, it may still be that the coincidence of several constituents and relationships may suffice, providing that the coincidence can be detected. It is this problem of coincidence detection that I believe to be crucial to solving the general situation recognition problem.

As an example, consider the following fragment of a protocol of a commander giving commands to an intelligent display system:

Cdr: Show me a display of the eastern Mediterranean.
[computer produces display]

Cdr: Focus in more on Israel and Jordan.
[computer does so]

Cdr: Not that much; I want to be able to see Port Said and the Island of Cyprus.

In the first clause of the third command of this discourse, (i.e. "not that much"), there is no single word that is strongly suggestive of the interpretation of the sentence. Moreover, there is nothing explicit to suggest the relationship of this clause to the one that follows the semicolon. The latter, if interpreted in isolation, would merely be a request for a display, or perhaps a succession of two displays, while in the context given, it is a request to modify a previous display.

There are two methods that I believe may be sufficient, either individually or in combination, to model coincidence detection. One is the use of factored knowledge structures that merge common parts of alternative hypotheses. The other involves the use of a markable classification structure in which the individual recognition predicates triggered by the ongoing discourse will leave traces of their having fired, so that coincidences of such traces can be efficiently detected. I have been investigating a structure which I call a "taxonomic lattice", that combines some features of both methods.

3.1 Factored Knowledge Structures

Given a knowledge-based system with large numbers of situation-action rules, where it is infeasible to find the rules that match a given situation by systematically considering each rule, one needs to have some way of reducing the computational load. As mentioned before, one approach is to index the rules according to some salient feature that will be easily detectable in the input situation and can then be used to find a much more limited set of rules to apply. This has been done in many systems, including the LUNAR system for natural language question answering [Woods, 1973, 1977]. In that system, rules for interpreting the meanings of sentences were indexed according to the verb of the sentence and rules for interpreting noun phrases were indexed by the head noun. Although this approach reduces the number of rules that need to be considered, it has several limitations still. The first is that there may be some values of the index key for which there are still a large number of rules to consider. In the case of the LUNAR system, for example, the verb "be" had a large number of rules to account for different senses of the word. Another is that there can be certain constructions for which there is no single easily detected feature that is strongly constraining as to possible meaning. In this case, there is no useful index key that can be used to select a sufficiently constrained set of rules to try.

Another limitation of this indexing approach as the range of language becomes more fluent is that in certain elliptical sentences, the constraining key may be ellipsed, and although one can have the rules indexed by other keys as well, the remaining ones may not sufficiently constrain the set of rules that need to be considered. Finally, even when the set of rules has been constrained to a relatively small set, there is frequently a good deal of sharing of common tests among different rules, and considering each rule independently results in repeating these tests separately for each rule.

One approach to solving all of the above problems is to use what I have been calling a "factored knowledge structure" for the recognition process. In such a structure, the common parts of different rules are merged so that the process of testing them is done only once. With such structures, one can effectively test all of the rules in a very large set, and do so efficiently, but never consider any single rule individually. At each point in a factored knowledge structure, a test is made and some information gained about the input. The result of this test determines the next test to be made. As each test is made and additional information accumulated, the set of

possible rules that could be satisfied by the input, given the values of the tests so far made, is gradually narrowed until eventually only rules that actually match the input remain. Until the end of this decision structure is reached, however, none of these rules is actually considered explicitly. This principle of factoring together common parts of different patterns to facilitate shared processing is the basic technique that makes ATN grammars [Woods, 1970] more efficient in some sense than ordinary phrase structure grammars. It has also been used by the lexical retrieval component of the BBN speech understanding system [Woods et al., 1976; Wolf and Woods, 1977] and accounts for the efficiency of the finite state grammar approach of the CMU Harpy system [Lowerre, 1976]. A recent innovative use of this principle appears in Rieger's "trigger trees" for organizing spontaneous computations [Rieger, 1977].

Whether factored together or not, the task of accessing rules is not a simple one. One problem is that rules don't match the input letter-for-letter: rather, they have variables in them with various restrictions on what they can match. For example a rule might say that whenever an access is made to a classified file, then a record of the person making the request should be made. The description, "an access to a classified file" needs to be matched against the user's request (or some subpart of it) and in that match, the description "a classified file" will be matched against some specific file name. In this kind of situation, there is no natural ordering of the rules, analogous to the alphabetical ordering of words, that will help in finding the rules that are satisfied by the given situation. Nor is a structure as simple as the dictionary tree above adequate for this case.

Another problem is that a given situation may be matched by several rules simultaneously with differing degrees of generality. For example, there may be a rule that says "whenever access is made to a top secret file (more specific than classified), then check the need-to-know status of the user for that information and block access if not satisfied". In the case of a request to a top secret file, both of the above rules must be found, while in the case of an ordinary classified file, only the first should. The actual input, however, will not explicitly mention either "top-secret" or "classified", but will merely be some file name that has many attributes and properties, among which the attribute "classified" is not particularly salient.

3.2 Markable Classification Structures

Another technique that holds promise for situation recognition is the use of a markable classification structure in which coincidences of relatively non-salient events can be detected. The keystone of this approach is a technique that Quillian proposed for modeling certain aspects of human associative memory [Quillian, 1966, 1968]. Quillian's technique of "semantic intersection" consisted of propagating traces of "activation" through a semantic network structure so that connection paths relating arbitrary concepts could be detected. For example, his system was able to connect concepts such as "plant" and "nourishment" by discovering the "chain" equivalent to "plants draw nourishment from the soil". If the appropriate information were in the network, this technique would also find chains of indirect connections such as "Plants can be food for people" and "People draw nourishment from food." The method was capable of finding paths of arbitrary length.

The problem of finding connections between concepts in a knowledge network is like the problem of finding a path through a maze from a source node to some goal node. At the lowest level, it requires a trial and error search in a space that can be large and potentially combinatoric. That is, if one element of the input could be connected to k different concepts, each of which would in turn be connected to k others, and so on, until finally a concept that connected to the goal was discovered, then the space in which one would have to search to find a path of length n would contain k^n paths. However, if one started from both ends (assuming a branching factor of k also in the reverse direction), one could find all the paths of length n/2 from either end in only $2.k^{n/2}$.

If one then had an efficient way to determine whether any of the paths from the source node connected with any of the paths from the goal node, such search from both ends would have a considerable savings. This can be done quite efficiently if the algorithm is capable of putting marks in the structure of the maze itself (or some structure isomorphic to it), so that it can tell when reaching a given node whether a path from the source or the goal has already reached that node. However, without such ability to mark the nodes of the maze, the process of testing whether a given path from the source can hook up with a path from the goal would involve a search through all the paths from the goal individually, and a search down each such path to see if the node at the end of the source path occurred anywhere on that path. If this were necessary, then all of the advantage of searching from both ends would be lost.

The use of the graph structure itself to hold marks is thus critical to gaining advantage from this algorithm. Essentially, the nodes of the graph serve as rendezvous points where paths that are compatible can meet each other. The coincidence of a path from the source meeting a path from the goal at some node guarantees the discovery of a complete path without any path requiring more than a simple test at the corresponding node in the graph as each link is added to the path.

What is needed for situation recognition in a generalization of Quillian's semantic intersection technique in which the source and goal nodes are replaced by a potentially large number of concept nodes, some of which are stimulated by immediate input, and some of which are remembering recent activation in the past. Moreover, what is significant is not just simple paths between two nodes, but the confluence of marks from multiple sources in predetermined patterns. Moreover, unlike Quillian, who considered all connections identically in searching for paths, we will consider marker passing strategies in which marks can be passed selectively along certain links. Recently, Fahlman [1977] has presented some interesting formal machine specifications of Quillian-type spreading activation processes which have this characteristic.

4. The Structure of Concepts

In building up internal descriptions of situations, one needs to make use of concepts of objects, substances, times, places, events, conditions, predicates, functions, individuals, etc. Each such internal concept will itself have a structure and can be represented as a configuration of attributes or parts, satisfying certain restrictions and standing in specified relationships to each other. Brachman [1978] has developed a set of epistemologically explicit conventions for representing such concepts in a "Structured Inheritance Network", in which interrelationships of various parts of concepts to each other and to more general and more specific concepts are explicitly represented. The essential characteristic of these networks is their ability to represent descriptions of structured objects of various degrees of generality with explicit representation of the inheritance relationships between corresponding constituents of those structures. A concept node in Brachman's formulation consists of a set of dattrs (a generalization of the notions of attribute, part, constituent, feature, etc.) and a set of structural relationships among them. Some of these dattrs are represented directly at a given node, and others are inherited indirectly from other nodes in the network to which they are related.

Let us assume that each concept that the system understands is represented as a node in one of these structured inheritance networks. The network, as a whole, then serves as a conceptual taxonomy of all possible "entities" that the system can perceive or understand. Each node in this taxonomy can be thought of as a micro schema for the recognition of instances of that concept. Each has a set of dattrs with individual restrictions and a set of structural conditions that relate the dattrs to one another. These restrictions and structural conditions may themselves be defined in terms of other concepts defined by other micro schemata, and so on until a level of primitively defined, directly perceivable concepts is reached.

Each concept in the taxonomy can be thought of as having a level of abstractness defined as the maximum depth of nesting of its constituent structure. Instances of primitively defined concepts have level 0, constellations of those concepts have level 1, a concept having level 1 and lower concepts as dattrs has level 2, and so on. If a taxonomy contained only level 0 and level 1 concepts, then the situation recognition problem would be greatly simplified, since one never needs to recognize portions of the input as entities that participate as constituents of larger entities. The general problem, however, requires us to do exactly that. More seriously, the general case requires us to recognize a concept some of whose dattrs may have restrictions defined in terms of the concept itself. This is true, for example, for the concept of noun phrase in a taxonomy of syntactic constructions. Such recursively defined concepts have no maximum level of abstractness, although any given instance will only involve a finite number of levels of recursion. This potential for recursive definition must be kept in mind when formulating algorithms for situation recognition.

5. The Need for Inheritance Structures

As a result of having different levels of abstraction in one's taxonomy, an input situation will often satisfy several situation descriptions simultaneously, no one of which will account for all of the input nor supplant the relevance of the others. For example, adding a ship to a display is simultaneously an example of changing a display and of displaying a ship. Advice for both activities must be considered. Moreover, a single description may have several different instantiations in the current situation, with situation descriptions becoming arbitrarily complex

by the addition of various qualifiers, by the conjunction and disjunction of descriptions, etc. For example, one might want to store advice associated with the situation [wanting to display a large ship at a location on the screen that is within one unit distance from either the top, bottom, or side of the screen when the scale of the display is greater than 1:1000]. Finally, situation descriptions may subsume other descriptions at lower levels of detail, and advice from both may be relevant and may either supplement or contradict each other. For example, displaying an aircraft carrier is a special case of displaying a ship, and there may be specific advice associated with displaying carriers as well as more general advice for displaying any ship. Thus, conventions will be required to determine which advice takes precedence over the other if conflicts arise.

The organization of large numbers of such situation descriptions of varying degrees of generality so that all descriptions more general or more specific than a given one can efficiently be found is one thing we require of an intelligent computer assistant. In order to build and maintain such a structure, it is important to store each rule at the appropriate level of generality, relying on a mechanism whereby more specific situations automatically inherit information from more general ones. That is, when one wants to create a situation description that is more specific than a given one in some dimension, one does not want to have to copy all of the attributes of the general situation, but only those that are changed. Aside from conserving memory storage, avoiding such copying also facilitates updating and maintaining the consistency of the data base by avoiding the creation of duplicate copies of information that then may need to be independently modified and could accidentally be modified inconsistently.

For example, one may want to store advice about displaying geographical features, about displaying such features that cover an area, about displaying bodies of water, about displaying lakes, etc. Thus, information about finding the area covered by a feature would be stored at the level of dealing with such area-covering features. information about displaying water in a certain color would be stored at the level of displaying bodies of water, and information about having inlets and outlets would be stored at the level of lakes. In any specific situation that the system finds itself, many such concepts at different levels of generality will be satisfied, and the advice associated with all of them becomes applicable. That is, any more specific concept, including that of the current situation, inherits a great deal of information that is explicitly stored at higher levels in the taxonomy.

In the case of the situation descriptions that we are dealing with, even the specification of what dattrs a given concept possesses is stored at the most general level and inherited by more specific concepts. Thus, for example, the descriptions of attribute dattrs for color and weight are stored for a general concept of physical object. These dattrs are then inherited by any more specific concepts of physical objects, such as planes, ships, desks, and pencils.

6. The Taxonomic Lattice

I believe that a general solution to the situation recognition problem can be obtained by the use of a classification structure in which traces of individual elements of complex concepts can intersect to facilitate the discovery of coincidences and connections that may not be strongly inferable from constraining expectations. The structure that I propose to use is a version of Brachman's structured inheritance networks, in which descriptions of all potentially relevant situations are stored with explicit indications of general subsumption of one situation by another, and explicit indications of the inheritance of dattrs and of advice by one concept from another. This structure, which I have called a taxonomic lattice, is characterized by a multitude of situation descriptions at different levels of generality.

We say that a situation description S1 subsumes a description S2 if any situation satisfying S2 will also satisfy S1. In this case, S1 is a more general description than S2, and is placed higher in the taxonomy. For example, [displaying a portion of country] is a more specific situation than [displaying a geographical area], which is in turn more specific than [displaying a displayable entity]. All of these are subsumed by a general concept [purposive activity], which in turn is more specific than [activity]. Moreover, a given description can subsume many incomparable descriptions and can itself be subsumed by many incomparable descriptions. For example, an instance of [displaying a geographical area] is also an instance of [accessing a geographical area], [displaying information], and [using the display], and may possibly also be an instance of [responding to a user command].

The space of possible situation descriptions forms a lattice under the relation of subsumption. At the top of the lattice is a single, most general situation we will call T, which is always satisfied and can be thought of as the disjunction of all possible situations. Anything that is universally true can be stored here. Conversely, at the bottom of the lattice is a situation that is never

satisfied, which we call NIL. It can be thought of as the conjunction of all possible (including inconsistent) situations. Assertions of negative existence can be stored here.

At the "middle" level of the lattice are a set of primitive perceptible predicates -- descriptions whose truth in the world are directly measurable by the "sense organs" of the system. All classes above this level are constructed by some form of generalization operation, and all classes below are formed by some form of specialization. At some point sufficiently low in the lattice, one can begin to form inconsistent descriptions by the conjunction of incompatible concepts, the imposition of impossible restrictions, etc. There is nothing to prevent such concepts from being formed; indeed, it is necessary in order for the organism to contemplate, store, and remember their inconsistency.

There are a number of specific relationships that can cause one situation description to subsume another. A given situation description can be made more general by relaxing a condition on a dattr, by eliminating the requirement for a dattr, by relaxing the constraints of its structural description, or by explicitly disjoining it (or'ing it) with another description. A given description can be made more specific by tightening the conditions on a dattr, by adding a dattr, by tightening the constraints of its structural description, or by explicitly conjoining (and'ing) it with another description. These operations applied to any finite set of situation descriptions induce a lattice structure of possible situation descriptions that can be formed by combinations of the elements of the initial set. We refer to this structure as the virtual lattice induced by a given set of situation descriptions. Note that only a finite portion of this lattice need be stored with explicit connections from more specific to more general concepts. By processing this explicit lattice, one can test any given description for membership in the virtual lattice and assimilate any new situation description into the explicit lattice in the appropriate place corresponding to its position in the virtual lattice.

In operation, any situation description about which information is explicitly stored will be entered into the explicit lattice. Any situation that the machine can understand is in some sense already in the virtual lattice and needs only be "looked up" in it. One task we have set for ourselves to develop efficient algorithms to tell whether a given situation can be understood in terms of the concepts of the lattice and if so, to construct its corresponding description and explicitly record its relations to other concepts in the explicit lattice.

7. An Example

As an example of the situation recognition process using marker propagation in a taxonomic lattice, let us consider a simple case of interpreting the intent of a simple English sentence. The example chosen is not complex enough to require all of the machinery discussed, but is presented here to illustrate the mechanism. The major features of the situation recognition mechanism only become critical in interpreting commands that require several sentences to build up, or which depend on the current context in complex ways, but such situations are difficult to illustrate.

For our example, suppose that the system contained a concept for requests to display a geographical region, and the user's input request were "Show me the eastern end of the Mediterranean." The concept [request] contains dattrs for the requestor, the requestee, a description of the state that the requestor desires, a form of request (demand, order, polite request, expression of preference, etc.), and perhaps others, Requests can take many forms. Assume that we have stored in the system a rule that says "Any sentence of the form: 'show me NP' is a request to display that NP." This rule could be stored in the lattice as a piece of advice associated with the concept "A sentence of the form: 'show me NP'," in such a way that when a sentence of the indicated form was found, an instance of a display request would be created. At that point, this resulting display request would be placed in the lattice in such a way that all more general concepts of which it is an instance would be activated, and in particular, the concept of a request to display a geographical region would be activated.

The parsing of the original sentence can either be done by an ATN grammar, or by a version of the taxonomic lattice itself (one that characterizes a taxonomy of sentence types). Let us assume here that it is done by an ATN grammar that is closely coupled to a taxonomic lattice, with the ATN representing the syntactic information about sentence form and the taxonomic lattice representing general semantic information. As the ATN grammar picks up constituents of the sentence, it reaches states where it makes hypotheses about the syntactic roles that those constituents play in the sentence (e.g. "this is the subject", "this is the verb", etc.). Such hypotheses are then entered into the lattice, where they begin to activate the recognition conditions of concepts in the network. For example, in the taxonomic lattice there is a concept of an imperative sentence whose subject is the system, whose verb is "show", whose indirect object is the user and whose direct object is a displayable object.

As the parsing proceeds, the ATN will make assertions about the sentence it is building up, and it will not only be building up syntactic representations of constituents of the sentence, but will also be building up representations of possible meanings of those constituents. In particular, it will be building up a list of those concepts in the lattice of which the current constituent may be a restriction or instance and a list of the dattr-value pairings that have been found so far. If a parse path succeeds (i.e. reaches a POP arc), then a node in the taxonomic lattice corresponding to that hypothesis will be found or constructed. This node will have links to more general and more specific concepts, and will have its constituents linked to appropriate dattrs of those concepts. At the point when this concept node is found/constructed, a process of activation spreading will be launched in the lattice to find any advice that may be inherited by that concept. This process will also leave "footprints" in the lattice that will facilitate the detection of concepts of which the current one may itself be a dattr (or part of a structural condition).

In the example above, when the parser has parsed the initial portion of the sentence "show me", it has built up in its internal registers the information corresponding to the hypothesis that the sentence is an imperative, with subject "you" and indirect object "me". Moreover, it knows that (in input sentences) "you" refers to the system itself, while "me" refers to the speaker. It also knows that the main verb is the verb "show". Let us suppose that at this point, the parser decides to activate the corresponding taxonomic lattice nodes for the concepts [the system], [the user], and [the verb show] (possibly with pointers to the syntactic hypothesis being constructed and/or the labels SUBJECT, OBJECT, VERB, respectively). Ignoring for now whatever information or advice may be found associated with these concepts or their generalizations, the footprints that they leave in the network will intersect at a node [display request] which has dattrs for requestor, requestee, form of request, and requested thing. They also intersect at other concepts such as [imperative sentence], [active sentence], [action], and a more specific kind of display request [region display request], whose requested thing is a geographical region. This latter concept was created and inserted into the lattice precisely to hold advice about how to display geographical regions, and to serve as a monitor for the occurrence of such situations. Fig. 1 is a fragment of a taxonomic lattice showing the concepts of interest. (For details of the notation, see Brachman [1978], Woods and Brachman [1978].)

When the final noun phrase has been parsed and given an interpretation, the footprints that its activation leaves in the network will awaken the [region display request] node, which will then be fully satisfied, and the parser will create a corresponding instance node, with appropriate bindings for its dattrs. In processing the noun phrase, the parser will discover the adjective "eastern" and the noun "Mediterranean" and will activate the corresponding nodes in the taxonomic lattice. The concept [east] is an instance of [direction], which, among other things, is the restriction for a dattr of a concept [directionally determined subregion] that defines the meaning of such concepts as "north eastern Idaho". Another dattr of this same concept has the restriction [geographical region], which is on the superc chain from Mediterranean. Hence, footprints from "eastern" and "Mediterranean" will intersect at the concept [directionally determined subregion], causing an instance of that concept to be constructed as a possible meaning of the noun phrase. The [directionally determined subregion] concept itself has a superc connection to [geographical region], which happens to be the restriction for the "requested thing" dattr of the concept [region display request] which has already received marks for its other dattrs. Thus, the intersection of footprints from the various constituents of the sentence at this concept node has served to select this node out of all the other nodes in the network. Since the more general concept [display request] is on a superc chain from [region display request], it will also be activated, and advice from both places will be considered.

8. Conclusion

In situation recognition, the nodes of a taxonomic lattice structure serve as rendezvous points where footprints from various constituent elements of a concept can meet. This facilitates the detection of coincidences of related events, which in many cases will not be suggestive in isolation. The implementation of the kinds of operations described above involves a system of marker passing conventions for propagating the various "footprints" around the network, detecting coincidences, creating instance nodes, and propagating further markers when coincidences are found. A major portion of our current research involves the discovery of effective conventions for such marker passing operations. Other issues include working out conventions for how far markers should propagate (amounting to decisions as to where to rendezvous), deciding how much information a mark carries with it and to what extent marks are inherited, developing ways to allow a node to remember partial

intersections of marks in such a way that it can incrementally extend them as additional marks accumulate, identifying implications of the marker passing strategies on representational conventions, etc.

9. References

Brachman, R.J. (1978)
"A Structural Paradigm for Representing Knowledge," Technical Report No. 3605, Bolt Beranek and Newman Inc., Cambridge MA.

Fahlman, S.E. (1977)
"A System for Representing and Using Real-World Knowledge," Ph.D. dissertation, Dept. of Electrical Engineering and Computer Science, M.I.T.

Lowerre, B.T. (1976)
"The HARPY Speech Recognition System," Technical Report, Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Pa.

Quillian, M.R. (1966)
"Semantic Memory," Report No. AFCRL-66-189, Bolt Beranek and Newman Inc., Cambridge, Ma.

Quillian, M.R. (1968)
"Semantic Memory," in Semantic Information Processing (M. Minsky, ed.). Cambridge, Ma:M.I.T. Press., pp. 27-70.

Rieger, C. (1977)
"Spontaneous Computation in Cognitive Models," Cognitive Science 1, No. 3, pp. 315-354.

Wolf, J.J. and W.A. Woods (1977)
"The HWIM Speech Understanding System," Conference Record, IEEE International Conference on Acoustics, Speech, and Signal Processing, Hartford, Conn., May.

Woods, W.A (1970)
"Transition Network Grammars for Natural Language Analysis," CACM, Vol. 13, No. 10, October (reprints available).

Woods, W.A. (1973)
"Progress in Natural Language Understanding: An Application to Lunar Geology," AFIPS Conference Proceeding, Vol. 42, 1973 National Computer Conference and Exposition (reprints available).

Woods, W.A., M. Bates, G. Brown, B. Bruce, C. Cook, J. Klovstad, J. Makhoul, B. Nash-Webber, R. Schwartz, J. Wolf, V. Zue (1976)
Speech Understanding Systems - Final Report, 30 October 1974 to 29 October 1976, BBN Report No. 3438, Vols. I-V, Bolt Beranek and Newman Inc., Cambridge, Ma.

Woods, W.A. (1977)
"Semantics and Quantification in Natural Language Question Answering," to appear in Advances in Computers, Vol. 17, New York: Academic Press. (Also Report No. 3687, Bolt Beranek and Newman Inc., 1977).

Woods, W.A. and R.J. Brachman (1978)
"Research in Natural Language Understanding" - Quarterly Technical Progress Report No. 1 (BBN Report No. 3742), Bolt Beranek and Newman Inc., Cambridge, MA

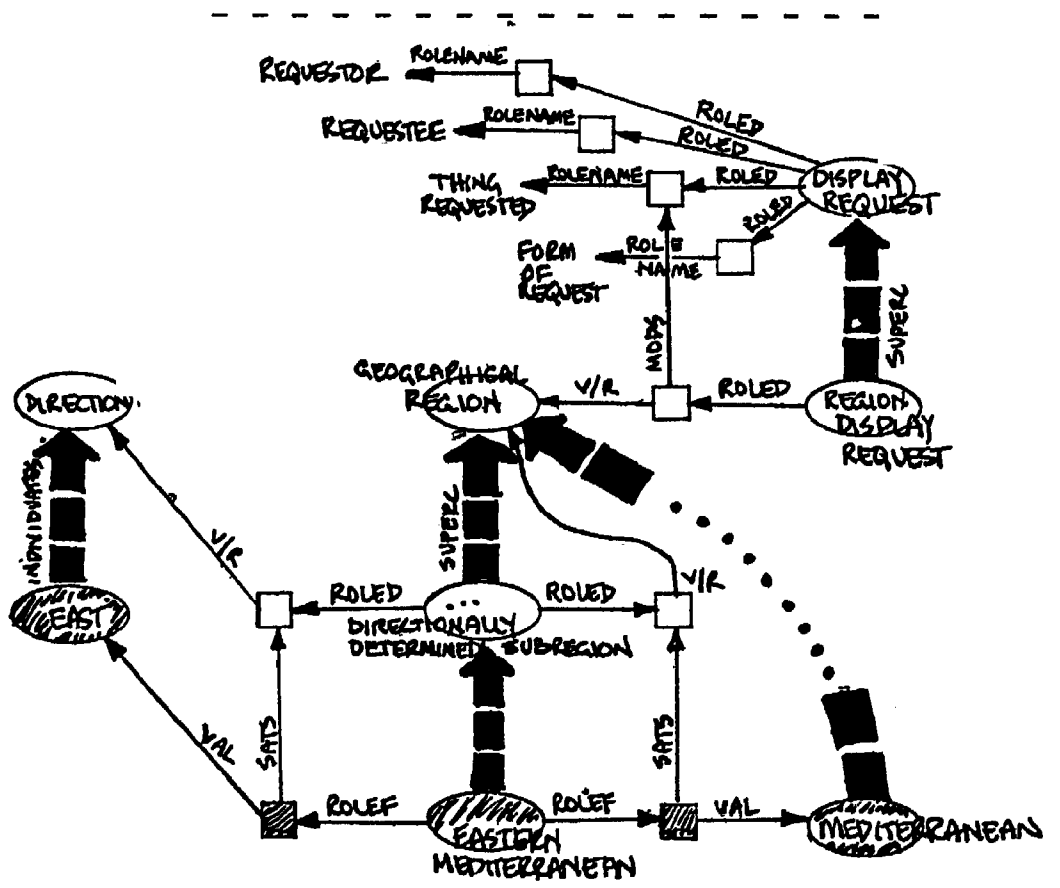


Fig. 1

Description Formation and Discourse Model Synthesis

Bonnie Lynn Webber
Bolt Beranek and Newman Inc.
50 Moulton Street
Cambridge MA 02138 <*1>

1. Introduction

Many researchers in linguistics, psychology, philosophy and artificial intelligence have recently begun to abandon a purely linguistic approach to **definite anaphora** <*2> in favor of a notion of reference into some kind of **model** of the discourse, cf. Karttunen [1976], Levin & Goldman [1978], Lyons [1978]. Stenning [1975]. My own research on definite anaphora (cf. Webber [1978a&b]) follows this approach, in particular making the following five assumptions:

1. One objective of discourse is to enable a speaker to communicate to a listener a **model** s/he has of some situation. Thus the ensuing discourse is, on one level, an attempt by the speaker to direct the listener in synthesizing a similar model.

2. Such a **discourse model** can be viewed as a structured collection of entities, organized by the roles they fill with respect to one another, the relations they participate in, etc.

3. The function of a **definite anaphoric expression** is to refer to an entity in the speaker's discourse model (DM_S). <*3> In using a definite anaphor, the speaker assumes (a) that on the basis of the discourse thus far, a **similar** entity will be in the listener's model (DM_L) as well and (b) that the listener will be able to access that entity via the given definite description or definite pronoun.

4. The **referent** of a definite anaphor is thus an entity in DM_S, which the speaker presumes to have a counterpart in DM_L. Discourse entities may have the properties of individuals, sets, events, actions, states, facts, beliefs, hypotheses, properties, generic classes, typical set members, stuff, specific quantities of stuff, etc.

5. In deciding which discourse entity a definite anaphor refers to, a listener's judgments stem in part from how the entities in DM_L are **described.** (When a discourse entity **E** is the referent of a definite anaphor **A**, one might distinguish that description of **E** conveyed to the listener by the immediately preceding text and consider it **A**'s **antecedent.**)

The point of making these assumptions explicit is to stress that insofar as reasoning about discourse entities is mediated by their descriptions, discourse entity descriptions are critical to anaphor resolution.

Now one consequence of these assumptions about discourse models and reference is that the task of understanding definite anaphora can be decomposed into several complementary parts:

1. deciding whether a definite pronoun or definite description is truly anaphoric (i.e., is intended to refer to some entity presumed to already be in DM_L) or whether the term fills some other role in the discourse;

2. synthesizing a discourse model which is similar to that of the speaker and inhabited by similar discourse entities;

<*1>. The author's current address is: Department of Computer and Information Sciences, The Moore School, University of Pennsylvania, Philadelphia PA 19174.
<*2>. Although I will soon explain what I mean precisely by "definite anaphora", the term basically denotes a **function** that some types of syntactic expressions can serve. Expressions which can function as definite anaphors include definite pronouns and definite descriptions. Other roles that definite pronouns and descriptions can fill are discussed in Geach [1962], Partee [1972], Norman & Rumelhart [1975] and Webber [1978a].
<*3>. A similar assumption is made by Karttunen [1976], Levin & Goldman [1978], Lyons [1978] and Stenning [1975].

3. constraining the possible referents of a given anaphoric expression down to one possible choice - the "anaphor resolution" problem;

4. determining what other functions a definite description is intended to fill besides enabling the listener to construct or get to its referent.

While I cannot hope in this short paper to cover even one of these four sub-tasks, what I shall try to do is illustrate how the explicit data - i.e., the actual sentences of the discourse, produced by a particular person (or a particular computer program) in a particular situation - provide material for the model synthesis process. In particular, I shall show (1) how indefinite noun phrases are associated with the evocation of new discourse entities, independently of any higher-level expectations, and (2) how those new discourse entities will initially be described. I will claim that such an initial description (ID) is critical to both model synthesis and anaphor resolution since it allows the listener to reason appropriately about the discourse entity in order to assign it to an appropriate role vis-a-vis his or her higher-level expectations. <*4> Moreover, since it is possible for a discourse entity's current role assignment to be found incorrect, it is the entity's ID that allows it to be re-assigned to another role with respect to the listener's revised expectations.

In Section 2 I will consider indefinite noun phrases vis-a-vis the discourse entities they evoke and how those entities are described. I will contrast them briefly with non-anaphoric definite noun phrases and then show that all determined noun phrases, including odd ones like "few orc eggs", "many lemon gum balls", etc. pattern after either definites or indefinites vis-a-vis the discourse entities they evoke and how those entities can be described. In Section 3 I will show how this approach to definite anaphora in terms of discourse entities and their descriptions can accommodate certain problematic cases of anaphoric reference that have been discussed in the linguistics and philosophic literatures - the famous "donkey" sentence (cf. Bartsch [1976], Edmundson [1976], Hintikka & Carlson [1977]) and the problem of reference in disjunctive contexts (cf. Karttunen [1977]). Finally, to show that it is not just definite and indefinite noun phrases that can evoke entities in the listener's discourse model, I will illustrate in Section 4 an example of deictically-evoked entities and comment on the problem of describing them appropriately.

<*4>. From different points of view, discussions of the relationship between the explicit text and higher-level organizing structures can be found in Collins, Brown & Larkin [1977] and Webber [1978b].

2. Indefinite Noun Phrases and Discourse Entities

Except after a copula, indefinite noun phrases <*5> may evoke a new discourse entity into a listener's discourse model. <*6> What I want to focus on here is appropriate IDs for them. Consider the following sentences.

1a. Wendy bought a yellow T-shirt that Bruce had liked.
 b. It cost twenty dollars.

2a. Each third-grade girl brought a pelican to Wendy's house.
 b. She is roosting them on her front lawn.

3a. If Bruce manages to catch a fish,
 b. he will eat it for dinner.

4a. John didn't marry a Swedish woman.
 b. She was Norwegian.

5a. Whether Bruce buys a mini-computer or an Advent TV,
 b. he will have to do the repairs on it himself.

6. Every man who owns a donkey beats it.

I claimed earlier that the initial description (ID) of a newly-evoked discourse entity is critical for both model synthesis and anaphor resolution, since the ID mediates all reasoning about the entity until its assignment to some role within the model. An entity's ID should imply neither more nor less about it than is appropriate. Now consider what an appropriate description would be for the discourse entity that "it" refers to in sentence 1b. It is not "the yellow T-shirt that Bruce had liked", since sentence 1a. can be uttered truthfully even if Bruce had liked several yellow T-shirts (and both speaker and listener were aware of that fact). Nor is it "the yellow T-shirt that Bruce had liked and that Wendy bought", since sentence 1a. can

<*5>. I will often refer to these as "existentials" because of their logical interpretation as existential quantifiers.
<*6>. An indefinite noun phrase following a copula functions together with the copula as a predicate, e.g.

Beverly is a bargain hunter.
Bruce became a librarian.

As such, it is purely descriptive and does not refer to any particular librarian or bargain hunter, cf. Kuno [1970].

be truthfully uttered even if Wendy had bought several such T-shirts. What is an appropriate description for the referent of "it" is something like "the yellow T-shirt that Bruce had liked and that Wendy bought and that was mentioned in sentence 1a."

What I am claiming is that in the case of a singular existential that is not within the scope of either negation, a universal quantifier, a hypothetical (e.g. "if", "suppose") or one of several other special contexts (cf. Webber [1978a]), the entity it evokes will be appropriately described via a conjunction of (1) the description inherent in the noun phrase (e.g. "yellow T-shirt that Bruce had admired"); (2) a predicate that embodies the remainder of the sentence (e.g. "which Wendy bought"); and (3) a predicate that relates that entity to the utterance evoking it (e.g. "which was mentioned in (or evoked by) sentence 6a."). This is the description that I am calling the entity's "initial description" or ID. Given how I specified its components then, it should not be surprising that I will claim that the ID of an existentially-evoked discourse entity can be derived from an appropriately structured sentence-level logical representation. Such a representation is independently motivated by its use in regular inference procedures.

Using a somewhat simplified version of the formalism described in Webber [1978a], a simple rule can be stated for forming the ID of an existentially evoked discourse entity - i.e.,

$$(Ex{:}C)\ .\ F_x \ ==>\ (Ez)\ .\ z = ix{:}\ Cx\ \&\ F_x\ \&\ \mathbf{evoke}\ S,x$$

Here (Ex:C) is an example of **restricted quantification**, in which C represents an arbitrary predicate which x satisfies. F_x represents an arbitrary open sentence in which x is free; i stands for Russell's definite operator, iota; and S is the label assigned to the proposition on the left-hand side of the arrow. Informally, this rule, which I shall call [RW-1], says that if a proposition S states that there is a member x of class C which makes F_x true, then there exists an individual describable as "the C which F's which was evoked by proposition S". This individual is taken to be the discourse entity evoked by the existential noun phrase. For example, let Y stand for the predicate corresponding to "yellow T-shirt that Bruce had liked". <*7> Then sentence 1a. can be represented simply as

$$(Ex{:}Y)\ .\ \mathbf{Bought}\ \mathbf{Wendy},x$$

Since this matches the left-hand side of the above rule, it follows that

$$(Ez)\ .\ z = ix{:}\ Y\ x\ \&\ \mathbf{Bought}\ \mathbf{Wendy},x\ \&\ \mathbf{evoke}\ S_{1a},x$$

That is, there is an individual describable as "the yellow T-shirt that Bruce had liked, that Wendy bought and that was evoked by sentence 1a." The discourse entity so described is the referent of "it" in sentence 1b.

Examples 2-6 illustrate singular indefinite noun phrases in some of the special contexts noted above. While I will only be discussing examples 5 and 6 in this paper, notice that in all five cases, the entity evoked by the indefinite noun phrase is appropriately described by taking into account at least the three factors mentioned above. That is, in example 2 the referent of "them" can be described uniquely as "the set of pelicans, each of which, mentioned in sentence 2a., some third grade girl brought to Wendy's house." <*8> In example 3, the referent of "it" can be described as "the fish mentioned in clause 3a. that Bruce has managed to catch, if Bruce has managed to catch a fish". In example 4, the negation appears intended to scope only "Swedish". Thus the discourse entity referent of "she" can be described as "the woman mentioned in sentence 4a. that John married". (We later learn in sentence 4b. that she is Norwegian rather than Swedish.) IDs for the two other existentially-evoked discourse entities in examples 5 and 6 will be discussed in Section 3.

Notice that a definite noun phrase in the same context as an indefinite noun phrase will also evoke a discourse entity, but one whose ID is somewhat different. To see this, consider the following sentences

---

<*7>. I will soon be more precise about the representation of relative clause containing noun phrases. Here, where the descriptive part of the noun phrase can be treated as an unanalyzed unit, the predicate name Y is an adequate representation.

<*8>. A rule similar to [RW-1] is given in Webber [1978a] for existentials scoped by universals. In all, six such rules are given covering

1. independent existentials (sg/pl)
   "I saw {**a cat, three cats**} on the stoop."
2. definite descriptions (sg/pl)
   "I saw **the** {**cat, cats**} **which hate Sam.**"
3. distributives
   "**Each cat on the stoop** hates Sam."
   "**The three cats each** scratched Sam."
4. universally quantified existentials
   "Each boy gave each girl {**a peach, three peaches**}."
5. class dependent definites
   "Each boy gave a woman he knew **the** {**peach, two peaches**} **she wanted.**"
6. class dependent distributives
   "Each boy I know loves **every woman he meets.**"

7a. Wendy bought the yellow T-shirt that Bruce had liked.
b. It cost twenty dollars.

8a. Each third grade girl has seen the pelican on Wendy's lawn.
b. They prefer it to the plastic flamingo she had there before.

9a. John didn't marry the Swedish woman.
b. He threw her over for a Welsh ecdysiast.

In each case, an appropriate description for the discourse entity evoked by the singular definite noun phrase is just that singular definite noun phrase itself "the yellow T-shirt that Bruce had liked", "the pelican on Wendy's lawn", "the Swedish woman". While it is certainly true that the definiteness of these noun phrases may be contingent on context (i.e., identifiability within the speaker's model of the underlying situation) nevertheless unlike entities evoked by indefinite noun phrases, those evoked by definites do not depend for their appropriate IDs on the particular sentences the definite noun phrases appeared in.

The same characteristic behavior of definites and indefinites discussed for singular noun phrases holds for plural noun phrases as well. That is, while both indefinite and definite plural noun phrases evoke discourse entities, the unique initial descriptions that can be assigned to those entities will differ in the two cases. To see this, consider the following example.

10a. I saw the guys from "Kiss" on TV today.
b. I saw the three guys from "Kiss" on TV today.
c. I saw all three guys from "Kiss" on TV today.
d. I saw some guys from "Kiss" on TV today.
e. I saw three guys from "Kiss" on TV today.

11. They were being interviewed by Dick Cavett.

Sentences 10a-c each contains a definite plural noun phrase. That noun phrase should evoke a discourse entity into the listener's model, one appropriately described as "the (set of) guys from 'Kiss'". This can be verified by following either of these sentences by sentence 11 and considering what is the referent of the definite pronoun "they". <*9>

<*9>. While sentences 10b&c. provide the additional information that the number of guys in "Kiss" is three [not actually true - BLW], that information is not needed in order to describe the set uniquely.

Sentences 10d&e, on the other hand, each contains an **indefinite** plural noun phrase. That noun phrase will evoke a discourse entity appropriately described as "the (set of) guys from 'Kiss' that I saw on TV today and that was mentioned in Sentence 10d(e)." This is because either sentence is consistent with there being other members of "Kiss" whom I didn't see on TV today, as well as other members whom I did see but whom I don't mean to include in my statement. <*10> Notice again that the set size information provided in sentence 10e. is not necessary for describing that set uniquely. However, it too may be useful later in resolving definite anaphora.

An interesting point is that there seem to be no other patterns that English determiners follow vis-a-vis discourse entity IDs. To see this consider the following sentences.

12a. Few linguists smoke since **they** know it causes cancer.
b. Few linguists were at the party, but **they** drank more than the whole Army Corps of Engineers.

13a. Many linguists smoke although **they** know it causes cancer.
b. Not many linguists smoke since **they** know it causes cancer.
c. Many linguists don't smoke since they know it causes cancer.

In sentence 12a, the referent of "they" is the discourse entity appropriately described as "(the entire set of) linguists". That is, "few <x>s" can evoke the same discourse entity as the definite noun phrase "the <x>s". However as

However, it should not be ignored, as it may be needed later in resolving a definite anaphor like "the three guys". <*10>. This latter point is a subtle one, and usage may vary from person to person. That is, some people intend an indefinite plural noun phrase contained in a sentence S - "Some <x>s P" - to refer to the **maximal** set - i.e., "the set of <x>s which P". Other people intend it to refer to some subset of that set - "the set of <x>s which P which I (the speaker) intended to mention in sentence S". For a system to cope with this variation in usage, it would be better for procedures to derive the latter, non-maximal set description, which is always appropriate. If a system is sophisticated enough to associate a "belief space" with the speaker (cf. Cohen [1978]), other procedures can later access that belief space (if necessary or desirable) to judge whether the maximal set interpretation might have been intended. (This will again become an issue when I discuss other determiners like "many" and "several".)

sentence 12b. shows, "few <x>s" can also pattern after the indefinite plural: the referent of "they" is the entity appropriately described as "the just-mentioned set of linguists who were at the party". (We learn from "few" that this set is small or smaller than the speaker expects.)

"Many", on the other hand, seems to pattern only after the indefinite plural. In sentence 13a., the referent of "they" is appropriately described as "the just-mentioned set of linguists who smoke" (We learn from "many" that this set of linguists is large or larger than the speaker expects.) Sentence 13b. shows that the reverse polarity "not many" acts like "few" vis-a-vis evoking discourse entities: the referent of "they" is the entire set of linguists. However as sentence 13c. shows, a NEG which occurs in the sentence auxiliary does not effect this same change in behavior: "they" refers to the just-mentioned set of linguists who don't smoke.

3. Two Interesting Reference Problems

Recall that the purpose of this paper is to point out the importance of description formation to both discourse model synthesis and reference resolution **and** to show that this process can, to an important degree, be formalized. I have taken as given the notion that a listener is using both the discourse and his or her knowledge of the world to synthesize a model of what s/he believes to underlie the discourse. Definite anaphora are viewed as means by which the speaker refers to entities in DM_S that are presumed to have counterparts in the listener's model. What I want to show in this section is that this approach to definite anaphora can accommodate not only straight-forward cases as discussed above, but certain problematic cases as well.

3.1 Parameterized Individuals

The problem of formally characterizing the referent of "it" in examples like 6 below has often been discussed in the linguistics and philosophy literatures cf. Bartsch [1976], Edmundson [1976], Hintikka & Carlson [1977].

6. Every man who owns a donkey beats it.

The problem has been taken to be that while "it" intuitively seems related to the embedded noun phrase "a donkey", there is no way to represent this logically in terms of simple quantifier scoping. What I shall show is that an approach in terms of discourse entities and their IDs makes this intuitive relationship simple both to explain and to represent.

First notice that this problem arises independently of how the matrix* noun phrase is determined.

14. A man I know who owns a donkey beats it.
15. The man who owns a donkey beats it.
16. Which man who owns a donkey beats it?
17. No man who owns a donkey beats it.

In all these examples, "it" seems intuitively related to "a donkey". Informally, one might describe its referent as "the just-mentioned donkey he owns", where "he" is bound to whatever value that "(each, a, the, which, no) man who owns a donkey" may take. But this is just a discourse entity of a rather special type - one with a parameterized ID, rather than a rigid one. I call such entities "parameterized individuals", borrowing the term from Woods & Brachman [1978]. <*11>

Notice that parameterized individuals behave somewhat differently from the "actual" discourse entities the sentences evoke. <*12> That is, parameterized individuals all have the **same ID**, independent of how the noun phrase containing the relative clause is determined. On the other hand, the actual discourse entities evoked by these sentences do not. For example,

18a. Each man who owns a donkey beats it.
it = the donkey he owns
b. However, **the donkeys** are planning to get back at **them**.
the donkeys = the set of donkeys, each of which some man who owns a donkey owns
them = the set of men, each of whom owns a donkey

19a. The man I know who owns a donkey beats it.
it = the donkey he owns
b. But **the donkey** is planning to get back at **him**.
the donkey = the just-mentioned donkey that the man I know who owns a donkey owns
him = the man I know who owns a donkey

20a. Which man who owns a donkey beats it?

<*11>. The phrase "parameterized individual" is being used somewhat loosely to include "parameterized" sets, stuff, etc. For example,
(i) No man who owns two donkeys beats them.
them = the two donkeys he owns

<*12>. By "actual" discourse entities, I mean ones that can be referred to anaphorically in subsequent sentences.

it = the donkey he owns
-- "None"

b.*Are **the donkeys** planning to get back at {**him**, **them**, ???}?
the donkeys = ???

c.*Is **the donkey** planning to get back at {**him**, **them**, ???}?
the donkey = ???

To show that this approach to definite anaphora in terms of discourse entities and their descriptions can explicate "donkey" sentences as well, I will have to introduce a bit more of the formalism described in Webber [1978]. That bit involves an extension of **restricted quantification**, cf. [RW-1] above. In restricted quantification, a quantification operator (e.g. ∀,E), the variable of quantification and the class it ranges over (noted implicitly as a predicate) constitute a structural unit of the representation. For example, "Every boy is happy" can be represented as

(∀x:Boy) . Happy x

This is truth functionally equivalent to

(∀x) . Boy x ==> Happy x

Similarly "Some boy is happy" can be represented as

(Ex:Boy) . Happy x

which is truth functionally equivalent to

(Ex) . Boy x & Happy x

The extension I will introduce will permit the representation of noun phrases with relative clauses as well as simple noun phrases. Semantically, a relative clause can be viewed as a predicate. One way to provide for arbitrary predicates is through the use of the abstraction operator, represented as " " by Hughes & Cresswell [1968], following Church [1941]. For example, the noun phrase "a peanut" can be represented as

(Ex:Peanut)

while the noun phrase "a peanut that Wendy gave to a gorilla" can be represented as

(Ex: λ(u:Peanut)[(Ey:Gorilla)
Gave Wendy,u,y])

In this case

λ(u:Peanut)[(Ey:Gorilla)
Gave Wendy,u,y]

names a unary predicate which is true if its argument is a peanut that Wendy gave to some gorilla.

Using this notation, sentence 6 can be represented as

(∀x:λ(u:Man)[(Ey:Donkey) . Own u,y])
Beat x,IT

By applying rule [RW-1] to the embedded clause [(Ey:Donkey) . Own u], the entity evoked by the existential can be identified as

iy: Donkey y & Own u,y & **evoke** $S_{6.1}$,u
"the just-mentioned donkey that u owns"
<*13>

As I mentioned above, the semantics of restricted quantification is such that the variable of quantification, here x, satisfies the predicate in the restriction. Thus if * satisfies λ(u:Man)[(Ey:Donkey) . Own u,y], there must be an entity identifiable as

iy: Donkey y & Own x,y & **evoke** $S_{6.1}$,y
"the just-mentioned donkey x owns"

This is a parameterized individual - parameterized by the variable in (∀x:...) - that is a possible referent for "it" in the matrix sentence - i.e.,

(∀x: λ(u:Man)[(Ey:Donkey) . Own u,y])
Beat x, iy: Donkey y & Own x,y
& **evoke** $S_{6.1}$,y
"Every man who owns a donkey beats the just-mentioned donkey-he owns"

I noted above that a sentence like "Every man who owns a donkey beats it" could sensibly be followed by a sentence like "However, **the donkeys** are planning to get back at them" (cf. example 18). Given that I have shown how to account for the referent of "it" in the first sentence in terms of discourse entities and their formally derivable descriptions, can the referent of "the donkeys" be account for in the same way? <*14>

To show that it can, I need to present the rule for dealing with class dependent definite descriptions that I mentioned in footnote 8. This rule is motivated by examples such as 21, where the referent of "them" is presumably the discourse entity evoked by the noun phrase "the flower she picked", where "she" stands for the variable bound by "each girl in the class".

<*13>. In labeling each clause of a complex sentence, I use the following convention: if the matrix clause is labelled S, its leftmost embedded clause will be labelled S.1, the leftmost embedded clause in S.1 will be labelled S.1.1, etc.

<*14>. I shall not take the time here to discuss the path from the phrase "every man who owns a donkey" to the discourse entity informally describable as "the set of men, each of whom owns a donkey", since it is rather straightforward, cf. Webber [1978a]. This entity is a possible referent for "them" in sentence 18b.

21a. Each girl in the class gave Ivan the flower she picked.
b. He arranged **them** artfully in an empty Glenfiddach bottle.

This is a definite noun phrase, but because of its binding to the distributively quantified noun phrase "each girl", it will evoke a discourse entity with the properties of a set rather than an individual (cf. example 8). In this case, it will be "the set of flowers, each of which was the flower that some girl in the class picked". Simplifying for brevity here, this rule can be written

$$(\forall x{:}K)\ .\ P\ x,iy{:}C\ x,y \Longrightarrow (Ez)\ .\ z = \{u \mid (Ex{:}K)\ .\ u = iy{:}C\ x,y\}$$

where K represents an arbitrary unary predicate which x satisfies and both P and C represent arbitrary binary predicates. The right-hand side of this rule implies that in case the left-hand side matches some sentence, there will be a discourse entity roughly describable as "the set of u's, each of which is the thing that stands in relation C to some member of K".

Notice now that after the "it" is resolved in "Every man who owns a donkey beats it" (see above), the sentence matches the left-hand side of the above rule - i.e., "Every man who owns a donkey beats **the just-mentioned donkey he owns.** Thus it follows that there is a discourse entity describable as "the set of donkeys, each of which is the just-mentioned donkey that some man who owns a donkey owns" - i.e.,

$$\{w \mid (Ex{:}\lambda(u{:}Man)[(Ey{:}Donkey)\ .\ Own\ u,y])\ w = iz{:}\ Donkey\ z\ \&\ Own\ x,z\ \&\ evoke\ S_{18},z\}$$

This is a possible referent for "them" in sentence 18b.

3.2 Disjunction

The other class of problematic examples that I want to discuss here in terms of discourse entities and their descriptions is one I first encountered in Karttunen [1977]. Karttunen presents examples like the following.

22. If Wendy has a car or Bruce has a bike, it will be in the garage.

23. Bruce can have either a bike or a car, but he must keep it in the garage.

24. Either Bruce has a new car or he has borrowed his brother's. In any case, it is blocking my driveway.

25. Whether Bruce buys a car or his brother buys a bike, he will have to keep it in the garage.

The problem is again to determine just what it is that "it" refers to.

I see two ways of approaching this problem in terms of discourse entities and their IDs. One way holds that in each sentence, each term of the disjunction evokes a different discourse entity into DM_L, each with a different ID:

(22) "the car that Wendy has (if she has a car)"
"the bike that Bruce has (if he has a bike)"

(23) "the bike that Bruce will have (if he chooses a bike)"
"the car that Bruce will have (if he chooses a car)"

(24) "the new car that Bruce has (if Bruce has a new car)"
"Bruce's brother's car"

(25) "the car Bruce will have bought (if he buys a car)"
"the bike Bruce's brother will have bought (if Bruce's brother buys a bike)"

The truth of the disjunction (which seems in each case to be interpreted as exclusive "or") then guarantees there being one and only one entity in the model to which "it" refers. Notice that if the terms were conjoined rather than disjoined, the truth of the conjunction would imply the simultaneous existence of two entities within the model. In that case, either the referent of "it" would be ambiguous or the sentence would just be bizarre.

The other, I think nicer, way of approaching the problem holds that each sentence evokes only a **single** discourse entity into the model, with the indecision (i.e., the disjunction) embodied in its **ID.** That **ID** is of the form "A if P, otherwise B". For example, the entity evoked by sentence 22 would be describable as "the car that Wendy has (if she has a car) or the bike that Bruce has otherwise" that evoked by sentence 23 would be describable as "the bike that Bruce will have (if he chooses a bike) or the car that Bruce will have otherwise"; that evoked by sentence 24, as "the new car that Bruce has (if he has a new car) or Bruce's brother's car otherwise"; and that evoked by sentence 25, as "the car Bruce will have bought (if he buys a car) or the bike Bruce's brother will have bought otherwise".

One advantage to this approach is that additional properties which truthfully follow from either **ID** can be ascribed to the entity without committing oneself to one description or the other. This can be useful in anaphor resolution. For example, in sentence 24, the subject

of "block my driveway" must be a physical object, preferably large and somewhat mobile. This condition is satisfied by the discourse entity evoked by sentence 24, independent of which ID is appropriate.

Although there may be other ways to approach the problem of disjunction, the "donkey" problem, and the whole problem of definite reference in general, what I hope to have shown in these two sections is the robustness of an approach based on notions of a discourse model, discourse entities and their formally derived descriptions.

4. Conclusion

In arguing for the importance of description formation to both discourse model synthesis and reference resolution, I concentrated on how indefinite noun phrases evoke new entities into the listener's discourse model and how their appropriate initial descriptions (IDs) could be derived from a formal sentence-level representation of the text. There are many other ways in which discourse entities can be evoked, and many interesting problems in forming appropriate descriptions of them. I will conclude therefore with a brief discussion of deictically-evoked discourse entities and the problem of describing them appropriately.

The example comes from the children's book **Babar Loses his Crown** by Laurent de Brunhoff, and involves the following situation: Babar, King of the Elephants, decides to take his wife Celeste and his family on a trip to Paris. In packing for the trip

> "Babar puts his crown in a little red bag." (p.3)

They travel by train and then by taxi to their hotel in Paris, and when they arrive

> "Celeste opens all the bags. Last of all, she opens the little red one. 'Look!' she cries. 'What is this? A flute! Babar! **This is not your bag!**' " (p.10)

Before this point in the story, there should have been one little red bag in DM_L. Now there should be two. The first is the existentially-evoked discourse entity (say, e_{43}) - "the little red bag mentioned in sentence <x> that Babar put his crown in". However if "this" on page 10 is not that entity, then it must be some other one (say, e_{48}). How should it be described? Since "this" presumably points to the little red bag Celeste is opening, e_{48} can appropriately be described as "the just-mentioned little red bag which Celeste is opening, which contains a flute and not Babar's crown, and which is not equivalent to e_{43}". <*15>

The problem here is to be able to articulate clearly what each of these properties derives from since they do not come from a single sentence. In this case one must determine what things relevant to the story do or do not follow from e_{48}'s not being Babar's bag.

* * * * * * * * *

In this paper, I have tried in as brief a way as possible to reveal an aspect of understanding definite anaphora that precedes the more frequently discussed problem of "anaphor resolution". This aspect involves accounting for what it is that definite anaphors refer to and how such things become available. I moved from the notion of reference into a model to problems of how that model is synthesized, and in particular, how the entities in it are appropriately described. In this endeavor, I focused on the initial descriptions (IDs) ascribed to existentially-evoked entities, briefly touching upon deictically-evoked entities as well. This paper has just skimmed the surface of a very large problem. In particular, one must still account for, **inter alia**, reference to actions, events, processes, stuff, quantities of stuff, etc.; relativization of descriptions to the speaker's beliefs (cf. Cohen [1978], Prince [1978]); additional descriptions derived from the various roles in higher-level situations that an entity is assigned to; effects of tense, modality, negation, etc. on description formation; and how descriptions change over time. Some of these problems (as well as others) are discussed further in Webber [1978a&b], and much interesting work remains to be done.

Bibliography

Bartsch, R. Syntax and Semantics of Relative Clauses. In R. Bartsch, J. Groenendijk & M. Stokhof (Eds.), **Amsterdam Papers on Formal Grammars.** The Netherlands: University of Amsterdam, 1976.

<*15>. Throughout this approach, I am making no assumptions about the separateness of discourse entities. That is, I am not assuming that two discourse entities are necessarily and for all times distinct, unless it is specified explicitly as was done here. Two discourse entities may have incompatible descriptions and as a consequence be **assumed** to be distinct. But I do not view it as impossible for two separately evoked discourse entities with different but compatible descriptions to later be found to be one and the same.

Church, A. **The Calculi of Lambda Conversion.** Princeton NJ: Princeton University Press, 1941.

Cohen, P. On Knowing What to Say: Planning Speech Acts. Technical Report No. 118. Department of Computer Science, University of Toronto, 1978.

Collins, A., Brown, J.S. and Larkin, K. Inference in Text Understanding. (CSR-40) Center for the Study of Reading, U. of Illinois and Bolt Beranek and Newman Inc., 1977. (Also in R. Spiro, B. Bruce and W. Brewer (Eds.), **Theoretical Issues in Reading Comprehension.** New Jersey: Lawrence Erlbaum Associates, 1978.)

Edmondson, J.A. Semantics, Games and Anaphoric Chains. In R. Bartsch, J. Groenendijk & M. Stokhof (Eds.), **Amsterdam Papers on Formal Grammars.** The Netherlands: University of Amsterdam, 1976.

Geach, P. **Reference and Generality.** New York: Cornell University Press, 1962.

Hintikka, J. & Carlson, L. Pronouns of Laziness in Game-theoretical Semantics. **Theoretical Linguistics,** 1977, **4**(1/2), 1-30.

Hughes & Cresswell, M. **An Introduction to Modal Logic.** London: Methuen Publishing Co., 1968.

Karttunen, L. Discourse Referents. In J. McCawley (Ed.), **Syntax and Semantics** (Volume 7). New York: Academic Press, 1976.

Karttunen, L. Whichever Antecedent. Squib Anthology, Chicago Linguistics Society, 1977.

Kuno, S. Some Properties of Non-Referential Noun Phrases. In R. Jakobson and S. Kawamoto (Eds.), **Studies in General and Oriental Linguistics.** Tokyo, Japan: TEC Company Ltd., 1970.

Levin, J. & Goldman, N. Process Models of Reference. Unpublished ms. Marina del Rey CA: Information Sciences Institute, 1978.

Lyons, J. **Semantics.** England: Cambridge University Press, 1977.

Norman, D. & Rumelhart, D. **Explorations in Cognition.** San Francisco CA: W.H. Freeman, 1975.

Partee, B.H. Opacity, Coreference and Pronouns. In G. Harman and D. Davidson (Eds.), **Semantics of Natural Language.** The Netherlands: D. Reidel, 1972.

Prince, E. On the Function of Existential Presupposition in Discourse. In J. Petofi (ed.), **Text vs. Sentence.** Hamburg: Buske Verlag, 1978.

Stenning, K. **Understanding English Articles and Quantifiers.** Unpublished doctoral dissertation. New York: The Rockefeller University, 1975.

Webber, B.L. A Formal Approach to Discourse Anaphora. (Technical Report 3761) Cambridge MA: Bolt Beranek and Newman Inc., 1978a.

Webber, B.L. Discourse Model Synthesis: Preliminaries to Reference. Presented at the Sloan Workshop on "Computational Aspects of Linguistic Structure and Discourse Setting", University of Pennsylvania, 25-28 May 1978b.

Woods, W.A. & Brachman, R.J. Research in Natural Language Understanding - Quarterly Technical Progress Report No. 1. (Technical Report 3742) Cambridge MA: Bolt Beranek and Newman Inc., 1978.

The Processing of Referring Expressions within a Semantic Network

John R. Anderson
Yale University

Frege (1892) is credited with emphasizing the distinction between sense and reference. His famous example involved the morning star and the evening star. Despite the fact that they both refer to the same object (i.e., Venus), they have different senses as witnessed by the fact that sentence (1) is not synonymous with sentence (2):

(1) The morning star is the morning star.
(2) The morning star is the evening star.

This philosophical issue has similarities to an issue that is of importance to understanding natural language processing: How do subjects process referring expressions to extract internal representations (a) of their meaning and (b) of their referents in the external world. The example sentence that we will be returning to in this paper is:

(3) The first president of the United States was a bad husband.

It is clear that in understanding this sentence we both process the subject as a description, and identify this as referring to George Washington. This paper will try to explain how this comes about. As I believe that all interesting questions about representation come down to questions about memory, I will approach this question from a human memory perspective.

Some "self-evident" truths about human memory.

To set up a framework for further discussions, I would like to list some of the facts that I think we know about human memory -- either because of a sophisticated common sense and self-observation or because of a mass of experimental data:

(1) Human memory can be conceived of as a network of associations among concepts.
(2) Some nodes in this network refer to individuals in the external world.
(3) Once information is deposited in memory it cannot be erased.

While there are a number of memory theories that embody these assumptions, I will be using the ACT model (Anderson, 1976) to present the theory and discuss the data in this paper. With this brief statement of the pre-theoretical biases, I would like to turn to an experimental paradigm which captures, in expanded time scale, the processes that I think are going on when we comprehend referring expressions.

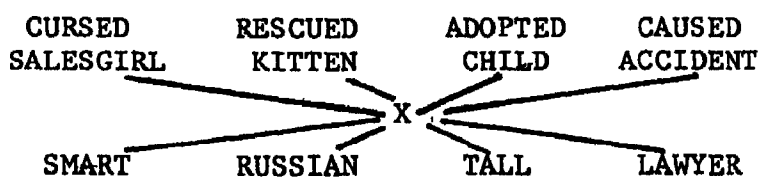
A Mock-up of the Morning Star-Evening Star Example

One of the experiments in this series (see Anderson, 1977; Anderson & Hastie, 1974 for a thorough report) had subjects study a set of facts such as (4) - (8):

(4) The smart Russian is the tall lawyer.
(5) The smart Russian cursed the salesgirl.
(6) The smart Russian rescued the kitten.
(7) The tall lawyer adopted the child.
(8) The tall lawyer caused the accident.

The critical manipulation was whether the identification sentence (4) was learned some time before or some time after sentences (5)-(8). For the identification before condition, part (a) of Figure 1 illustrates, very schematically, the network structure we thought was created. There is a node X set up to represent the individual and attached to that node are the various facts learned about this person. Part (b) of Figure 1 illustrates the network situation in the identification after condition. Because the subject did not learn of the identity between the two individuals until after learning sentences (5)-(8), he was led to create two nodes in memory which turn out to refer to the same individual. It would seem optimal if he could merge nodes X and Y together but this would amount to erasing memory structures, violating principle 3. Rather we assume that the subject encodes a separate proposition to the effect that the two individuals are identical. This is represented in Figure 1b, by the link between X and Y labelled with an '='

(a) Identification Before



(b) Identification After

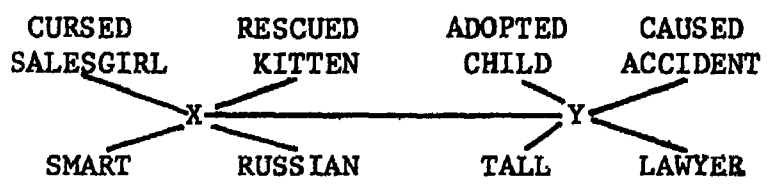


Figure 1. Memory representations at the beginning of the reaction time verification phase.

The memory representations in Parts (a) vs. (b) make different predictions about time to verify statements (9) vs. (10):

(9) The smart Russian cursed the salesgirl.

(10) The smart Russian caused the accident.

Statement (9) is referred to as a direct statement because it is identical to a study statement, while statement (10) is referred to as an inference as it can be inferred from statements (4) and (8).

Table 1 displays subjects' reaction times to verify direct statements and inferences in the identification before and identification after condition. We would expect subjects to show very little advantage for direct statement over inference in a representation like Figure 1a since there is no special connection preserved between the predicates and the referring expressions they

Table 1

Reaction Times (in msec) to verify Statements like 9 and 10

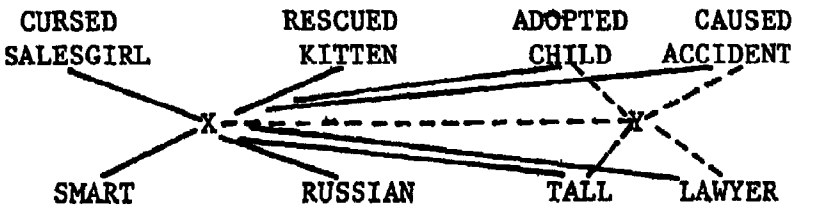
	Identification Provided	
	Before	After
Direct Statement	2310	1978
Inference	2388	2634

were studied with. In fact the verification times are almost identical in the two conditions. In contrast the after condition in Part b of Figure 1 each referring expression is only directly connected to the predicates it was studied with. To verify an inference requires an extra step of activating the path encoding the equality of X and Y. Correspondingly, we find an advantage for direct statements over inference. Finally, note that there are many more links attached to node X in part (a) than to either X or Y in part (b). This means there are more irrelevant paths that can interfere with finding the desired connection. Correspondingly, we find subjects faster to direct statement trues in the after condition.

The data reported in Table 1 come from the first block of reaction time test trials. There were four such blocks of trials. The reaction time data for all four blocks are displayed in Figure 2. Besides illustrating a general speed-up over the course of the experiment, the figure illustrates the differences among the conditions gradually disappear over the course of the experiment. Specifically, the differences between inference and direct statements in the after condition disappears and the differences between identification before and identification after condi- disappear.

To account for this across-block trend we propose that the subject begins a process of copying the predicates from one of the nodes in Figure 1b to the other node. That is, one node is chosen to be abandoned and the other to receive all information. Therefore, supposing the subject choses to copy from node Y to X, everytime he encounters a fact attached to X he will attempt to copy it to Y. Figure 3 illustrates our belief about the memory representation by the end of the experiment. Note that the node X has been attach-

Figure 3: Memory representation in the identification after condition after much practice at verfying inference questions.



ed to all the facts learned of Y. Also the connections involving Y are dotted to indicate that they have become weak through disuse. The after

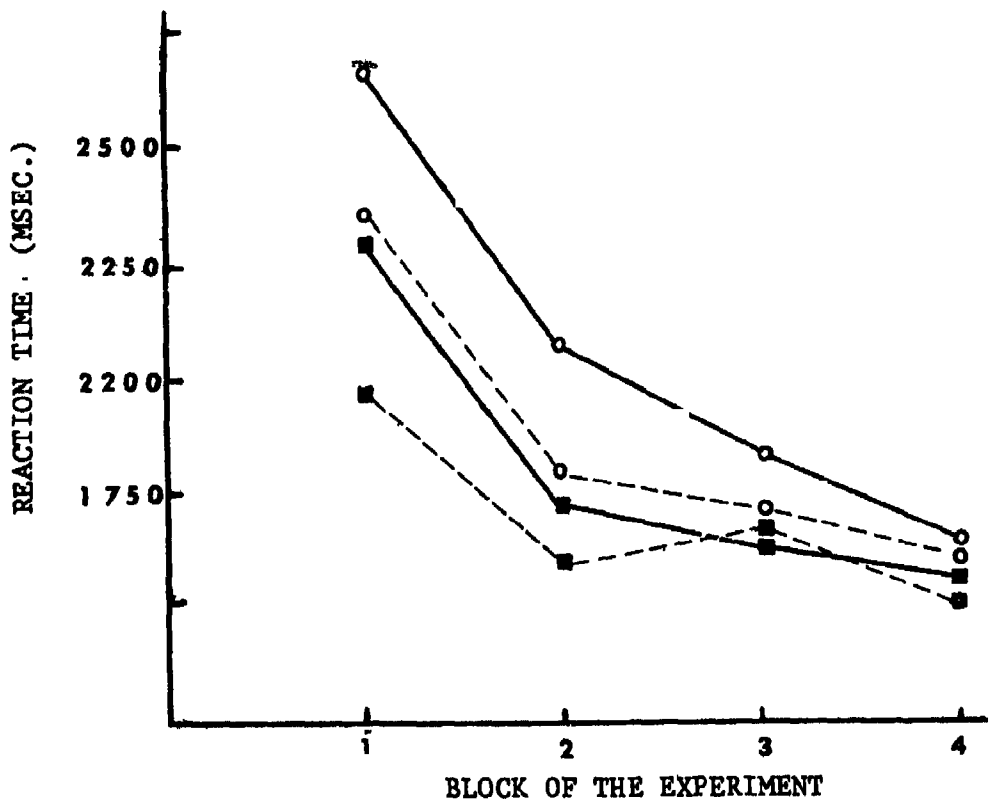


Figure 2: Verification times for various kinds of probes as a function of practice.

representation in Figure 3 has become functionally almost equivalent to the before representation in Figure 1a. Thus there is little difference between inference and direct statement or between the after and before condition.

One might wonder why the subject did not perform this copying when he learned about the identity between the two referring expressions rather than later in the verification phase of the experiment. In the ACT memory model such copying operations cannot be performed unless the data to be copied is active in working memory. At the time of studying the identification statement (4) the predicates needed for copying would not be active in memory. It is only when inferential statements like (10) are encountered in the test that the copying can take place. The referring expression could be copied while learning the identification statement. So the expression tall lawyer might be immediately attached to X. Thus, Figure 1b might be an oversimplification of the state of memory in the identification after condition. But in any case, the inference effect will not go away until the predicates are copied and this will not occur until the reaction time test phase.

Why should we believe this copying explanation rather than any of the multitude of alternative mechanisms that might be offered the explain the data in Figure 2. First, it satisfies the constraint that the subject not be able to erase information from memory and many of the mechanisms would not be. Second, unlike many of the other mechanisms, it assumes an asymetry in the fate of the two individual nodes in Figure 1b. One node is fated to receive all the information and the other node is to be abandoned. It seems reasonable that a subject would choose to preserve that node which had the more information attached and/or had this information attached more strongly. We have been able to demonstrate that subjects do abandon the "weaker" node.

The evidence for this asymetry comes from experiments that use a proper name rather than one of the definite descriptions. That is, the material is the same as in the example except that wherever tall lawyer appears a proper name like James Bartlett would be used. There is evidence (Anderson, 1977) that subjects learn material less well involving the proper name than the definite description. Correspondingly, we would expect subjects to choose to abandon the proper name node and maintain the definite description node. Evidence for this comes from the following analysis: We would propose that, in the initial drilling on the sentence James Bartlett is the Russian, in the identification after condition subjects copy the James Bartlett name to the Russian node. Figure 4 illustrates the memory representation with this asymetry. Note that,

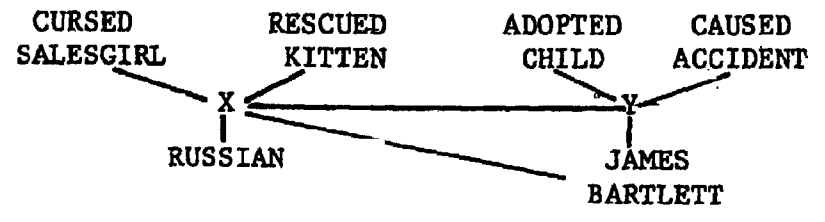


Fig. 4 : Verification times in the identification before and identification after conditions in an experiment that used both proper names and definite descriptions as referring expressions.

according to this representation, subjects should be as fast when verifying an inference predicate of James Bartlett as a direct statement predicate. This is because the proper name is directly attached to both. In contrast, subjects should be much slower for an inference predicate to a definite description because those predicates have not yet directly been attached to node X to which the description is attached. To verify these questions involves the extra retrieval of the proposition that node X equals node Y. Figure 5 presents the data from one of the experiments (Anderson & Hastie, 1974) contrasting definite descriptions and proper names. As predicted there is a large inference effect only for definite descriptions in the after condition.

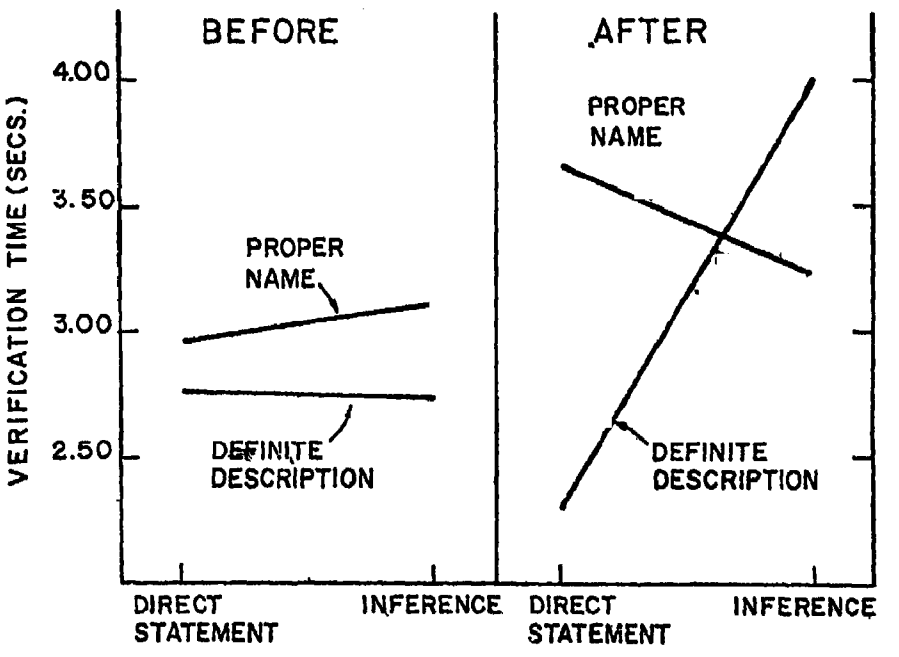


Figure 5.

Application to Recognition of Referring Expression

The advantage of the paradigm just reviewed is that the sequence of states of memory is sufficiently spread out over time that it is possible to map out the changes in memory. I will be proposing that there is a similar sequence of memory states when subjects process referring expressions as in (3):

(3) The first president of the United States was a bad husband.

However, the processing happens so rapidly it is not as easy to verify each state in the sequence.

Figure 6 illustrates two possible sequences of information processing. Part (a) illustrates the state of memory right after comprehension of

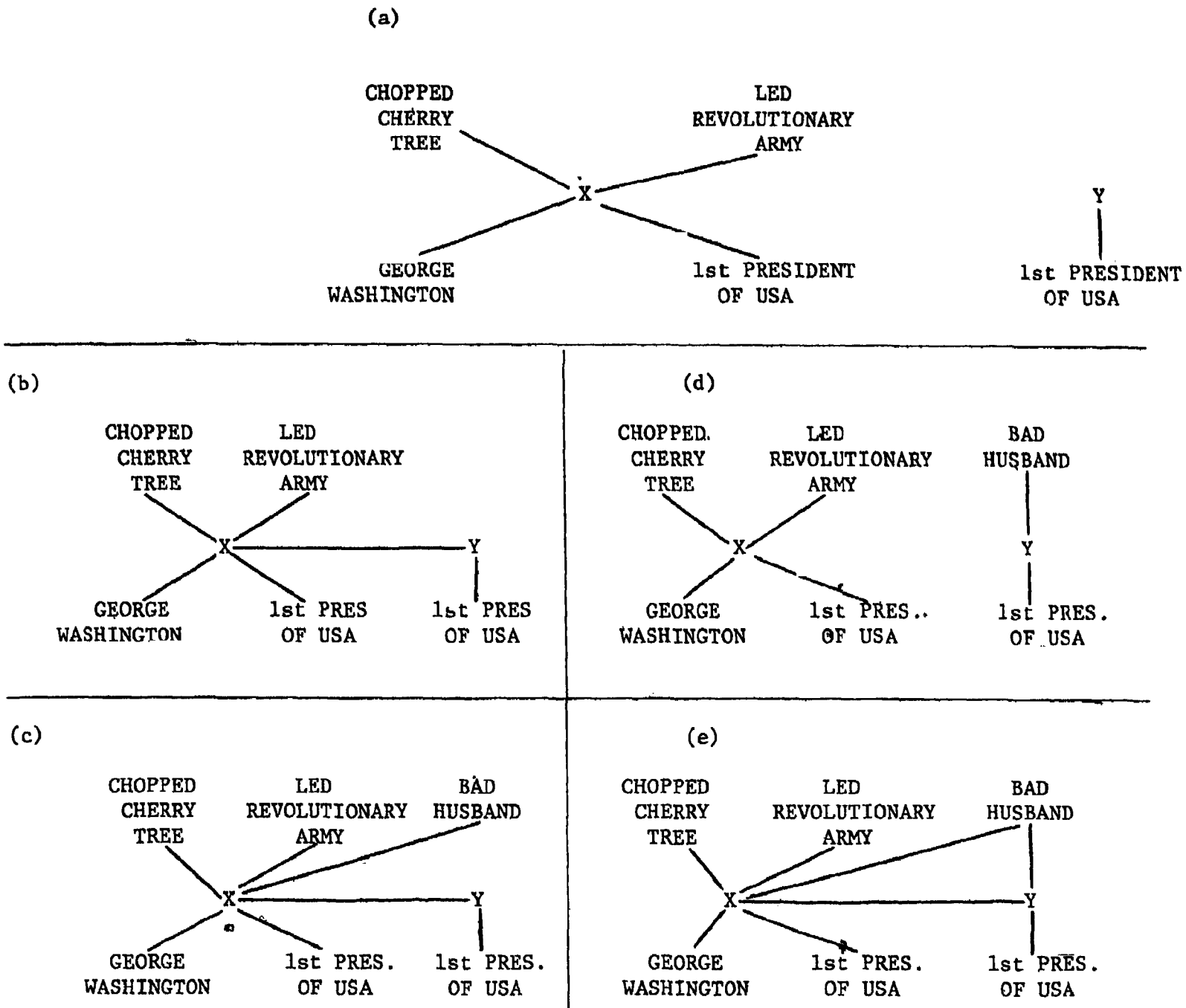


Figure 6. Possible states of memory representation during the processing of sentence (3).

the definite description. A node Y has been created to which there has been attached the "first president of USA" description. A separate node, X, in memory encodes permanent information about George Washington. Part (a) of Figure 6 illustrates a situation analagous to the identification after condition, prior to the identification statement. There are two distinct nodes, unconnected, that refer to the same individual. Introspectively, it seems clear that at least sometimes I comprehend definite descriptions before recognizing their referents. For instance, I understand the description The president of France in 1970 long before I decide that this is George Pompidou. The structure surrounding Y in Part (a) not only provides an embodiment of this pre-identification comprehension, it serves as an encoding of the information that is to guide the search for a referent. The ACT theory would use this representation to build a pattern that could be matched to memory to retrieve the referent. In the case of a description like first president of the USA a direct pattern match should suffice to retrieve the referent. In my case for the President of France in 1970 description, a more complex problem solving strategy had to be evolked.

Once the description of sentence (3) has been comprehended two things can happen: The subject can proceed to recognize the referent of the definite description and he can go on to comprehend the "was bad husband" predicate. Depending on the order of these two events we will wind up with slightly different representations in memory. Part (b) of Figure 5 illustrates the state of memory after recognition of the description and before comprehension of the predicate. As in the after condition (Part b of Figure 1) a link is introduced encoding the identity of X and Y. When the predicate is comprehended a representation of its meaning can be attached directly to X, giving the representation in Part (c) of Figure 6.

Part (d) illustrates the state of memory when the predicate has been comprehended but the definite description has not been identified. In this case the meaning representation of the predicate has been attached to node Y. Part (e) of Figure 6 illustrates the state of memory when the definite description is subsequently recognized. Again a link is introduced indicating the identity between X and Y. The bad husband predicate, which is active in memory, is copied from Y to X. The difference between the final state of the recognize-description-then-comprehend-predicate sequence (Part c) and the comprehend-predicate-then-

recognize-description sequence (part e) is that in the latter case the predicate is attached to both nodes. This latter situation is like the situation in the after condition of the previous memory experiments.

What determines which occurs first -- recognition of description or comprehension of predicate? In the ACT model both processes can go on independently. It would simply be a race between two independent processes. Factors such as how quickly the predicate is presented (if spoken) or how quickly the subject turns to the predicate (if printed) will determine the speed of the comprehension success. The speed of recognizing the description will vary with the difficulty of finding its referent. It is clear that neither process waits on the other as witnessed by the sentences:

(11) The first prime minister of Canada was a bad husband.

(12) The first president of the United States pilacked gibs.

In (11) we comprehend the predicate although we never find a referent for the subject. In (12) we find a referent for the subject although we never comprehend the predicate.

Evidence on the Recognition of Referring Expressions.

Right now the contentious reader might be thinking "Yes, that is a possible model for the processing of referring expressions. Yes, it is consistent with the model for your earlier memory experiments. Yes, you presented evidence for that model. But, is there any independent experimental evidenoe for this model when applied to the real-time recognition of definite descriptions?" Because of its rapid real-time characteristics it is hard to provide particularly direct evidence for this process. But there are some consistent experimental results:

A relevant feature to note about Figure 6c is that it preserves no record that the bad husband predicate was asserted via the first president of USA description. In contrast Figure 6e does preserve this information. Both representations are possible depending on the exact timing of description recognition versus predicate comprehension. To the extent that there is a mixture of these representations we predict both a tendency to make confusions about what referring expression was used (representation 6c) and that subjects will have some residual ability to make this discrimination (representation 6e). An experiment reported by Anderson and Bower (1973) supports this dual prediction. They had subjects study sentences like:

(11) The first president of the United States was a bad husband.

(12) Abraham Lincoln was a good husband.

After studying such sentences subjects were asked to chose among alternatives such as the following:

(13) The first president of the United States was a bad husband.

(14) George Washington was a bad husband.

(15) The first president of the United States was a good husband.

(16) George Washington was a good husband.

These alternatives were presented to the subject randomly ordered but I present them here systematically. Subjects were instructed to indicate the exact sentence that they had studied in which case (13) would be the correct choice. To the extent that subjects false alarm more to (14) over (15) or (16), this is evidence for a representation like Figure 6c where no information is retained about the referring expression used. To the extent that subjects prefer (13) over (14) this is evidence for a representation like Figure 6e. Thus, our predictions in terms of preference is (13) > (14) > (15) = (16). The evidence clearly confirms this prediction with subjects saying that they had seen sentences like (13) 65.2% of the time, like (14) 21.4% of the time, like (15) 7.2% of the time, and like (16) 6.3% of the time. An earlier memory model, HAM (Anderson & Bower, 1973) predicted total confusion in this situation rather than an intermediate level of confusion. In the recognition model for HAM there was no separate memory structure to encode the referring expression. Rather the referent node was directly retrieved from memory without the intermediate step of calculating a representation of the referring expression in memory.

Recently Ortony and Anderson (1977) report a study which replicated and extended this result. They noted that some predicates seemed more appropriate to a proper name and other predicates seemed more appropriate to a definite description. Consider their examples:

(17) The first man on the moon became a national hero.

(18) Neil Armstrong has several children.

(19) The first man on the moon has several children.

(20) Neil Armstrong became a national hero.

Ortony and Anderson point out that the uses in (17) and (18) are somewhat more natural than the uses in (19) and (20). Correspondingly, they found subjects made fewer errors in remembering what the referring expression had been for sentences like (17) and (18) than for sentences like (19) and (20). The error rates were 19.6% versus 30.7%. Note, however, that in both cases subjects identified the original referring expression better than chance (50%).

The Ortony and Anderson result would be expected under the current theory. To the extent that the predicate fits the referring expression subjects might attach it to the new node (e.g., node Y in Figure 5) which has the referring expression attached to it. As Ortony and Anderson noted, the HAM theory had no way to explain this affinity between certain referring expressions and certain predicates. To explain the Ortony and Anderson results in the HAM framework we had to attribute them to a response bias.

In the current ACT theory we can explain this result in terms of the frequency with which subjects chose Part (c) versus Part (e) of Figure 6. The claim is that subjects use representations like Part (e) more frequently when the referring expression is appropriate. This is because it is easier to elaborate on the connection between the referring expression and the predicate.

Opaque and Transparent References

This analysis of reference has a natural extension to analyzing the difference between opaque and transparent reference. For instance, contrast:

(21) I am looking for the best lawyer in town.

(22) I am looking for my little old mother.

While both (21) and (22) might be considered ambiguous, the more apparent interpretation of (21) is that I am looking for someone who fits the description "the best lawyer in town" and that I do not have a particular person in mind. In contrast, the more apparent interpretation of (22) is that I do have a particular person in mind. The former is an instance of opaque reference and the latter is an instance of transparent reference. Our discussion has so far focused on transparent reference. To correctly remember an instance of opaque reference it is critical that it not be treated in the same manner as transparent reference. That is, even if the listener knows the reference of "the best lawyer in town", he should not use the node for this reference in representing the meaning of (21). Rather he should create a new node, attach the description to it, and put this node in the representation of (21). Figures 7a and 7b illustrate the different representations appropriate for (21) and (22). In Part (a) there are two dis-

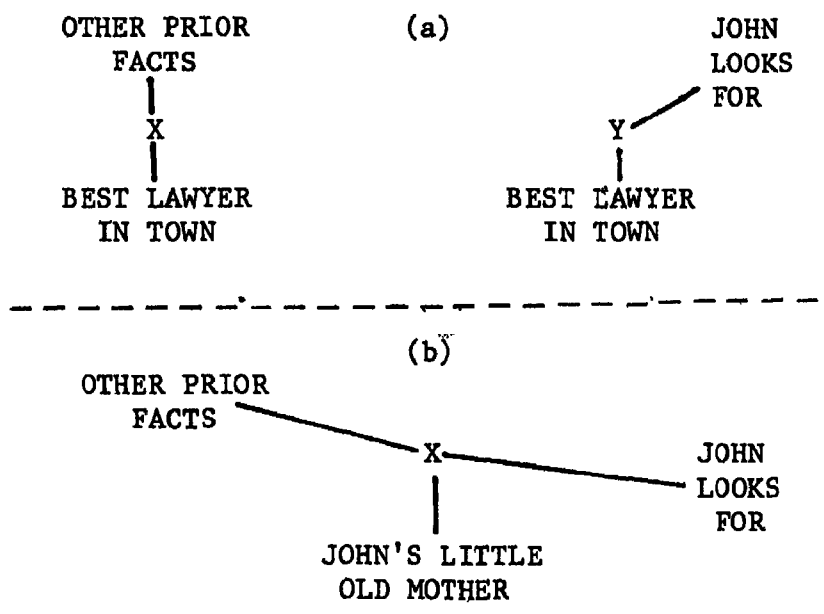


Figure 7: Memory representation for any instance of opaque reference (a) and transparent reference (b).

tinct nodes preserved to represent the best lawyer in town. One node (X) has the prior facts known about the person while the second node (Y) stores information about the opaque reference in sentence (21). There is no such distinction in Part (b) of Figure 7 for the transparent case in (22). All information is attached to the original node X. So, the difference between transparent and opaque reference is whether the new information is copied to an existing node.

Conclusions

In concluding this paper I would like to return to Frege's morning star-evening star example. The discovery that the morning star was the evening star was an important scientific result. Frege used this fact to make clear the important distinction between sense and reference. The first half of this paper reported experiments where we basically recreated Frege's example and discovered that subjects dealt with that dilemma by the process of copying from one referring node to another. The argument in the second half of the paper was that Frege's examples are not isolated to discoveries of science or to bizarre psychological experiments. Rather, every time we recognize a transparent referring expression we go through a discovery like that of the identity between the morning star and evening star. We create a node to represent the referent of the referring expression and only then discover, with varying difficulty, that this node has the same reference as an established node in memory.

Acknowledgments: Preparation of this manuscript and the research reported was supported by Grant BNS 76-00959 from the National Science Foundation. I would like to thank Paul Kline for his comments on this manuscript.

References

Anderson, J. R. Language, memory and thought. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Anderson, J. R. Memory for information about individuals. Memory & Cognition, 1977, 5, 430-442.

Anderson, J. R. & Bower, G. H. Human associative memory. Washington, DC: Hemisphere Press, 1973.

Anderson, J. R. & Hastie, R. Individuation and reference in memory: Proper names and definite descriptions. Cognitive Psychology, 1974, 6, 495-514.

Frege, G. Über Sinn und Bedetung. Zeitschrift für Philosophie und philosophiche Kratik, 1892, 100 (all). (Translation available in P. Geach and M. Black (Eds.), Translations from the philosophical writings of Gottlob Frege. Oxford: Basil Blackwell, 1960, 56-78).

Ortony, A. & Anderson, R. C. Definite descriptions and semantic memory. Cognitive Science, 1977, 1, 74-83.

Reference Diaries[1]

Herbert H. Clark
and
Catherine Marshall

Stanford University

When we make a definite reference to a thing, we normally make sure that our audience "shares" with us certain knowledge about that thing. To refer to a woman as she, the woman, or Nancy, we usually have good evidence that our audience knows about her too. But exactly what "shared" knowledge is required? This question is critical if we are ever to discover how people make or interpret definite reference--how they represent knowledge in memory and consult it in uttering and interpreting expressions like she, the woman, or Nancy. The question is critical if we are ever to characterize the mental archive people have for storing the facts they need to know for definite reference. We will argue that this archive has to be another detailed diary, or reference diary, supplemented by atlases, histories, and certain other reference texts. To make this argument, and to see what goes into the archive, we will examine the prior question, What "shared" knowledge is required for definite reference? As it happens, this question leads directly to a puzzle we will call the mutual knowledge paradox. It is in the solution of this puzzle that we get our best clues as to what the reference diary must be like.

The Mutual Knowledge Paradox

Imagine that there is a Marx brothers retrospective on at the local theater for which there are two or three movies a night for several evenings. Against this background consider the following scenario:

> Version 1. On Wednesday night Ann and Bob go to see Monkey Business. The next morning Ann meets Bob and asks, "What did you think of the movie?"

What we are interested in is Ann's definite reference the movie, which she intends to refer to Monkey Business. What facts does Ann have to assure herself of before she can felicitously make this reference? Our interest here is in only those facts that are involved in "shared" knowledge. As a first condition, for example, Ann must herself have a certain awareness of Monkey Business. For now we will express that awareness as "knowing about R" (where R stands for the referent Monkey Business). Thus one fact Ann must assure herself of is this:

(1) Ann knows about R.

But is this enough? Of course not, for (1) provides no assurance that Bob knows about Monkey Business. The way it fails can be made clear in a variation on the original scenario that goes like this:

> Version 2: On Wednesday night Ann and Bob go to see Monkey Business, but neither knows that the other went too. The next morning Ann meets Bob and asks, "What did you think of the movie?"

Although version 2 satisfies condition (1), Ann has clearly made her definite reference without the right assurances. If (1) were all that had to be satisfied, version 2 would lead to a felicitous definite reference. Since it does not, we must add another condition, and the obvious one is that Ann must also assure herself that Bob knows about the movie, condition (2):

(2) Ann knows that Bob knows about R.

(If it seems too strong to require knowledge instead of belief, each know can be replaced by believe; without legislating on the argument, we will stick with know).

At first, conditions (1) and (2) together seem enough, but it is easy to show that they are not. Consider this version of the original scenario:

> Version 3: On Wednesday night Ann goes to see Monkey Business, and there she sees Bob. But he doesn't see her, and she realizes this. Furthermore, she realizes that Bob, unlike herself, might have seen A Day at the Races and A Night at the Opera, which are also showing that night. The next morning Ann meets Bob and asks, "What did you think of the movie?"

Although Ann has satisfied conditions (1) and (2)--she knows about Monkey Business and she knows that Bob knows about Monkey Business--she has not yet assured herself of enough. She cannot be sure Bob won't take the movie as referring to A Day at the Races or A Night at the Opera or even some other movie. Why? Because he couldn't be sure, uniquely, which movie she had in mind that he knew about. Bob must know not only about Monkey Business, but also that Ann knows about Monkey Business. At least, this is something Ann must try to assure herself of. This leads directly to the next

condition:

(3) Ann knows that Bob knows that Ann knows about R.

With condition (3) we must surely have strong enough conditions for the success of Ann's definite reference. But that isn't so, as we can show in still another variation on the original scenario:

Version 4: On Wednesday night Ann goes to see Monkey Business, and there she sees Bob. As she walks down the aisle, she notices that he sees her, but as she is about to wave he turns and moves to another part of the theater. So she does not believe that he realizes that she has seen him. The next morning Ann meets Bob and asks, "What did you think of the movie?"

This version satisfies conditions (1), (2), and (3). Ann knows about Monkey Business; she knows that Bob knows about it; and she knows that he knows that she knows about it. But Ann doesn't believe that he knows that she knows that he knows about it. This piece of negative knowledge should be enough to keep Ann from using her definite reference. What if Bob had gone to A Day at the Races and A Night at the Opera too?, she should ask herself. He might think that while he is sure she didn't see him at Monkey Business, she might have seen him at one of the other two. If so, she might be referring to one of the other two. He couldn't be sure. According to Ann's reasoning, therefore, she must assure herself of something more--that Bob realizes that Ann realizes that he had been to see Monkey Business. That is, she must satisfy the following condition too:

(4) Ann knows that Bob knows that Ann knows that Bob knows about R.

With condition (4) it looks as if we have gone far enough (see Kempson, 1975, p. 165; Stalnaker, 1977, p. 137), but can we be sure? Only if we cannot dream up another variation that satisfies conditions (1) through (4) but still doesn't work. Indeed, with a little difficulty, we can:

Version 5: On Wednesday night Ann goes to see Monkey Business and there she sees Bob and Charles. Because she sits down a few rows in front of them, she believes that they see her there, but because she doesn't turn around while they are there, she believes that they don't realize that she has realized that they have seen her there. On the way home, however, she meets Charles, who tells her that Bob did realize that she had seen them there, but because she hadn't waved at them, Bob was certain that she didn't realize that they had seen her notice that they were there too. The next morning Ann meets Bob and asks him, "What did you think of the movie?"

Complicated as this version is, we realize that Ann in good conscience shouldn't have made this definite reference. Although conditions (1) through (4) are all satisfied, Ann should have considered this possible reasoning on Bob's part. What if Bob had seen A Day at the Races and A Night at the Opera too. He might think that she had seen him at, say, A Day at the Races and that she thought he had seen her there too. even though he hadn't. He would then have reason to think she was referring to A Day at the Races, since to have referred to Monkey Business she would have been sure that he knew that she knew that he knew that she was there (Bob's equivalent to condition (4)). So despite all of the conditions she has already assured herself of, she must add one more:

(5) Ann knows that Bob knows that Ann knows that Bob knows that Ann knows about R.

Is condition (5) enough? Hardly. What these versions show is that there is a way in principle of demonstrating that the last piece of iterated knowledge is insufficient. The method is this. Corresponding to Ann's condition (1) is an analogous condition that Bob must assure himself of if he is to uniquely identify the referent for Ann's definite reference, and it is this:

(1') Bob knows about R.

For Ann to be sure that her reference goes through, she must put herself in Bob's shoes, reason as Bob would, and make sure that he would identify the intended referent uniquely. What we did in constructing version 2 was create a scenario in which (1) and (1') held, but Ann couldn't know that (1') held. This led us to add condition (2), Ann knows that Bob knows about R, the equivalent of Ann knows that (1'). But just as Ann needs to assure herself of (2), Bob needs to assure himself of (2'):

(2') Bob knows that Ann knows about R.

But then (2') is something else Ann needs to know, as we showed in creating version 3 of our scenario, and this led to condition (3). Corresponding to (3), however, is Bob's (3'), which we used in creating version 4. In principle, we could use this procedure to construct countermanding versions ad infinitum.

The successive versions and the conditions they give rise to eventually become absurdly complicated, but they do bring out a general point. In principle, one must satisfy oneself of an infinite number of conditions either to make or to interpret a definite reference. Hence the mutual knowledge paradox. If each condition takes a finite amount of time to check, no matter how small, and if these checks cannot all be made in parallel, then making or interpreting a definite reference like the movie should take an infinite amount of time.

Mutual Knowledge

In common parlance, "shared knowledge" has several definitions. Ask your aunt what it means for the two of you to share knowledge that the mayor is an embezzler, and she would probably say, "It means that you know he is an embezzler, and that I do too." If we express the proposition that the mayor is an embezzler as p, then the first definition of shared knowledge comes out like this:

A and B share$_1$ knowledge that p $=_{def.}$

(1) A knows that p.

(1') B knows that p.

However, she might give you a more complicated answer: "It means that both of us know that he is an embezzler, and furthermore, I know that you know he is, and you know that I know he is." This leads us to a second definition of shared knowledge:

A and B share$_2$ knowledge that p $=_{def.}$

(1) A knows that p.

(1') B knows that p.

(2) A knows that B knows that p.

(2') B knows that A knows that p.

Indeed, we can define a series of types of "shared" knowledge merely by extending the list of statements. None of these finite definitions, of course, describes the "shared" knowledge required of Ann and Bob in her reference to Monkey Business. For that we need something more.

What is required, apparently, is the technical notion of mutual knowledge. It has been defined and exploited by Lewis (1969) and Schiffer (1972) for dealing with close cousins of the problem we have raised here. Mutual knowledge is Schiffer's term, while Lewis' term for the same thing is common knowledge. We have opted for Schiffer's term since it seems more transparent and less open to misinterpretation. In any case, mutual knowledge is defined as follows:

A and B mutually know that p $=_{def.}$

(1) A knows that p.

(1') B knows that p.

(2) A knows that B knows that p.

(2') B knows that A knows that p.

(3) A knows that B knows that A knows that p.

(3') B knows that A knows that B knows that p.

et cetera ad infinitum.

Heuristics for Assessing Mutual Knowledge

So far two conclusions seem firm. First, definite reference requires a certain amount of mutual knowledge. Other simpler notions of "shared" knowledge will not do. Second, it is unthinkable that speakers and listeners assess mutual knowledge by working serially, statement by statement, through the infinity of statements that make up mutual knowledge. But they surely assess it somehow, as the first conclusion seems to require. The inevitable conclusion is that they use some sort of heuristics. We will consider two families of such heuristics--truncation heuristics and co-presence heuristics.

Truncation Heuristics

The stickler in assessing mutual knowledge statements is that there is an infinity of such statements, and that is too many to check. What if people checked only a few of them--like the first four? The task could then be carried out in a finite, even short, amount of time. There would be errors, of course, but they would probably be neither very serious nor very frequent. If Ann has verified the statement (4), Ann knows that Bob knows that Ann knows that Bob knows that p, it is extremely likely, on actuarial grounds, that the higher order statements would check out too. And when she does make an error, Bob will often look puzzled or ask for clarification, which will allow her to repair her reference. Indeed, repairs are quite frequent in spontaneous speech as if speakers might be doing just that. So people could assess only a truncated part of mutual knowledge. Heuristics of this kind will be called truncation heuristics.

Are these heuristics plausible as the way people normally assess mutual knowledge? We believe not. Our doubts lie in two areas. First, it is not easy to deal with statements as complicated as (4). It is implausible that people check these statements per se. Second, the evidence needed to verify such statements anyway suggests a radically different family of heuristics.

In version 4 of our movie scenario, Ann didn't believe that Bob knew that she knew that he knew about Monkey Business, a violation of knowledge statement (4). Version 4 is complicated. Not only did we have a hard time creating it, but people have a hard time grasping it, for it is difficult to keep track of who knows what. Statements like (4) are difficult not because of their syntactic form, but because they describe reciprocal relations between two people. Whereas John Dean knew that Nixon knew that Haldeman knew that Magruder knew that McCord had burgled O'Brien's office is fairly comprehensible, John Dean knew that Nixon knew that John Dean knew that Nixon knew that McCord had burgled O'Brien's office is not. Although when we need to we can figure out fourth order reciprocal relations--not just the statements themselves, it seems highly implausible that we do so routinely.

But what counts as evidence for the truth of statements like (1), (2), (3), and (4)? Take statement (3), Ann knows that Bob knows that Ann knows about R. Obviously, Ann won't have this statement per se already stored in memory. She doesn't go through life creating statements like this for every object she or anyone else might want to refer to. Rather, what she needs to verify (3) is a piece of evidence from which she can deductively or inductively infer it. Imagine that she and Bob had gone to Monkey Business together. It is hard to think of better evidence than this that she could appeal to for the truth of (3). Of course, the inductive rules by which she infers (3) from this evidence need to be spelled out, but that doesn't sound impossible.

The fact that Ann and Bob saw the movie together, however, is more useful evidence even than that. It is also about the best evidence we could imagine for the truth of (1), and of (2), and of (4), and so on ad infinitum. It is a piece of evidence that allows Ann, in one quick jump, to be sure of the truth of all the statements. Why, then, would she want to check the statements one by one--even a truncated list of them? She would be better off looking for that single piece of evidence that could in principle confirm them all. Indeed, that is the foundation assumption of the next family of heuristics we will take up, the co-presence heuristics.

Consider the following strategy. When people make or interpret a definite reference, they try to assure themselves of mutual knowledge of the referent by searching for evidence of what we will call triple co-presence. This is evidence of a particular event in which the speaker, listener, and referent are "co-present," i.e., are "present" simultaneously, as when Ann, Bob, and Monkey Business are openly "present" together on Wednesday night. Strategies like this will be called co-presence heuristics. To see how they are reasonable, we will look at first principles.

When Lewis and Schiffer hit on the notion of mutual knowledge, both recognized the need for a finite means of handling the infinity of statements. Their solutions were essentially the same. If A and B make certain assumptions about each other's rationality, they can use certain kinds of evidence, or states of affairs, to infer that each one of the infinite number of statements in mutual knowledge is true. But how? We get some hints from a concrete illustration of mutual knowledge devised by Schiffer.

The scene: Ann and Bob are sitting across a table from each other, and there is a single candle between them. Both are looking at the candle, and both see the other looking at it too. The proposition p is that there is a candle on the table. Consider the scene from Ann's point of view. Clearly, she has direct evidence for the truth of (1):

(1) Ann knows that p.

But she also sees that Bob has his eyes open and is looking simultaneously at her and the candle. That is, she has evidence that she and Bob are looking at each other and the candle simultaneously. We will call this the simultaneity assumption. Indeed, she assumes that he is not only looking at her and the candle, but also attending to them. We will call this the attention assumption. Finally, she assumes that Bob is normal and in her shoes he would be drawing the same conclusions she is. We will call this the rationality assumption. But if Bob is attending to the candle and is rational, he has evidence for (1'):

(1') Bob knows that p.

This, however, is Ann's conclusion, and so she has evidence for (2):

(2) Ann knows that Bob knows that p.

But if Bob is rational, he will be drawing the inference that corresponds to hers--his equivalent of (2)--namely (2'):

(2') Bob knows that Ann knows that p.

Once again, this is Ann's conclusion, and so she has evidence for (3):

(3) Ann knows that Bob knows that Ann knows that p.

In like fashion, Ann would be justified in iterating this process through the remaining knowledge statements (4) through infinity, and Bob would be justified in doing the same for his

So Ann has reason to believe that she and Bob mutually know that there is a candle on the table. First, there is the "direct" evidence. She directly perceives that there is a candle on the table and that Bob is simultaneously looking at both her and the candle. Second, there are her assumptions about the situation. She assumes that Bob is consciously attending to her and the candle, that he is doing so at the same time she is, and that he is rational. The upshot is that she has no reason to believe that she couldn't confirm the knowledge statements as far down the list as she wanted to go. She is therefore justified in claiming mutual knowledge. Indeed, since nothing she doesn't know herself can be mutual knowledge, and since she can assume Bob is chronically rational, all she needs to do normally is search for evidence of her and Bob simultaneously attending to each other and the candle on the table. With this we have the essence of the co-presence heuristics: To assess mutual knowledge, people search for evidence of triple co-presence--an event in which A and B are simultaneously attending to each other noting the same evidence for p. In equation form:

Co-presence + Assumptions = Mutual knowledge

The co-presence heuristics both solve the mutual knowledge paradox and make intuitive sense. When we assure ourselves of mutual knowledge, it is unlikely that we check for a series of pieces of evidence, even as few as the truncation heuristics might let us get away with. More likely, we check for a single piece of evidence of just the right kind. The candle example suggests that what we check for is evidence of triple co-presence.

Varieties of Triple Co-presence

There are many different kinds of evidence people may use for the triple co-presence of the speaker, listener, and rererent. Some of these constitute strong evidence for triple co-presence, and others constitute weak evidence. That is, some kinds rightly give people a lot of confidence that the referent is mutually known, whereas other kinds do not. As reflected in our equation, there is a trade-off between the evidence and the assumptions. The stronger the evidence is, the fewer assumptions are needed to infer mutual knowledge. Conversely, the fewer assumptions that are needed, the stronger the evidence is considered to be. The strongest evidence requires the fewest, or weakest, assumptions.

The cornerstone of our argument is this. The prototypical kind of evidence for mutual knowledge is physical co-presence, very much as illustrated in Schiffer's candle example. It is the strongest possible evidence, the one requiring the fewest auxiliary assumptions, and all other kinds are weaker in one way or another. What follows is a tentative classification of these varieties of triple co-presence.

1. Physical co-presence. Ann, Bob, and the candle are an example par excellence of physical co-presence. Not only are the three of them physically present together, but Ann can readily assume that Bob is attending to this fact, is doing so at the same time she is, and is rational. All three auxiliary assumptions are necessary. If she believed Bob was catatonic, or hypnotized the right way, or very near-sighted, for example, she wouldn't want to assume physical co-presence. Once Ann has assured herself of the direct evidence and these assumptions, she is warranted in inferring mutual knowledge of the candle and can refer to it as the candle.

There are two distinct types of physical co-presence. Ann may refer to the candle while it is still physically co-present with them, as in The candle is romantic, isn't it? Or she may refer to the candle some time after it has been co-present with them, as in The candle was romantic, wasn't it? These two types could be called immediate and delayed physical co-presence. The first kind, on the face of it, is the stronger evidence. When physical co-presence is synchronous with the definite reference, Ann can be sure that she and Bob mutually know about the candle at the time she is referring to it. She doesn't have to count on Bob's remembering the past incident of physical co-presence, as she does in the delayed kind.

The assumptions Ann would need in order to infer mutual knowledge from immediate physical co-presence are these: simultaneity, attention, and rationality. She would need an additional one for the delayed case: simultaneity, attention, rationality, and memory. Simultaneity, attention, and rationality refer to the assumptions we have described earlier. Memory refers to the additional assumption for delayed physical co-presence: Ann must assume that Bob can and will recall the earlier incident of their physical co-presence. So far so good. The stronger the evidence, the fewer assumptions Ann needs in order to make her definite reference. Immediate physical co-presence has one fewer requirement than delayed physical co-presence.

2. Linguistic co-presence. Many things we refer to have never been physically co-present. They are often things we or someone else has mentioned in conversation. Imagine Ann saying to Bob I bought a candle yesterday. Her utterance of a candle is a locutionary act that posits for Bob the existence of a particular candle in the real world. If Bob hears and understands a candle correctly, he knows about the candle's existence at the very same time as she posits it. It is as if Ann has placed the candle on the stage before the two of them so that it would be physically co-present. So when Ann utters a candle and Bob simultaneously understands it, the two of them can be said to be in the linguistic co-presence of the candle. Once Ann has established this, of course, she can make a definite reference to it, as in The candle cost me plenty.

Linguistic co-presence is weaker evidence for mutual knowledge than physical co-presence. Seeing is believing--hearing about something isn't. To begin with, linguistic co-presence requires the assumptions of simultaneity, attention, and rationality. Ann and Bob must be attending to Ann's utterance of a candle simultaneously, and both must be rational. And like delayed physical co-presence, linguistic co-presence requires memory. For Ann to refer to the candle, she has to count on Bob's recalling the earlier incident of linguistic co-presence with her uttering of a candle. But there is an additional assumption we will call understandability. Ann must assume that Bob will penetrate her indefinite reference, a candle, and understand that she is sincerely positing the candle's existence. She must assume that Bob understands her, and he must assume that she believes he does.

3. Indirect co-presence. Imagine Ann saying to Bob I bought a candle yesterday; the wick is made of cotton. In uttering a candle Ann has established the linguistic co-presence of him, her, and the candle, but not of him, her, and the wick. How, then, can she refer to the wick? She has to assume that when Bob accepts the existence of the candle, he will also accept the existence of its wick. This way, by uttering a candle; Ann has established what we will call the indirect co-presence of her, Bob, and the wick.

The inferences required in indirect co-presence are often much stronger than those needed for wick (see Clark, 1977; Clark & Haviland, 1977). Ann can refer to something that is only likely to be associated with a thing she has already established, or even only possibly associated with it. She can tell Bob: I bought a candle yesterday, but the wrapper was torn; or I bought a candle yesterday, and the bayberry smelled great. Candles don't necessarily come in wrappers nor are they often made of bayberries, yet these are parts she expects Bob to infer on the basis of her definite references to them. So what is established may be only the likelihood or possibility of a thing being co-present with the speaker and listener. Its certain existence is established only with the definite reference itself.

Indirect co-presence is parasitic. It has to be established via some other type of co-presence--for example, physical or linguistic co-presence. Before Ann can say The price was $3 of a candle, she must already have established the candle's co-presence. She and Bob could be looking at it, for physical co-presence, or she could have just mentioned it, for linguistic co-presence. For the moment we will assume that indirect co-presence is always established via either physical or linguistic co-presence.

There is both a strong and a weak case of indirect co-presence. Instead of saying The price was $3, Ann could have said The price of the candle

was §3, providing a much more certain reference. She would have made it explicit that the price referred to is that of the candle and not of something else. Bob would then have had no trouble inferring that there was one and only one price associated with the candle. They both could then assume that they mutually knew about the price. This case may be so direct that it ought to be placed in a separate category. For now we will treat it as a very strong kind of indirect co-presence.

To infer mutual knowledge from indirect co-presence, Ann and Bob need all the assumptions of physical or linguistic co-presence, whichever is the parasite's host, plus one we will call associativity. They have to assume that each other is capable of entertaining the certainty, likelihood, or possibility of a particular part or role being associated with the thing whose co-presence has already been established. The hierarchy still works as expected. Indirect co-presence, because of its added assumption, is weaker evidence for mutual knowledge than either physical or linguistic co-presence.

4. Cultural co-presence. Even when Ann is not acquainted with Bob, she can assume there are particulars the two of them mutually know. The basic idea is that there are things everyone in a culture knows about. She reads newspapers, and so does everyone else in her culture. So Bob and she can mutually assume that they both read newspapers. Ann can then take the fact that John Dean, Michael Doonesbury, and Billy Jean King have been prominently mentioned in the newspaper as good evidence that she and Bob mutually know about these people. This is an instance of what we will call cultural co-presence. Certain particulars are assumed to be universally known in a cultural milieu--they are culturally co-present for everyone in it--and that is taken as evidence that everyone in the milieu knows about them.

The trick, of course, is to judge cultural milieus. Ann may think that she and Bob mutually realize that they are both high school graduates, or drug dealers, or nineteenth century history buffs, or New Yorkers, or telephone operators, or some combination of these, and her assumptions about cultural co-presence will change accordingly. If her assessments are accurate, her definite reference is likely to succeed, and if not, it isn't.

Cultural co-presence doesn't appear to belong to the same hierarchy as the previous three types of co-presence. For one thing, it is relatively permanent, whereas the other three are relatively transitory. Culturally known particulars take time to become familiar and to lose familiarity. Teddy Roosevelt is familiar to Americans today, just as he was 75 years ago. Particulars known by physical, linguistic, or indirect co-presence have only fleeting familiarity and then only to specific pairs of people. Mutual knowledge about these particulars is easily established, but also easily lost. For another thing, cultural co-presence is parasitic on other forms of mutual knowledge. For Ann to establish that she and Bob mutually know that they belong to the same cultural subgroup, she must find evidence of triple co-presence of that fact. She might establish it, for example, through linguistic co-presence, as in, What do you know--we're-both New Yorkers.

To infer mutual knowledge from cultural co-presence, therefore, people need assumptions that are not comparable with those of the other three types. Take Ann's reference to Hoover Tower in a conversation with Bob. First, she must assume that she and Bob mutually know that they belong to a particular cultural subgroup, say Stanford University students. We will call this assumption cultural membership. How Ann justifies this assumption, however, will not be simple. Like other types of mutual knowledge, it must be based on evidence of some kind of co-presence. Second, she must assume that virtually everyone in this cultural milieu takes it for granted that they all know about Hoover Tower. We will call this assumption universality of knowledge. The paucity of these assumptions should not fool us into thinking that cultural co-presence is strong, for they hide a tangle of complex justifications based on other pieces of evidence and other assumptions. It is best to treat cultural co-presence as incommensurate with the other three.

With cultural co-presence we have come to the last of the major kinds of co-presence. Not every kind of evidence for mutual knowledge, however, can be neatly classified as one of these four types. Some appear to require a complex combination of them, and not surprisingly, they provide intuitively weaker evidence for mutual knowledge.

Reference Diaries

If people assess mutual knowledge via triple co-presence, they must have a memory full of facts about triple co-presence. What do these facts look like? How are they represented? How are they assessed? If mutual knowledge is critical to definite reference--as we have suggested--then questions like these ought to be central to any theory of speaking, listening, or memory. Indeed, the arguments we have offered lead to a rather provocative conception of memory representation and memory search. It is provocative in that some of its critical properties are absent from most current models of comprehension and memory.

Most investigators have assumed that in processing definite reference people search memory for the particulars actually referred to. Take Ann's reference to Monkey Business. On hearing this Bob would search memory for a referential index to the intended referent Monkey Business. This index is a stand-in, so to speak, for the movie itself. Although the current models of comprehension differ in their specifics, virtually all of them assume this kind of search for the intended referent. That includes Anderson (1976), Clark and Haviland (1977), Kintsch (1974), Rumelhart, Lindsay, and Norman (1972), Schank and Abelson (1977), and Winograd (1972), to name just a few.

But if people use some kind of co-presence heuristics, then all of these models are incorrect--or at least incomplete. The point is that

Bob cannot search memory for the referent alone. That would hardly guarantee that it was mutually known to him and Ann, as it must be for her reference to be legitimate. Rather, he must search for an event that involves not only the referent but also Ann and him. That is, it must be an event of triple co-presence--of physical, linguistic, indirect, or cultural co-presence, or of some combination of these four types. In none of the current models just mentioned does the listener search for such an event.

Previous models of comprehension have treated search through memory as if it were a search through a telephone book. In a definite reference like the man in the red shirt we are told the name and address of the individual we want to get hold of. Our task is to search the telephone book for his number, our direct connection to him, his referential index. With the co-presence heuristics, memory must be more like a diary, more like the personal log Nixon kept of everything he did and experienced during his years at the White House. As before, in the man in the red shirt we are told the name and address of the individual we want to get hold of. But to find him we must search our diary for an entry that provides evidence of the co-presence of the speaker (say, Gertrude), us, and an individual of that description. The diary entry must show that we were physically or linguistically co-present, or that we were co-present in some other sense. That is, we must search in every case for an event. This is far more complicated than searching the telephone book, with or without yellow pages, for the right number.

The diary, of course, cannot be used alone. We also need histories and atlases to refer to John Dean, the Second World War, the decline and fall of the Roman Empire, and China, particulars that are culturally co-present. And for indirect co-presence we will also need texts on science, medicine, engineering, and law. To know that candles have wicks we need to look up facts about the engineering of candles.

What we need, in summary, is a diary of the significant events in our own personal experience, supplemented by cultural histories and atlases for cultural co-presence and by various reference texts for indirect co-presence. Such a diary contains a record of the events we will need for assessing co-presence. Anything less than a diary will be too little.

Footnote

[1]This paper is an abbreviated version of "Definite reference and mutual knowledge," presented at the Sloan Workshop on Computational Aspects of Linguistic Structure and Discourse Setting, University of Pennsylvania, May 1978. We thank Eve V. Clark for her helpful comments on the manuscript.

References

Anderson, J.R. Language, memory and thought. Hillsdale, N.J.: Erlbaum, 1976.

Clark, H. H., & Haviland, S. E. Comprehension and the given-new contract. In R. O. Freedle (Ed.), Discourse production and comprehension. Norwood, N.J.: Ablex Publishing, 1977, pp. 1-40.

Clark, H. H. Inferences in comprehension. In D. LaBerge & S. J. Samuels (Eds.), Basic processes in reading: Perception and comprehension. Hillsdale, N.J.: Erlbaum, 1977.

Clark, H. H., & Marshall, C. Definite reference and mutual knowledge. Paper presented at the Sloan Workshop on Computational Aspects of Linguistic Structure and Discourse Setting, University of Pennsylvania, May 1978.

Kempson, R. M. Presupposition and the delimitation of semantics. Cambridge: Cambridge University Press, 1975.

Kintsch, W. The representation of meaning in memory. Hillsdale, N.J.: Erlbaum, 1974.

Lewis, D. K. Convention: A philosophical study. Cambridge, Mass.: Harvard University Press, 1969.

Rumelhart, D. E., Lindsay, P. H., & Norman, D. A. A process model for long-term memory. In E. Tulving & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972, pp. 197-246.

Schank, R., & Abelson, R. Scripts, plans, goals and understanding; An inquiry into human knowledge structures. Hillsdale, N.J.: Erlbaum, 1977.

Schiffer, S. R. Meaning. Oxford: Oxford University Press, 1972.

Stalnaker, R. Pragmatic presuppositions. In A. Rogers, B. Wall, and J. P. Murphy (Eds.), Proceedings of the Texas conference on performatives, presuppositions, and implicatures. Arlington, VA: Center for Applied Linguistics, 1977, pp. 135-148.

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

Subsequent Reference: Syntactic and Rhetorical Constraints

David D. McDonald
MIT Artificial Intelligence Laboratory
Cambridge, Massachusetts 02139

Abstract

Once an object is introduced into a discourse, the form of subsequent references to it are strongly governed by convention. This paper discusses how those conventions can be represented for use by a generation facility. A multistage representation is used, allowing decisions to be made when and where the information is available. It is suggested that a specification of rhetorical structure of the intended message should be included with the present syntactic one, and the conventions eventually reformulated in terms of it.

Introduction

Whenever a speaker wants to refer in text or speech to some object, action, state, etc., she must find phrase which will both provide an adequate description and fit the context. What governs her choice? One way to find out might be to look at the selected phrase after the fact and try to develop a static characterization of the relation between it and its context. This is what most non-computational linguists do. However, relations derived from finished texts are at best incomplete. They will not tell us how the choice was made or even guarentee that the relation(s) was apparent when the choice had to be made.

To get a clear picture of what people know about making references, we have to focus our attention of the process that they go through. It must involve making decisions on the basis of some contextual evidence. What is the evidence? How and when is it computed? How is it described? Is the decision of what phrase to use made all at once or as a gradual refinement? How is this process interleaved with the larger process of constructing the rest of the utterance?

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defence under Office of Naval Research contract N00014-75-C-0643.

We can narrow the research problem by distinguishing two kinds of references: initial and subsequent. This classification divides instances of reference by their position in a discourse. "Initial" references introduce new entities into the discourse, while "subsequent" references are another mention of one already introduced.

An initial reference must be an encompassing enough description of the new entity that the audence will be able to recognize it. This requires matching goals with evidence from a model of what the audience is likely to already know and how likely they are to understand various choices of wording (e.g. which of its properties should be emphasized? - why is it being introduced?). This is not easy. People talking or writing about unfamiliar things or to unfamiliar audiences are not particularly good at it.

Subsequent references are another matter. They are very highly grammatisized. While an initial reference may take almost any form: noun phrases with unrestricted numbers of adjectives and qualifying phrases, nominalized clauses, verb phrases (for actions), etc., subsequent references must use very specialized forms: personal, reflexive, and personal pronouns; special determiners like "*this*" or "*my*"; class nouns like "*thing*" or "*one*"; and so on. Here, grammatical convention dictates most decisions and leaves only some details to free choice.

* * *

My observations in this paper are based on experiences with a program for generating English texts from the goal-oriented, internally represented messages of other programs. My program, and the state of the art in general, can deal much better with the representation of a grammar than with then representation of an audience model. Hence the focus here on subsequent references.

The next section looks at the course of the whole generation process as my program models it, and fits the sub-process of finding phrases for references within it. Then the process of deciding whether or not to use a pronoun will be examined in some detail. This will lead to the problem of

accessing audience models and, the idea that the relevant information should be computed outside the linguistic construction process per se. That idea is expanded to include "rhetorical structures" like the relation "all of a set" that leads to a phrases like "*...a square, ...the other square*". Finally, a design for this rhetorical structure is sketched.

Internal representation

Suppose we had a logically minded program that wanted to make the statement:

$$\forall x\ man(x) \rightarrow mortal(x)$$

People who have worked on language generation have almost universally factored out all of the program's knowledge of language into a temporally and computationally distinct component. Once the rest of the program has compiled a description of what it wants to say like the formula above - it passes it off to its "linguistic generation component" and lets it come up with the actual text.

But before moving on to that component, let us look closer at this formula. I am presuming that the speaker's primary (non-linguistic) representation, be it predicate logic, semantic nets, or whatever, uses a totally unambiguous style of representation - something equivalent to always refering to an object, etc by its unique name For example, the three "x"'s in the formula all denote the same object (albeit local). The two predicates, the quantifier and the implication sign all denote different objects.

We usually think of objects - noun phrases - as being the only things that might be refered to more than once, but that is not the case. Consider the formula **mortal(Romeo) ∧ mortal(Juliet)**. That could be rendered in any of several ways including: "*Romeo is mortal and so is Juliet*". Here the second instance of **mortal()** was realized by a special, highly restricted grammatic device - exactly the characteristics of a "subsequent reference". From the point of view of the language generation component, the important thing will be the repetition of some name from the input formula not, at first glance at least, the kind of object that name denotes. (The set of descriptive formulas supplied to the linguistics component is called the program's "message" Subformulas or terms within a message are called "elements" or "msg-elmts".)

The internal objects that appear in a speaker's descriptions will have defining and incidental properties associated with them which are accessible through their names. This will include a property (actually a packet of properties and procedures) which records what the program knows about realizing the object as an English phrase. I refer to this property as the object's "entry" - as in an entry in a translating dictionary. An entry specifies what are the set of possible English phrases that could be used for the object, and includes a set of context sensitive tests that will indicate which phrase to choose. Breaking down the speaker's "how to say it" knowledge into such small chunks facilitates the use of a general recursive process for turning messages into texts by following the compositional structure of the formula(s) from top to bottom.

Besides pointing to permanent properties, a object's name will also be the repository of more or less temporary annotations. In particular, when the generation component realizes an instance of an object as phrase, it can add an annotation to it marking what kind of phrase was selected, where in the text this occured, what the immediately dominating clause was at the time, and so on. The next time there is an instance of that same object the annotation can be found and used to help decide what kind of subsequent reference should be made.

Before the linguistic processing is begun, is it possible to examine the input formula and determine what subsequent references it will educe? The bound variable x appears three times, once with the quantifier and once with each predicate. It would be a candidate for some subsequent references if, in fact, the formula was rendered into English literally.
"For any thing, if that thing is a man, then it is mortal."

But other, more fluent, renderings of that formula will not give the x's a separate status:
"Being a man implies being mortal"

"All men are mortal"

In short, it is not possible to predict which objects will be explicitly refered to and which not just on the basis of a formula in the internal representation language. You would have to know (1) how the terms that dominate the object in the formula are going to be rendered; and (2) whether the object was mentioned earlier in the discourse and how it was described there. Then you would still have to, in effect, duplicate the reasoning process that the generation component would go through itself.

As we will see later, the generation component will often need "advice" as to whether or not the audience would understand certain phrasings. The audience model which makes these decisions will presumably prefer to work from pre calculated observations so as to avoid delay. The implication of the fact that you cannot whether that there will be a subsequent reference to a particular object until it actually happens is that you cannot make special preparations for it The audience model, or any other effected part of the program, will have to be generally prepared for whatever

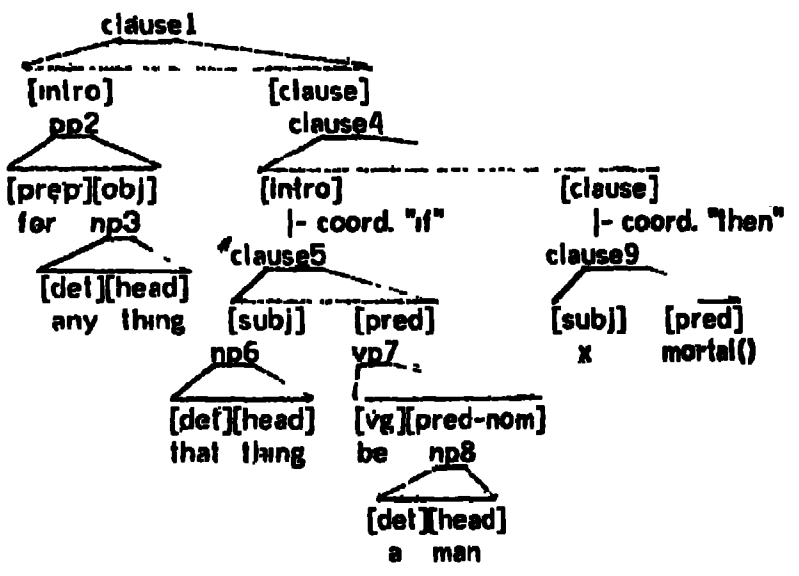
objects might be asked about.

The possibility of three different renderings for the same formula implies that the formula *per se* does not contain enough specification to pick out just one of them. If you consider the three sentences for a moment, you will appreciate that what distinguishes them are differences in rhetorical emphasis and in how to interpret ∀x. These are things that Frege deliberately omitted from the predicate calculus. To direct the generation component so as to arrive at a particular one of those sentences, more formulas would have to be added to the message or else found in the larger context (e.g. the formula might be part of a proof), and the entries for quantifiers, implication, etc. would have to be augmented to notice them.

Upgrading the predicate calculus enough to motivate the use of fluent English is a facinating problem, but one which I will gloss over in this paper. See McDonald [1978a] for more details. For now, I will assume that the decisions made by the various entries come out so as to give the literal version of the formula with the explicit references just so that we can use it for an example.

Syntactic Context

Below is my program's representation of the situation just as it is about to choose a phrase for the third instance of x in the formula. The point of showing this constituent structure is to demonstrate that while the program has a great deal of data to bring to bear on the choice, it also has a great deal of data which is utterly irrelevant to it. The packaging of the data - the size of the search space - is at least as important as having the data available in the first place.



In the diagram, the names of grammatical categories: clause1, pp, etc, denote the syntactic nodes of an annotated surface structure. Each node has a set of immediate constituents, organized by a list of named constituent slots. A slot can be empty, hold another node, hold a word or idiom, or hold an element of the input formula which has yet to be processed, e.g. x, or mortal(). The words at the leaves of the tree are given in their root form. A morphology subroutine specializes them for number, tense, etc. when they are spoken (printed on the console).

The choice of what syntactic categories, descriptive features and constituent slots to maintain is tied up with the choice of actions associated with them by the linguistics component. The [intro] constituent, for example, will act to insure that any introductory clause is realized as a participle. There are many trade-offs involved in the design of this grammar, and I will again gloss over them for this paper.

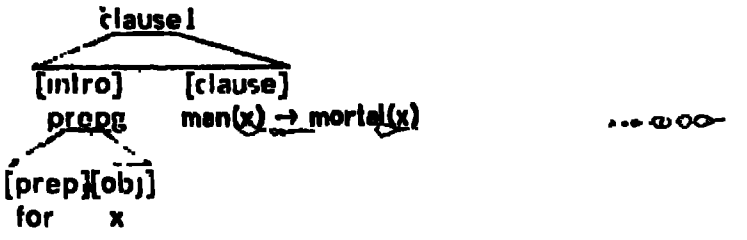
The choice of refering phrase for a subsequent reference is determined largely by the syntactic relationship between the current instance and the previous instance to the same object. In a static, after the fact analysis, we would determine this relationship by examining their positions in a tree like the one above. This is a simple enough operation for a person using her eyes, but it is an awkward mark and sweep style search for a computer program.

My program uses a much more efficient, and I would say more perspicious approach based on recording potentially relevant facts at the time they are first noticed by the linguistics component. The wording of the heuristics that are used for the decisions are similar to the wordings used in static analysis. (They almost have to be, given that that is how the bulk of linguistic research has been done to date.) But the data for the heuristics is acquired in a more natural manner.

Before discussing the program actual pronominalization heuristics, I will first digress to describe the workings of the generation process which collects (and creates) the data.

* * *

The tree in the previous column was developed incrementally. Clause1 is the result of realizing the conceptually topmost part of the input formula - the quantification. Its argument - the implication - was then positioned in the new syntactic structure but not yet realized itself. This is what the constituent tree looked like at that point.



All of the generation component's actual knowledge is spread about many small, local routines: dictionary entries for the object that will appear in input formulas; "realization strategies" - the construction routines that those entries execute to implement their decisions; or "grammar routines" -

associated with the names of categories or constituents and in charge of effecting conventional details not involved in conveying meaning These routines are all activated and organized by a simple controller.

The controller works by walking the constituent tree, top down through the syntactic nodes and from left to right at each level of constituents, The process begins with the top node of the tree just after it is built by the entry for the the topmost element of the input formula.

Outline of the Controller

Examine-node

(1) call the grammar routine for this category node
(2) rebind the node recursive state variables
(3) call Examine-constituents

Examine-constituents

- For each constituent slots of the current node in order do:

(1) call the grammar routine for that slot name
(2) call Examine-slot-contents

Examine-slot-contents

- Cases:

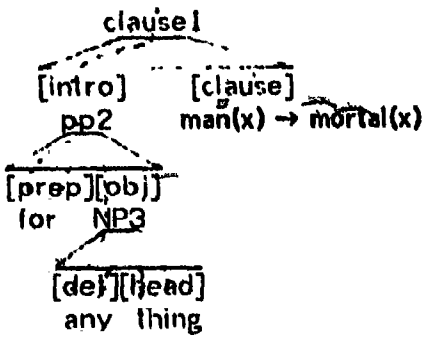
contents = nil do nothing

contents = <word>
call the morphology subroutine with the word
print the result

contents = <node>
call Examine-node

contents = <msg-elmt>
use the dictionary entry for the element to find a phrase for the element; replace the element with that phrase as the contents of the slot;
loop through the cases again.

So, having generated clause2, in effect by starting the controller on the last case of Examine-slot-contents, the controller will loop around. The contents will now be clause2; the third case will be taken and the clause "entered". Its first constituent contains another node; the controller recursively re-enters Examine-node and enters the prepositional phrase. Its first constituent contains the word "*for*", which is immediatedly printed out with no changes from the morphology subroutine; the second contains the first instance of x which is processed with the dictionary entry common to "issolated variables". The noun phrase it constructs replaces the x in the constituent tree; the controller then loops thrqugh the cases once more, recursively calling Examine-node on NP3. It is now three invocations deep. The dotted line shows its path.



spoken: "*For any thing,* "

After processing np3, the controller will leave the np and thepp, go to the next constituent of clause1, use the dictionary entry for implications, and so on, et cettera.

The design of this generation component is oriented around the decision making process of the dictionary entries (see [McDonald 1978b] for more discussion). The principle reason that the process is deterministic and indelible, for example, is to simplify the conditions that the entries will have to test for. A more relevant example here is the use the controller to "pre-calculate" certain relations about the context and make them available through the values of recursive state variables maintained by Examine-node. For example, the controller keeps pointers to the "current-main-clause", "current-verb-phrase", etc.. It keeps track of whether it is in a subordinate context, of what the last constituent was, last sentence, and so on.

Any of these relations could be calculated independantly by directly examining the form of the constituent tree and the annotations on its nodes and embedded message elements. But the point is more than just efficiency. By making certain relations readily available and not others, one says that just those relations are the important ones for making linguistic decisions. A one of a kind operation like subject-verb agreement will have a special predicate written for it that "knows" where to find the relevant subject constituent in the constituent tree. But relations that are often used, particularly those needed for evaluating pronominalization, are maintained by the controller, and, as a corollary, are only available in their pre-computed form when the controller is present at that point in the tree.

The design of the controller guarentees that the generation process will have these properties: (1) It is done in one pass - the controller never backs up. (2) Therefore decisions, choices of phrasing, must be made correctly the first time. (3) It is incremental. When the first part of the text is being printed out, later parts will be in their internal form. (4) Therefore very specific facts about the linguistic characteristics of earlier parts of the text are available to influence the decisions made about the later parts. (5) In particular, when the time comes to render any particular

associated with the names of categories or constituents and in charge of effecting conventional details not involved in conveying meaning. These routines are all activated and organized by a simple controller.

The controller works by walking the constituent tree, top down through the syntactic nodes and from left to right at each level of constituents, The process begins with the top node of the tree just after it is built by the entry for the-the topmost element of the input formula.

Outline of the Controller

Examine-node

(1) call the grammar routine for this category node
(2) rebind the node recursive state variables
(3) call Examine-constituents

Examine-constituents

- For each constituent slots of the current node in order do:

(1) call the grammar routine for that slot name
(2) call Examine-slot-contents

Examine-slot-contents

- Cases:

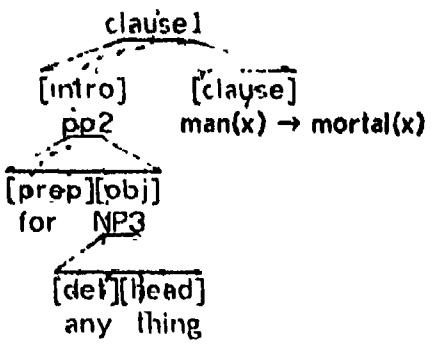
contents = nil do nothing

contents = <word>
call the morphology subroutine with the word
print the result

contents = <node>
call Examine-node

contents = <msg-elmt>
use the dictionary entry for the element to find
a phrase for the element; replace the element with
that phrase as the contents of the slot;
loop through the cases again.

So, having generated **clause2**, in effect by starting the controller on the last case of Examine-slot-contents, the controller will loop around. The contents will now be **clause2;** the third case will be taken and the clause "entered". Its first constituent contains another node; the controller recursively re-enters Examine-node and enters the prepositional phrase. Its first constituent contains the word "*for*", which is immediatedly printed out with no changes from the morphology subroutine: the second contains the first instance of x which is processed with the dictionary entry common to "issolated variables". The noun phrase if constructs replaces the x in the constituent tree; the controller then loops through the cases once more, recursively calling Examine-node on NP3. It is now three invocations deep. The dotted line shows its path.



spoken: "*For any thing,* "

After processing **np3**, the controller will leave the **np** and the**pp**, go to the next constituent of **clause1**, use the dictionary entry for implications, and so on, et cettera.

The design of this generation component is oriented around the decision making process of the dictionary entries (see [McDonald 1978b] for more discussion). The principle reason that the process is deterministic and indelible, for example, is to simplify the conditions that the entries will have to test for. A more relevant example here is the use the controller to "pre-calculate" certain relations about the context and make them available through the values of recursive state variables maintained by Examine-node For example, the controller keeps pointers to the "current-main-clause", "current-verb-phrase"; etc.. It keeps track of whether it is in a subordinate context, of what the last constituent was, last sentence, and so on.

Any of those relations could be calculated independantly by directly examining the form of the constituent tree and the annotations on its nodes and embedded message elements. But the point is more than just efficiency. By making certain relations readily available and not others, one says that just those relations are the important ones for making linguistic decisions A one of a kind operation like subject-verb agreement will have a special predicate written for it that "knows" where to find the relevant subject constituent in the constituent tree. But relations that are often used, particularly those needed for evaluating pronominalization, are maintained by the controller, and, as a corollary, are only available in their pre-computed form when the controller is present at that point in the tree.

The design of the controller guarentees that the generation proces will have these properties: (1) It is done in one pass - the controller never backs up. (2) Therefore decisions, choices of phrasing, must be made correctly the first time. (3) It is incremental. When the first part of the text is being printed out, later parts will be in their internal form. (4) Therefore very specific facts about the linguistic characteristics of earlier parts of the text are available to influence the decisions made about the later parts. (5) In particular, when the time comes to render any particular

was-a-thing, vs. was-a-proposition once and for all and makes it unnecessary for the heuristics that refer to this distinction to repeatedly include all of the particular cases. For that matter, it is also unnecessary to rewrite the code for the heuristics every time there is a new definition for a feature.

Other syntactic features currently computed include measures of relative position like **same-simplex, same-sentence,** or **stale,** and **proceed-and-command,** which are computed from the several position indexes in the record. The record of what constituent slot the last instance was in, in conjunction with the clause indexes, is used to check for features such as whether the last instance was the **previous-subject.** Also, parallel positions within conjoined phrases are noted.

Once the list of features is computed, the heuristics are run. At the moment, they are implemented as simple conditionals. Here again, there can be an immediate yes or no decision, or else a yet more involved process is invoked (see below). The grammar forces an immediate decision when **proceed-and-command** applies. Otherwise, a number of heuristics will immediately cause a pronoun to be used if there are no "distracting" references to other object in that vicinity of the discourse. For example, if the last instance of the object was itself realized as a pronoun, this will cause an immediately decision to use one again.

In the case of this example, the third instance of "x" will be described as:

same-sentence, last-subject, was-a-thing

As there are no other similar references in the vicinity to distract the audience, the heuristics will immediately decide that a pronoun should be used. The subroutine for computing the correct print name for pronouns is then consulted, and the result, "*it*" is returned to be inserted in the constituent tree and "spoken" on the next loop of the controller.

Reasoning about distracting references

Except when instance and anaphor are in the same simplex clause, syntactic relations alone are never enough to dictate whether or not a message element should be pronominalized. The linguistics component must to be able to tell if there are any other elements with which this one might possibly be confused. The problem is, of course, that the "confusion" will be a semantic or pragmatic one, i.e. it will be based on cognitive facts about the message elements which the linguistics component, **per se,** knows nothing about.

Given an oracle to tell it which message elements would compete with current one for the interpretation of a pronoun in that position, the linguistics component can use a simple procedure to decide whether to go ahead with the pronoun, namely to run those other elements through the pronominalization heuristics as well and see which accumulates the best reasons for being pronominalized.

Consider this example sentence. Imagine that the linguistics component has reached the point in brackets and must make the choice whether to say "*her*" or "*Candy's*".

"Candy asked Carol to reschedule {her, Candy's} meeting for earlier in the day."

Whether or not two objects will be ambiguous depends on what the audience knows. In this case, an audience that knows who both Candy and Carol are will know that Candy is a graduate student who might well organize a meeting and that Carol is a group secretary, someone who would probably make the arrangements needed for changing a meeting's time. For such an audience, it would be not at all confusing to say "*her meeting*". An audience that didn't know who they were however would at best be confused and would in fact probably make the wrong choice.

This kind of information is much too specific to imagine encoding as part of general purpose dictionary entries. But because of the general unpredictability at the message level of whether an object will have subsequent references made to it in the eventual text, the linguistics component will have to make its query to the main program "oracle" at the very last minute as part of pronominalization heuristics.

The oracle will presumably be some kind of audience model. But for present purposes, we can think of it as a function that takes the object we are interested in ("Candy") as its argument and returns a list of those objects that appeared in this and recent messages which the audience might confuse with it. So, in this case, if the audience knew Candy and Carol, then the oracle would return a null list, and the pronominalization option would go through. If they didn't know them, then it would return "(Carol)", and a further round of heuristics would be tried.

To compare the relative "pronominalizability" of several message elements, Pronoun? runs them separately through the analysis and evaluation procedure. But instead of acting on the evaluation directly, it makes a list of the names of the individual heuristics that each passes and then compares the two lists. In the current program these would be:

Candy
- same-sentence
- proceed-and-command

Carol
- same-simplex ;*via a trace*
- proceed-and-command
- upstairs-subject
- no-interveening-distraction

In this case, the relative number of heuristics alone would indicate that Carol would make a "better" interpretation for a pronoun in that position, and that, therefore, the possibility of a using a pronoun for Candy should be rejected. But actually, the different heuristics are given weightings. *Same-simplex*, for example, is much better evidence than *same-sentence*.

Non-pronominal subsequent references

Every subsequent reference is first checked for the possibility of using a pronoun. If this check fails, a summary vector of the features analysed and of heuristics passed and failed is passed along to the message element's dictionary entry. Entries may have their own idiosyncratic procedures for dealing with these situations, but they may also make use of general procedures packaged by the grammar.

As explained in [McDonald 1978b], the "thinking" part of a dictionary entry consists of a set of "filters", which, if their conditions are met, will execute one or more "realization strategies" which assemble the phrase or modifer that the filter set decided upon. Because entries are not evaluated directly but instead are interpreted, it is possible for the interpreter to dynamically add or subtract filter sets according to the grammatical (or rhetorical - see below) circumstances.

One of the more common reasons for rejecting the use of a pronoun is that it might be missinterpreted as refering to some other object. The form of subsequent reference eventually choosen in these cases must distinguish the object from the one it is potentially ambiguous with, but does not have to recapitulate any more detail.

In particular, one frequent pattern for an initial reference is a noun phrase with the name of a class of objects as its head word, with a series of adjectives, classifiers, or qualifying phrases surounding it. There is a simple formula for constructing a non-pronominal, subsequent reference to follow this kind of NP namely to repeat the class name as the head word and use either "*that*" or "*the*" as a determiner.

Part of an element's discourse record is a list of the realization strategies that were used in the construction of previous phrases. This is a technique for smoothing over the irrelevant detail of the actual phrase that what used. As the realization strategies are refered to by name, can be annotated with properties describing what they do, and entered into abstraction hierarchies, Routines that have to think about what other routines have done or might do can do so at whatever level of generality is appropriate. In particular, this is a way to describe patterns of noun phrase construction so that general purpose filter sets can recognize them.

The initial references pattern above is recognized by a filter set that the entry interpreter can add. The filter's predicate checks for the name of the realization strategy **head<-classname** being included as one of the **"strategies-used"** of the anaphor. If it is found, this filter set will take precedence over any others in the entry. The filter set's action will assemble a new noun phrase with the same class name, as used for initial references (it is recorded with the entry), and either *the* or *that* as the determiner depending on a heuristic measure of the distance between this instance and the last. This is the process operating in a sentence like:

"There is room for a block on a surface iff that surface is a table or has a clear top."

Subsequent references to the same kind of object

The controller makes only one pass through constituent tree, turning internal, message level structures into linguistic structures as it passes. While the amount of information available for material behind the controller is limited only by how much annotation the designer cares to record, material in front of the controller is only megerly described. The (potential) linguistic properties of an object embedded in the constituent tree in front of the controller can be explored to a limited extent by "querying" its dictionary entry. However, this is limited as a practical matter because the interveening text has not been finished and any filters in that entry which depended on the discourse context will be undefined.

This means that if you want the realization of two separated objects to be coordinated, the coordination has to be planned for well in advance and somehow marked. Otherwise the first object will be realized freely, since it would not be able to "see" that there is even a second object present. The phrases below are examples of where coordination is required. (The first two are from the tic-tac-toe talking program of [Davey 1974]. He used special purpose routines to handcraft the pairs.)

".. my edge and yours .."

"...a corner ...the opposite one "

". will enclose X's in square brackets and Y's in angle brackets"

"..a big block and a little one"

In each of these cases, the two objects were both of the same "sort": edges, corners, brackets, or blocks. By the usual criteria, this would mean that they share dictionary entries, and, indeed, the paired phrases have much in common, and could be seen as only differing in the choice of strategy for their adjectives and/or determiners. This means that the coordinating mark must be something other than the "kind-of"

pointer that links objects with their entries. It will also probably have to be a temporary structure, since *"the opposite corner"* is a transient phenomena, defined only at particular moments in each game of tic-tac-toe.

The simplest way to mark the pairs is with an additional formula in the input message, e.g.

(all-of-a-set corner1 corner9)
or (contrast-by-size B6 B3)

When the message is initially processed, formulas like these are indexed by their arguments so that, e.g., the dictionary entry for blocks will be able to notice them and choose its strategies accordingly.

Indicators like all-of-a-set are a part of the common grammar, and operate in the same way that the earlier filter set for subsequent references by classnames does. The dictionary entry interpreter keeps track of the arguments to the formula and when the last of them is being processed, it "interupts" and preempts the choice of determiner to insure that it is *the*, indicating that the speaker intends for the audience to appreciate that there is no other corner (or whatever) left. (This is a simplification.)

Rhetorical context

Rhetoric is the art of persuasion [Aristotle]. Stylistic variations in ordering, word choice, use of function words, ellipsis, etc. are potentially rhetorical techniques, if the speaker program (or rather its designer) knows when their use would have a particular desired effect, i.e. when their use would make the text more persuasive.

The rhetorical context will typically be just an additional parameter to be noticed by the entires and grammatical routines. The dimension that it adds, however, greatly increases the fluency of the linguistic component's output. The only problem is that rhetorical phenomena have not been studied much at all - they have been sweep under the rug of stylistic variations".

Goals about how to express the message's content can be specified in the message. They will have their own dictionary entries and end up determining part of the rhetorical context that accompanies the syntactic context. (At this writing, the details of the structure of the rhetorical context are still being implemented. What follows is a sketch.) Consider:

All of the pronominalization heuristics mentioned earlier were based on syntactic relations. However, there are other relations governing the understanding and generation of texts, which have to do with their rhetorical" or "discourse" structure. In particular, each region of text will have a focus - loosely speaking the object or action that that text is "about" (see [Sidner 1978] for an elaboration).

Pronominalization of subsequent references to the focused object is almost always obligatory. (There can be exceptions if the last several references to the object were pronominalized, and the intention is to "refresh" the audience's memory.) In the example with "Candy" and "Carol", if the previous part of the discourse had been saying things about Candy, then she would have been established as the focus of that sentence. Then the presence of a current-focus heuristic in Candy's list of sucessful heuristics would have outweighed all of the syntactically based heuristics in Carol's list and the pronoun would have been used.

The only question is how to mark and monitor focus or any other rhetorical indicator. It is not a natural or even consistantly definable part of a syntactic constituent structure. Therefore it will have to be "tacked on" somehow. The technique I am experimenting with is to implement a focus "register" which is explicitly set and reset by any dictionary entries that effect focus. A new message could also effect the focus register via an explicit directive included with it - say, when the topic of conversation is being changed. An explicitly dictated focus would cause the linguistics component to "transform" the realization of the content parts of the message to insure that the new focus is properly marked as such by the syntactic form of the text.

* * *

The rhetorical context could be very domain specific. Consider the sentence:

The black queen can now take a pawn."

Notice that it is not necessary to say "*a white pawn*" because immediate inference that one makes about what pieces it is legal for a piece of a given color to "take".

Since the criteria for constructing a refering expression for any chess piece will overlap, they will likely share a dictionary entry. Thus we have a sort of subsequent reference phenomena. The entry for chess pieces will be looking for the mention of a piece's color earlier in the text. If it finds one, or rather if it finds one of the complementary color, and if the situation is right, it can omit any mention of color from the phrase it has assembled.

How to determine that the situation is "right" is a matter for the rhetorical context to specify. The problem is the color of contrasting piece can be omitted only if the choice of verb or some other device indicates that, in fact, a constrasting context is present. But there are too many suitable verbs to imagine listing them in the entry and explicitly looking for them.

Instead, the rhetorical context will include a list of "relations" that currently hold. What relations there should be is a matter of the rhetorical roles that different parts of a message might play and whether the recognition of these roles by the audience could be facilitated by a choice of wording (i.e. it is a matter of research and experiment). For a program that talked about chess games, one of these relations would be:

opposing-pieces
    piece1 = xxx
    piece2 = xxx
    relation-name = {attack, defend, pin, ...}

To decide whether to include the name of a piece's color, the entry looks to see if there is an **opposing-pieces** relation holding at the moment. If there is, it looks to see if its piece is part of the relation and whether it is the second of the two to be mentioned. If so, it omits the color name.

The power of this representational technique is that it compiles its record of the needed facts at the time when they easily determined, i.e. as the message is being compiled, well before the relation name has been rendered into English and the simplicity of the relation obscured.

This technique should be applicable to many more phenomena than simply subsequent reference. Consider sentences like these:

"Brian also wants to come to the meeting."

"Mitch as a class then and so does Beth."

"The meeting might run overtime, but I don't expect it."

The underlined words are not a part of the "literal" content of those sentences. They represent rhetorical relations between parts of the sentence or between the sentence and earlier parts of the discourse.

If the source messages for those sentences described only their literal content, it would be impossible to motivate the use of *also*, *so*, or *but* in those ways, yet they are what give the sentences their naturalness. But if those rhetorical relations are included as part of the linguistic context, with their links to specific phrases and dictionary entries, including these "little" words becomes simple.

References

Aristotle, Rhetoric and Poetics, translated by Roberts, Modern Library edition, Random House, Newy York, 1954.

Davey [1974] **The Formalization o Discourse Production**, Ph.D. thesis, Edinburgh University.

McDonald [1978a] *"How MUMBLE translated the barber proof"* manuscript being readied for publication, MIT A.I. Lab. Cambridge, Mass.

McDonald [1978b] *"A Simultaniously Procedural and Declarative Representation and Its Use in Natural Language Generation"* in the proceedings of the 2d Annual Meeting of the CSCSI/SCEIO, Toronto, Canada, July 19-21, 1978.

Sidner [1978] *"The Use of Focus as a Tool for Disambiguation of Definate Noun Phrases*, this volume.

Some Psycholinguistic Constraints on the Construction and Interpretation of Definite Descriptions[1]

Andrew Ortony
Center for the Study of Reading
University of Illinois at Urbana-Champaign

Introduction

It is a curious and important fact about natural languages that they can be and often are used nonliterally. Whereas in artificial languages it is normally impossible to distinguish between the meaning of an expression on the one hand, and the intended meaning the user wishes to convey with that expression on the other (the two are identical), in natural languages a parallel distinction between sentence meaning and speaker meaning (see, for example, Searle, 1975, in press) underlies all nonliteral uses. One of the central concepts in the analysis of nonliteral uses of language is that of indirectness. It is a notion that has begun to attract the attention of linguists, philosophers, psychologists, and computer scientists in their various efforts to come to a better understanding of natural languages and of human linguistic performance. One of my purposes in this paper is to show how central a concept indirectness is with respect to the production and comprehension of definite descriptions.

The main problem with which I am concerned is a multi-level one. At the most general level it concerns the way in which people determine the referents of definite descriptions, and how language users choose the definite descriptions they do. More specifically, I am concerned with the question of the constraints that exist upon how a thing can be referred to. What makes this an interesting problem is the fact that it seems not to be necessary for a referring expression to be based on either information that has already been made explicit in the preceding discourse, nor even on information that is entailed by what has. Yet clearly, there are constraints on the expressions that can be used if there is to be a realistic hope of communicative success.

The question of what is to count as a reasonable way of referring to something in part depends for its answer on what counts as a reasonable indirect use of language. When, for example, one refers to the 1977/78 Seattle basketball team as The Cinderella of the NBA one is using a definite description based on a predicate that is not literally true of the intended referent but that is metaphorically applicable to it. As one thinks about the processes that might be involved in the production or comprehension of such an expression they appear to be very complex, yet however complex they may be, people usually engage in them without any apparent difficulty. At present there appears to exist no adequate theoretical account of what these processes are like, perhaps because a comprehensive treatment of definite descriptions has as a prerequisite a theory of indirectness, and that in turn seems to hinge on a more comprehensive theory of speech acts than is currently available. My own proposals are not intended to fill all these gaps, but they are intended to sketch a possible direction for doing so. The main goal that I have is to suggest a way of imposing limits on indirectness, and then, to show how those same limits are needed to account for some important constraints on successful definite descriptions.

Definite Descriptions and their Textual Relations

I shall take it as axiomatic that every definite description is based upon a predicate that is supposed to be uniquely applicable (at least within the context of the discourse) to some entity relevant to the discourse. Thus, the definite description The first man on the moon is based on the predicate is/was the first man on the moon, and it is assumed to be applicable to some entity (e.g. Neil Armstrong) relevant to the discourse. It is important to note at the outset that coreferential expressions cannot always be substituted for one another without a change of meaning. For instance, if, on arriving in a strange unfamiliar hotel in a foreign land one were to utter (1a), it hardly makes sense to say that it is equivalent to uttering (1b).

(1a) I feel like the first man on the moon.
(1b) I feel like Neil Armstrong.

However, if the speaker can safely assume that his audience knows that Neil Armstrong was the first man on the moon, (1b) could be used as an indirect way of achieving the communicative intent of (1a). Notice, it is not necessary to know who the first man on the moon was in order to fully understand (1a), whereas it is necessary to know that Neil Armstrong was the first man on the moon in order to properly understand (1b) (although one might not understand why the speaker used (1b) with its unnecessary demands on additional knowledge and inferences in preference to (1a).) In any event, it seems that even the relationship between definite

descriptions and proper names may sometimes depend on a notion of indirectness (see Ortony & Anderson, 1977).

The interpretation of definite descriptions often relies heavily on the establishment of inferential relationships of various kinds in order to determine which entity is being referred to. Such inferences tend to be forced jointly by the desire of the hearer or reader (hereafter referred to simply as "the hearer") to make sense of the discourse, and the assumption that the speaker or writer (hereafter, simply, "the speaker") is communicating in accordance with the cooperative principle (see Grice, 1975). This latter assumption is critically important in cases where the predicate underlying the definite description is not obviously true of the intended referent--and since these cases appear, at least on the surface. to constitute the most difficult ones, I shall concentrate on their analysis, to some extent at the expense of simpler examples. I shall call such cases "definite descriptions of inference." The overriding logic of the determination of the referents of such definite descriptions is that if the speaker is communicating in accordance with the cooperative principle certain assumptions have to be recognized in order for the expression in question to successfully identify the intended referent. These assumptions often serve to "sneak in" new information about the referent (in much the same way as appositive relative clauses introduce new information). The identification of these assumptions is based on inferences of various kinds.

Definite descriptions of inference can be contrasted with definite descriptions based on entailment relations. Definite descriptions based on entailment are those for which the referent can be determined either by transforming a predicate that has already appeared in the discourse into a definite description (descriptions based on the principle of identity), or by relating the description to predicates that have appeared earlier, on the basis of rules of formal logic (e.g. modus ponens) applied to them.[2] The important difference between a definite description of entailment and a definite description of inference is that the interpretation of the former does not depend on the provision of suppressed premises drawn from the comprehender's general world knowledge. In a definite description of inference it does. An example of a definite description of entailment can be found in (2), where the underlined expression is entailed by the content.

(2) A well-dressed man entered the room and greeted the hostess. Although everyone else was drinking sherry, he asked the waiter for a scotch. The waiter gave him one. The man with the scotch walked over to his host.

In this case, if the waiter gave the man a scotch, it entails that the man had the scotch, and so, within the constraints imposed by the context, he can be uniquely identified by the definite description the man with the scotch.

With definite descriptions of inference, as with communication in general, success often depends on the speaker and the hearer sharing a common background of knowledge (see, for example, Stalnaker, 1974). Definite descriptions of inference are more complex. For example, suppose that in (2) the sentence The waiter gave him one is omitted. Then, the definite noun phrase The man with the scotch only succeeds in referring to the right man if it is assumed that the man who asked for a scotch was given one. Unfortunately, only in biblical circles is it true that asking for something guarantees being given that thing. So, in order for the hearer to identify the intended referent he has to assume that the man got his scotch. Of course, this assumption comes easily for it can be made on the basis of a plausible inference requiring only the introduction of plausible suppressed premises, such as that when a guest asks a waiter for a particular kind of drink at a cocktail party, the waiter normally obliges if that drink is available. This constitutes a simple example of a definite description of inference.

On encountering a definite description, the hearer has to assume that the description does indeed refer to some already mentioned person or thing, in this case, say, the guest. In doing so, he makes inferences that fill in what went before--that is he makes inferences about what might have been asserted to enable the predicate underlying the description to be both applicable and relevant. The comprehender might reason as follows in the present example: "If this expression refers to the guest, then it must be the case that the waiter gave him a scotch. This is quite plausible since it is customary for waiters at cocktail parties to give guests the drinks they request if those drinks are available. It is plausible that scotch was available, since it is a frequently served drink at such occasions. So I shall assume that this is what happened and that is why the guest was referred to as 'the man with the scotch' " Whether or not people normally construct such chains of reasoning in order to identify the antecedents of definite descriptions is not the issue here. What is the issue, as we shall see later, is that it be possible to construct such a chain. Certainly, one has to suppo that the kind of general world knowledge required to do so is normally available during the comprehension process. The relevant frames, scripts, schemata, or whatever other knowledge structures are supposed, are presumably activated.

Definite descriptions of inference involving indirectness, like (3) and (4) below, tend to be more complex. They are characterized by the fact that the applicability of the predicates underlying them often depends on the utilization of knowledge that Morgan (1978) calls knowledge about the language, as opposed to knowledge of the language. These turn out to be cases of inferences involving knowledge about illocutionary forces and perlocutionary effects (see Austin, 1964).

(3) The hostess offered the guest some cake. He told her that he was on a diet. His brother, who was with him, told her that he personally was not

on a diet. The man who had refused the cake walked over to his host.

(4) The hostess asked the man where his wife was. He replied "Mind your own business, you old bag." The hostess was furious that the man who had insulted her had been invited to her party.

The interpretation of (3) requires not only semantic and general world knowledge in the way that (2) does, it also requires the knowledge that saying that one is on a diet can count as rejecting an offer to eat something. In the case of (4) it requires the knowledge that the violation of certain language-use conventions can count as offensive behavior. Of course, in a sense, this kind of knowledge about the conventions of language use and the social/communicative consequences of their violation is knowledge of the world, just as knowing that waiters normally serve the drinks they are asked to is. But, insofar as it is knowledge of conventions about the use of language, and insofar as this is an area which has been singled out as being of core concern in pragmatics, it is worth separating such cases from the other kinds of cases, like (2). In fact, I think, the distinction is difficult to uphold because the mechanism required to deal with indirectness is the same kird of inferential mechanism as is required to deal with "ordinary" knowledge of the world.

The question that eventually has to be answered concerns the constraints that there are on the predicates employed in definite descriptions. My view is that the answer to this question depends on finding an answer to a more general question about the pragmatics of language, namely the question: what constraints are there on what is relevant (in the sense used by Grice, 1975 and others). Staying, for the moment, with definite descriptions, compare (5) and (6) below

(5) The hostess offered the guest some cake. He told her that he was on a diet. His brother, who was with him, told her that he personally was not on a diet. The man who thought he ought not to eat fattening things walked over to his host.

(6) * The hostess offered the guest some cake. He told her that he was on a diet. His brother, who was with him, told her that he personally was not on a diet. The man who was not hungry walked over to his host.

It seems to me that whereas (5) is perfectly coherent, (6) is not. It becomes coherent, however, if the context is changed so that instead of (he) told her that he personally was on a diet it reads (he) told her that he personally had just eaten, then both (3) and (6) seem perfectly acceptable. It seems, then, that the appropriateness of the definite description depends on the appropriateness of its underlying predicates. Telling someone that one has just eaten is an appropriate, relevant, piece of information for permitting the inference that one is not hungry, and/or that one does not want the offered food. By contrast, telling someone that one is on a diet is an appropriate piece of information for permitting the inference that one does not want what is being offered, but it is not appropriate for the inference that one is not hungry.

Indirect Speech Acts

A major part of my thesis is that the predicate underlying a definite description of inference is constrained by the relevance relation in just the same way as that relation constrains what counts as an indirect speech act in a discourse. If this is so, then it will help to have a working hypothesis about the constraints that exist on indirect speech acts.

Suppose the situation is that described in (3), namely, one in which someone is offered some cake and in uttering (7) intends to refuse the cake.

(7) I am on a diet.

The question we have to answer is this. Since it does not follow logically from (7) that the intention was to refuse the cake, on what basis does a listener come to the conclusion that indeed that was the intention? Furthermore, why does, for example, (8) not succeed in communicating the refusal?

(8) My mother is an opera singer.

Perhaps one should reject (8) on some very general grounds. For example, on the grounds that one cannot randomly assign a sentence to an intention and expect to be understood. But the same old question arises about what constitutes a random versus a non-random assignment as arises about appropriateness and relevance.

The solution I propose is based on the notion of a "plausible chain of reasoning." It is this. For an indirect speech act to be understood as being relevant, or appropriate, it must be able to participate as a premise, or as a sub-conclusion, in a quasi-logical, or better, psycho-logical, chain of reasoning that plausibly relates the event that initiates it to its intended illocutionary force. To see the full implications of this proposal, let us see how it works with the example. The man is offered some cake, and this offer is the event that initiates his response. From the perspective of the man, (9) is true.

(9) I am being offered some cake.

Such an event calls for one of two responses, an acceptance or a refusal, appropriately modified by politeness conventions. Let us assume that "Yes, please" and "No, thank you count as direct, literal speech acts for accepting and refusing, respectively. They certainly are conventionally regarded as direct ways of accepting and refusing. Now we can see that in this particular case, the proposal is this for (7) to be understood as a refusal, it must be able to participate as a premise, or as a subconclusion, in a psychological chain of reasoning that plausibly relates

the original offer to its acceptance or rejection. Such a chain of reasoning might look something like (9) - (15).

(9) I am being offered some cake
(10) I am on a diet
(11) People on diets ought not to eat fattening things
(12) Cake is fattening
(13) (It follows logically that) I ought not to eat any cake
(14) (It follows deontically that) I will not eat any cake
(15) (It follows conventionally that) I will refuse the cake

This chain of reasoning, including the intermediate and final conclusions does not constitute a deductively valid argument in the usual logical sense. The relationships that exist between (13) and (14), and between (14) and (15) are not entailment relations, but they are characteristic of human reasoning.

A number of important observations have to be made about the chain of reasoning--observations that amount to constraints on what it normally is. First, there are no unnecessary premises in it. Every premise is needed for the establishment of the first subconclusion, (13), which in its turn is needed for establishing the final conclusion. Second, although the order of the premises that are introduced from the speaker's general knowledge can be manipulated, the most natural order is one in which each premise invokes a concept that has been foregrounded (in the sense of Chafe, 1972) by the preceding one. If this were not the case, the possibility of introducing irrelevant premises would arise--a possibility that could serve no useful purpose in the present context. In fact, this constraint probably needs to be a little more liberal than I have described, but for the reasons I have indicated, something close to it needs to operate. Third, the conclusion of the chain contains the information appropriate for a direct response to the initiating event, an event that need not itself be a linguistic one (as it is in the present example). The initiating event might be an observed event to which an appropriate response might be a description of it, or of a reaction to it. Consequently, in the general case, "response" should not be taken to mean "reply."

There are doubtless other constraints that a more detailed analysis would reveal, but for the moment I want only to suggest that the conjunction of these (or some comparable set of) constraints constitutes what I mean by "plausibility" in the context of my requirement that the chain of reasoning be a psycho-logically plausible one.

We are now in a position to consider what happens from the perspective of the hearer. The most important thing is that the hearer assumes that the speaker is constrained in what he says in just the kind of way that I have indicated. The hearer, therefore, attributes to the speaker some plausible chain of reasoning. However, the hearer may not have all the knowledge that is available to the speaker (he may not know that he is on a diet, for example). Consequently, he may have to make inferences of his own in order to reach some of the premises required. This would be true if, for example, the response to the offer of cake had been (12) rather than (10). Sometimes these inferences are incorrect and one understands correctly what was intended, but for the wrong reasons, or one misunderstands it altogether. As we shall see, this fact, that the hearer's interpretation is only probabilistically determined, has some important consequences for the speaker's selection of his utterance.

The most crucial claim that I wish to make about the chain of reasoning is this. Assuming that the speaker does not choose to express himself directly (for whatever reason), then within the limits of the context, any of the premises or subconclusions in the chain from the initiating event to the (direct) conclusion can function as more or less easily interpretable surrogates for the conclusion--any of the steps can constitute an indirect speech act appropriate to the direct speech act that constitutes the conclusion. Thus, any of (10), (11), (12), (13), and (14) can serve as indirect response to the offer. And, if some other response is made, it must be able to serve as a step in a similar chain of plausible reasoning. If it cannot, it is an inappropriate response. It is precisely these constraints that prevent (8) from being a possible indirect response to the offer, since there is no basis of shared knowledge that will normally permit a hearer to reconstruct an argument in which (8) figures to be relevant on the chain from initiating event to conclusion.

An important question that now needs to be answered is why do people use language indirectly in the first place, and why, given that they can choose from a restricted range of indirect communicative acts, do they select the ones they do. Why, for example, would a speaker choose (10) instead of, say (12)? The answer to the first part of the question depends on exactly what kind of indirect language act is being used. For example, metaphors may be used for purposes of communicative economy, communicative vividness, or even communicative possibility (see Ortony, 1975). With indirect speech acts, the answer is very often that the speaker gets "two for the price of one." For example, he can, with one utterance, not only refuse the offer, but also satisfy certain social conventions by providing a good reason for his refusal, or at least hinting at one. As Searle (1975) points out, in an indirect speech act the speaker intends both the sentence meaning and the speaker meaning to be recognized by the hearer. So, indirectness affords economy as well as, often, politeness and sensitivity.

There remains the question of why a speaker should select one form over another. The answer again lies in the fact that the communication of the literal meaning of the indirect language act is intended. Some of the knowledge that is needed to construct the reasoning chain may be more publicly available than other knowledge required. Thus, most people know that people on diets ought not to eat fattening things

(depending on the purpose of the diet, of course). Consequently it can be assumed that a hearer has more ready access to that fact than to the fact that the speaker is on a diet (which possibly very few people know). Thus, the speaker's selection of the particular language act can take advantage of his beliefs about what the hearer is likely to know. It can also take advantage of the fact that some of the choices seem to have a stronger force than others. This is a complex issue. My intuitions are that (13) leaves open the possibility of ultimately accepting some cake rather more readily than does (10), perhaps because once (10) is used it must be relevant to the chain of reasoning, whereas if (13) is used, it could be used to reach a different conclusion. After all, most people occasionally do things that they ought not to do, and that possibility seems wide open if the response to the offer is (13). This is not the place to explore these issues further, but it is worth noting that many jokes capitalize on expectations of plausible reasoning chains of the kind I have been discussing--the trick is to make them go awry!

My proposal shares certain characteristics with that of Searle (1975) in that it suggests a not necessarily conscious chain of reasoning. It differs from Searle's account insofar as it makes claims about the constraints on what can be said and understood. Searle's chain of reasoning contains many metalinguistic premises about indirectness that I have treated as background assumptions. My focus, by contrast, is on the content of the chain. What I have proposed is a possible answer to the question "How indirect can an indirect speech act be?" I have suggested that it cannot be so indirect that it could not participate in a chain of plausible reasoning relating a representation of the initiating event to an appropriate direct response to that event. I have also suggested that the illocutionary effect of all steps within such a chain will be appropriate for that initiating event.

Definite Descriptions of Inference

When speakers and writers produce, and hearers and readers comprehend definite descriptions, they do so against a background of knowledge that includes their tacit knowledge about indirectness. This knowledge is often brought to bear in dealing with definite descriptions of inference--descriptions, that is, in which the underlying predicate could appear on the reasoning chain and that could constitute a direct or indirect speech act. Thus, for example, (3) and (4) are cases in which the underlying predicate could constitute the conclusion of a chain of reasoning--i.e. a direct language act, while (5) is a case in which the underlying predicate could constitute a premise in a plausible chain of reasoning--i.e. an indirect language act.

In a sense, what I have proposed is a partial account of relevance in Grice's (1975) sense, or perhaps better yet, a partial account of when apparent violations of relevance are indeed only apparent, and why. It is quite clear that the predicates underlying definite descriptions have to be relevant to the discourse just as any other comparable meaningful components of it must be. This is the sense in which I claim that the constraints that govern what definite descriptions can be used by a speaker who hopes to be understood are the same as those that govern what a speaker can in general say, if he has those same aspirations. There is no doubt that a detailed translation of my proposals about indirectness into comparable ones about definite descriptions is no easy matter. One reason is that the reasoning process that underlies the determination of a referent may be from a conclusion to an initiating event, as in (3). Another is that not all the steps in the chain can be employed, but only those that contain information applicable to the referent--not, for example, generalizations like (11). In such cases, if the premise is to be incorporated, it has to be embedded as the complement of an appropriate verb of propositional attitude. Nevertheless, it seems to me that some of the notions that I have laid out might prove helpful, if only by virtue of the fact that they may eventually lead to better proposals by others.

Earlier, I suggested that perhaps descriptions of inference involving indirect speech acts and those not involving them, really hinge on fundamentally the same kind of processes. On the surface, the basic difference concerns whether or not they involve the addition of pragmatic knowledge. It turns out, however, not to be an easy matter to decide what is pragmatic knowledge and what is merely semantic or factual. For example, (16) is a description of inference:

(16) The navigator had heard that the weather might be unpleasant. He had always been concerned for the comfort of the passengers. He proposed taking a more indirect route to avoid the possible storms. The captain disagreed. He felt that the sooner they arrived at their destination the better--he wanted a drink and a decent meal. After a long argument the cautious one got his way.

Now, for a hearer to determine that the cautious one and the navigator are coreferential, it is necessary for him to invoke general knowledge about what constitutes a cautious act (contrast this with an offensive act). This in turn requires inferences to be made about human actions and intentions. To be sure, the actions in question are not linguistically performed acts, but that appears to be the only difference. Furthermore, had the pilot asserted that he wanted to take the shortest route because of his frivolous desires, would he not, thereby have been indirectly recommending a (possibly) reckless act? Surely, what is pragmatic and what is not cannot come down to performative verbs. Yet, if it is to be broader than that, what criteria are to be used to separate the semantic from the pragmatic? The old notion of semantics as entailment is certainly too restrictive to be useful as a model of natural language processing, but the new notion of pragmatics seems to amount to little more than the notion that language processors are rational beings who engage their reasoning processes in language comprehension and production

just as they do in perception and action. Even the notion of a speech act seems to have very fuzzy boundaries unless it is trivialized by invoking psychologically uninteresting surface structural aspects like the presence or absence of performative verbs.

Distinctions between different classes of linguistic phenomena are usually difficult to maintain in any rigid way, particularly if they are supposed to have psychological correlates. This is true of the distinction between syntax and semantics, of that between semantics and pragmatics, of that between literal and nonliteral, and of that between descriptions of entailment and descriptions of inference. As usual, clear cases are easy to recognize, but there is always a large grey, undecided area in the middle where the classification seems sterile. In the case of the distinction between descriptions of entailment and descriptions of inference, the problem is exactly the same as the classical philosophical one that plagues the analytic/synthetic distinction. This is hardly an accident since my distinction is really no more than the analytic/synthetic distinction in disguise. Maybe all that needs to be said is that some inferences (e.g. ones based strictly on the rules of logic) are generally easier to make than others. If this is right then it merely means that some relationships between descriptions and their intended referents are more transparent than others. Nobody could object to that.

The last question I want to deal with is the psychological status of my claims, particularly with respect to the inference patterns that I have proposed. My position is not that it is a necessary condition for the comprehension or production of a definite description of inference that a person actually construct such a chain of reasoning. My claim is only that it should be possible to do so--there has to exist some determinable connection between the predicate underlying the definite description and the discourse in which the description occurs. But, being determinable and being determined are different things. As a matter of fact, there are often other clues that will permit the hearer to make a good guess about the referent's identity, discourse topic being one of them. It is almost certainly the case that people sometimes do go through some such reasoning process as I have outlined, and if and when they do not, they could probably be induced to do so by being asked suitable questions about what they took the referent to be, and why it was reasonable or plausible to do so.

References

Austin, J. L. How to do things with words. Oxford Oxford University Press, 1962.

Chafe, W. L. Discourse structure and human knowledge. In J. B. Carroll and A. O. Freedle (Eds.), Language comprehension and the acquisition of knowledge. Washington, D.C.: V. H. Winston, 1972.

Grice, H. P. Logic and conversation. In P. Cole and J. L. Morgan (Eds.), Syntax and semantics (Vol. 3): Speech acts. New York: Academic Press, 1975.

Morgan, J. L. Two types of convention in indirect speech acts. In P. Cole (Ed.), Syntax and semantics (Vol. 9): Pragmatics. New York: Academic Press, 1978.

Ortony, A. Why metaphors are necessary and not just nice. Educational Theory, 1975, 25, 45-53.

Ortony, A., & Anderson, R. C. Definite descriptions and semantic memory. Cognitive Science, 1977, 1, 74-83.

Searle, J. R. Indirect speech acts. In P. Cole and J. L. Morgan (Eds.), Syntax and semantics (Vol. 3): Speech acts. New York: Academic Press, 1975.

Searle, J. R. Indirect speech acts. In A. Ortony (Ed.), Metaphor and thought. Cambridge: Cambridge University Press. In press.

Stalnaker, R. C. Pragmatics. In D. Davidson and G. Harman (Eds.), Semantics of natural language. Boston: D. Reidel Publishing Co., 1972.

Footnotes

[1]This research was supported in part by the National Institute of Education under Contract No. US-NIE-C-400-76-0116, and by a Spencer Fellowship awarded to the author by the National Academy of Education. I am grateful to my colleagues Glenn Kleiman, Robert Kantor, and Jerry Morgan for their helpful comments on earlier drafts of this paper.

[2]I call them definite descriptions of entailment because technically they both are. The principle of identity, that $p \supset p$, represents an admittedly trivial entailment. It is important in the present context because it represents the case in which some predicate is literally transformed into the body of a definite description. More complex cases are still based on the usual rules of propositional logic such as *modus ponens*, $((p \supset q).q) \supset q)$.

[3]Caution is needed here. Some cases of giving do not entail having. One can give somebody a pat on the back, or a kick in the teeth; the recipient gets it alright, but he doesn't have it! However, if we specify the appropriate constraints on the object the entailment will hold.

BOUND VARIABLES AND OTHER ANAPHORS

Barbara H. Partee
Univ. of Mass., Amherst

When a noun phrase or a pronoun occurs in a sentence, it is frequently appropriate to ask what entity it refers to, but it is well known that not all uses of noun phrases and pronouns are referential in this simple sense. In computational approaches to language processing, I believe the main thrust in this area has been toward understanding those referential uses of NP's and pronouns which require the use of both linguistic and non-linguistic inferences to determine the most plausible referent for the expression. My emphasis in this paper will be somewhat different. I believe that recent work by linguists, logicians, and philosophers is leading to convergence on the view that there are two fundamentally distinct uses of pronouns which have to be treated quite separately: (i) a use that corresponds to the logician's use of bound variables, and (ii) a use which I will call, for want of a better name, a pragmatic use. It can be argued that bound variable pronouns are restricted to occurrences in syntactic construction with their antecedents, and are fully interpreted at the level of semantics, while pragmatic pronouns need not have linguistic antecedents at all, and require pragmatics as well as semantics for their interpretation.

1. The basic distinction.

The clearest cases of bound variable anaphora involve antecedents like every man and no man which are singular in form but do not refer to individuals, as in (1) and (2).

(1) Every man put a screen in front of him.
(2) No child will admit that he is sleepy.

When the he of (2) is understood as anaphorically related to the noun phrase no child, the he clearly does not refer to a particular individual Rather, the sentence can be understood as the result of binding an open sentence, (3), with a quantifier phrase, no child.

(3) He_0 will admit that he_0 is sleepy.

(It is immaterial for the purposes of this paper whether we view the process in question as a generative one, as in Montague (1973) or Lakoff (1971) or as an interpretive one, as in Jackendoff (1972) or the I-grammar Montague variant of Cooper and Parsons (1976). The use of subscripted pronouns

rather than x's and y's follows Montague's practice, but that distinction is also immaterial here.)

The semantics of variable binding is well studied in logic; a particularly clear and brief account can be found in Quine (1970). The crucial point here is that the semantics involves consideration of a whole range of possible values for the variables, not the determination of any single value or referent. Equally crucial is that the interpretation of (2) involves an open sentence with two occurrences of the free variable he_0, one in the position of the antecedent noun phrase, the other corresponding to the surface pronoun.

Using these clear cases, we can discover strong syntactic constraints on the occurrence of bound variable pronouns. With few exceptions, it appears that bound variables must be in construction with their antecedents (the observation is made by Evans (1977); the notion "in construction with" comes from Klima (1964): a constituent A is in construction with a constituent B if and only if A is dominated by the first branching node which dominates B. The term c-command is a more recent alternative name for the same notion.) Thus the following do not permit a bound variable reading:

(4) (a) Every man walked out. He slammed the door.
(b) John loves every woman, and he hopes to date her soon.
(c) If no student cheats on the exam, he will pass the course.

By contrast, the bound variable reading is permitted in cases like (1) and (2) above, in which the pronoun is in construction with its antecedent.

The clearest cases of what I am calling pragmatic uses of pronouns are cases where a pronoun is used with no linguistic antecedent at all, as in (5), or where the antecedent occurs in an earlier sentence of a discourse, as in (6).

(5) (On walking into a room) Why is he [pointing] here?
(6) I couldn't reach Elliot last night. He is probably in Boston.

These are cases where the pronoun is being used to refer to a particular individual, and the determination of which individual the intended referent is requires making use of the linguistic and non-linguistic context. Ignoring some complicated cases that I will discuss later, we may say that at the level of purely linguistic description, such pronouns function like free variables which are not bound at all at the semantic level. A sentence containing one expresses a determinate proposition only relative to a particular choice of value for the variable, much as a sentence containing the word now expresses a determinate proposition only relative to a particular time of evaluation. Such choices depend on the context of use of the sentence, which is why I call this a pragmatic use of pronouns.

I believe that there are no absolute rules governing the choice of referent for pragmatic uses of pronouns, but that there are discoverable strategies and principles governing the relative likelihood or or preference among choices. The other participants in this panel know much more than I do about what those principles and strategies are; I hope they would agree that the output of such principles is a probable or expected referent rather than an absolute referent for the pronoun. For example, in most contexts, the probable referent of the he in (6) is Elliot; but one can easily enough imagine a context where speaker and hearer are most interested in figuring out where Max is, and being unable to reach Elliot is a good clue to Max's being in Boston; then he may be intended and understood as referring to Max. What matters most seems to be the salience and relevance of a particular individual, and I see no reason to draw any theoretical line between cases where that salience comes from the linguistic context as opposed to the non-linguistic context.

Where I do want to draw a sharp line is between the bound variable use and the pragmatic use of pronouns. The bound variable use is best described at the level of syntactic form and semantic interpretation of single sentences, and the relevant question is not what the pronoun refers to, but what quantifier phrase is binding it. The pragmatic use is best described at the pragmatic level, where the full context of the sentence in use is considered; on the syntactic level, these pronouns are really no different from proper names, and at the semantic level, they can be viewed as free variables or as dummy names.

2. Structurally ambiguous pronouns.

I have begun with the clearest examples of the distinction; if all uses of pronouns fell unambiguously into these two categories, I could stop here. All the rest would be a matter of improving the description of the syntactic constraints on bound variable anaphora and unravelling the processing mechanisms that we use to determine the referents of the pragmatic uses of pronouns. But the clear cases do not provide a set of necessary and sufficient conditions for telling the two kinds of pronouns apart. All we can conclude so far by way of conditions is the following:

(i) A pronoun can function as a bound variable only if it is in the same sentence as its antecedent.
(ii) Any pronoun can be used pragmatically.[2]

If these are the only conditions, we would expect many occurrences of pronouns to be ambiguous as to which use they have, and indeed many are. The pronouns in (1) and (2) are ambiguous in this way and the sentences have sharply different interpretations in the two cases. But now consider a sentence like (7):

(7) The prosecutor believed that he would win the case.

This example can be analyzed either way; if the pronoun is analyzed as a bound variable, the sentence is interpreted as in (7a), and if the pronoun is treated pragmatically, we can represent it as in (7b).

(7a) (The prosecutor: he_0) [He_0 believed that he_0 would win the case.]
(7b) The prosecutor believed that he_5 would win the case.

On the pragmatic pronoun reading, the free variable he_5 will be interpreted as some salient individual determined by the context; and one likely choice will be the prosecutor. This looks at first as if we are predicting an ambiguity where there is none. And this is not just an isolated example, since the same situation will arise whenever we have an antecedent noun phrase that picks out a particular individual. But it turns out that there is striking evidence that this is a real structural ambiguity, and not just an artifact of the analysis. I believe that Keenan (1971) was the first to point this out; Sag (1976) and Williams (1977) discuss such cases extensively. The evidence comes from verb phrase deletion. and involves examples like the following:

(8) The prosecutor believed that he would win the case, and so did the defense attorney.

The missing verb phrase can be understood in just two ways, corresponding to the two structures we have posited for the first clause. On each reading, sentence (8) predicates the same property of the defense attorney as it predicates of the prosecutor: either the property of being an x such that x believed that x would win the case (the bound variable reading), or the property of being an x such that x believes that he_5 (the prosecutor) would win the case (the pragmatic pronoun reading).[3] Thus the examples of so-called "sloppy identity" (Ross 1967) of pronouns are really examples of strict semantic identity of predicates. This important generalization can be captured only by recognizing that apparently unambiguous sentences like (7) are in fact structurally ambiguous.

Cases with proper names as antecedents to pronouns work just like (7) and (8), the unified treatment of all noun phrases, including proper names, as quantifier phrases proposed by Montague (1973) is an important aid in permitting the treatment of pronouns advocated here.

Another major source of pronoun ambiguity is the systematic ambiguity of most plural noun phrases as between a "group" reading and an "individual" reading, as in (9).

(9) Three men lifted the piano.

When the plural pronoun they is used as a pragmatic pronoun, it always refers to a group; but when it is used as a bound variable, it may be either a variable over individuals or a variable over groups. Thus we get two bound variable readings plus a pragmatic pronoun reading for (10).

(10) The Democrats voted for their wives.

On the group-level bound variable reading, the Democrats as a group voted for their wives as a group. On the individual-level bound variable reading, each of the Democrats voted for his own wife. On the pragmatic pronoun reading, the Democrats voted for some group's wives; that group might be the Democrats themselves, but might be some other group determined by the context. Again the three readings lead to corresponding readings in sentences with verb-phrase deletion:

(11) The Democrats voted for their wives before the Republicans did.

I will not enumerate the readings, but it can be seen that the positing of the three structures for the first clause plus the requirement that verb phrase deletion be interpreted as semantic identity of predication makes the correct predictions about the possible interpretations of the full sentence.

Yet another source of structural ambiguity is the fact that noun phrases may have other noun phrases embedded within them, and a pronoun may have either the whole noun phrase or a subpart as antecedent. Sentences (12a) and (12b) do not have this particular ambiguity because of the number difference, but (13) is ambiguous as between (13a) and (13b).

(12) (a) One of the prisoners believed that she could escape.
(b) One of the prisoners believed that they could escape.[5]
(13) Two of the prisoners believed that they could escape.
(a) Two of the prisoners believed that they could escape.
(b) Two of the prisoners believed that they could escape.

Each of these sentences is ambiguous between a bound variable use and a pragmatic use of the pronoun; and sentence (13a) permits either the individual-level bound variable reading (each of the two believed she could escape) or the group-level reading (both believed that both could escape). However, (13b) on the bound variable reading must be a group-level pronoun, because the antecedent is in a partitive construction, which requires a group-denoting noun phrase. A fuller discussion of plural noun phrases and bound variable pronouns can be found in Bennett (1974), although Bennett does not specifically discuss the pragmatic uses of pronouns. No new principles of pronoun interpretation are needed for these cases beyond the important observation that they can function semantically as an individual-level pronoun, that is, just like a singular pronoun. The complexities of these examples result simply fron the joint interaction of several individually simple phenomena: bound variable vs. pragmatic uses of pronouns, individual vs. group readings of plurals, and the possibility of either a whole noun phrase or a subpart of it serving as antecedent for a pronoun.

The examples discussed so far are summarized and extended in Table I below. The column headed "Pragmatic Pronoun" should be understood as follows: the given pronoun can be interpreted as referring to an individual or group determinable on the basis of the interpretation of the given "antecedent" as the relevant linguistic context. Thus, for example, while every man does not refer to the group of all man, it can promote that group into salience, as can no man and no men.

(14) No students came to the party. They thought they weren't invited.6/

TABLE I

Antecedent	Bound Variable Pronoun	Pragmatic Pronoun
every man	he	*he, OK they (group)
no man	he	*he, OK they (group)
the man	he	he
John	he	he
one man	he	he
more than one man	he	*he, OK they (group)
at most one man	he	he
three men	they (ind), they (group)	they (group)
one of the men	he	he
{one / two} of the men	they (group)	they (group)
two of the men	they (ind), they (group)	they (group)
no men	they (ind)	they (group)
John and Bill	they (ind), they (group)	they (group)
John or Bill	he	they (group)7/

3. Are there "pronouns of laziness"?

Both traditional grammar books and early transformational accounts such as Lees and Klima (1963) suggest a treatment of pronouns different from either of the two I have described. This is the view that a pronoun is a substitute for a linguistically identical noun phrase;(15b) would on this view be derived from (15a).

(15) (a) John spoke to Mary when John walked in.⇒
(b) John spoke to Mary when he walked in.

But such a view requires that semantic interpretation operate on surface structure, since the application of the rule changes the meaning whenever the repeated noun phrase is anything other than a proper noun or a definite description.

(16) (a) John lost a watch and Bill found a watch ⇒
(b) John lost a watch and Bill found it.

Given that pragmatic pronouns must be generated directly anyway because of cases where there is no linguistic antecedent, there is then no work left for such a transformation to do; it simplifies neither the syntax nor the semantics. Hence it has been abandoned by linguists of just about every theoretical persuasion.

But there are some cases that look as though they might be better handled via a syntactic substitution rule than by either the bound variable or the pragmatic treatment. One class was introduced by Geach (1962), who provides examples like (17):

(17) Every man who owns a donkey beats it.

On the defensible assumption that a donkey should be analyzed here as an existential quantifier phrase having narrower scope than the every, this it cannot be analyzed as a bound variable (see Partee 1975a). But it also does not refer to any specific donkey, and so does not appear to be functioning as a pragmatic pronoun. Geach suggests that a sentence like (17) be analyzed in terms of (18):

(18) Every man who owns a donkey beats the donkey he owns.

Thus the it is viewed as standing for a description recoverable in a complex way from the initial part of the sentence. Geach may or may not have called this an example of a "pronoun of laziness"; the term is his, but it has been used by him and others in a variety of ways. What all uses of the term have in common is the idea that some pronouns should be analyzed neither as bound variables nor as directly referential, but in terms of some syntactically definable relation to an antecedent noun phrase.

Another example for which a "pronoun of laziness" treatment has plausibility is (19), from Karttunen (1969):

(19) The man who gives his paycheck to his wife is wiser than the man who gives it to his mistress.

This it is also not a bound variable nor directly referential; it seems to be a substitute for the expression his paycheck. In both Partee (1970) and Partee (1975b). I argued for the existence of a syntactic pronoun-of-laziness rule, intended to cover both these examples and those cases of what I am now calling pragmatic pronouns in which the antecedent is itself a directly referring expression such as a proper noun or a definite description. However, neither I nor anyone else that I know of ever succeeded in stating a version of the rule which covered all of these cases without generating clearly unacceptable results as well. Recent arguments by Terry Parsons (personal communication), Robin Cooper (forthcoming), Gareth Evans (1977), Emmon Bach (personal communication), and others have convinced me that there is no way to make the notion of "pronoun of laziness" coherent without reducing it to one which covers only a small subclass of the pragmatic pronouns and hence does no useful work.

What then can we say about the paycheck sentences and the donkey sentences? Many lines of attack are being explored currently; one that I find particularly promising is proposed by Cooper (forthcoming), who suggests a rather natural extension of the notion of pragmatic pronoun to handle them. Before describing his proposal, I need to fill in some background.

Russell's analysis of singular definite descriptions (Russell 1905) requires that there be a unique object satisfying the description in order for the expression to denote anything, and hence notoriously fails to account for the successful reference of a noun phrase like the clock in (20).

(20) Did you wind the clock?

That the missing ingredient is pragmatic has long been recognized; Cooper (forthcoming) proposes a mechanism that brings in pragmatics in a simple way that parallels the account of pragmatic pronouns given above (which is also basically Cooper's). He proposes for definite descriptions a semantic interpretation like Russell's but with the addition of a free property variable P: the clock then denotes (the property set of) the unique individual x such that clock(x) and P(x). At the

semantic level, P is just a free variable; it is left to the pragmatic interpretation of the sentence in context to determine an appropriate choice for P. In a context where there is no salient distinguishing property, the singular definite description would indeed be inappropriate or uninterpretable. Cooper's treatment can be seen as a formalization of the informal gloss of the (by Katz and others) as "contextually definite"

As a second background step toward Cooper's proposal, consider the interpretation of genitive phrases like that in (21).

(21) John's team lost again.

As is well known, John's team may be the team John owns, or plays for, or roots for, or collects trading cards of, or writes news stories about; there are virtually no limits on the relevant relation. I propose that such constructions be analyzed at the semantic level as definite descriptions containing a free relation variable R, whose value is to be determined at the pragmatic level, by looking for an appropriately salient and relevant relation in the linguistic or non-linguistic context. Thus John's team would be interpreted as (22):

(22) the x such that team (x) and R (John, x).

What is common to these analyses of pragmatic pronouns, definite descriptions, and genitive constructions is the use of semantic free variables that are pragmatically assigned particular values. Introducing the free variables allows a complete specification of the form of the interpretation to be given for each sentence at the semantic level, while providing an appropriate division of labor between semantics and pragmatics in the determination of the content.

Cooper's proposal for the donkey and paycheck sentences is that pronouns can be analyzed not only as free variables, but alternatively as expressions composed of more than one free variable, utilizing free property or relation variables much as in the examples just discussed. The logical formalism is complex, but I will give it for completeness and then try to paraphrase it less formally. A singular pronoun (he, she, or it) may have any translation of the following form:

(23) $\lambda K \exists x [\forall y [[\check{} \pi] (y) \equiv y = x] \wedge K (x)]$, where π is a property-denoting expression containing only free variables and parentheses.

What this says is that e.g. it may be interpreted as (the property set of) the unique individual x which has property π . For the paycheck example, an appropriate π will be R (u), where R is a free relation variable and u is a free individual variable that will be bound by the second occurrence of the man. The second clause of (19) will then say "the man u such that u gives the x such that R (x,u) to u's mistress." The pragmatically appropriate R will be "being the paycheck of". The computational complexity of the analys is justified, I believe, by the fact that only very salient relations permit the kind of pronoun use evidenced by the paycheck example.

Cooper's analysis of the donkey sentences uses the identical device; for details see Cooper (forthcoming).

The conclusion of this section is that there are no pronouns of laziness; the cases which seemed to require them can be handled by an extension of the notion of pragmatic pronouns. The extension is somewhat complex, but (a) it makes use of the same kind of property and relation variables that are needed for an account of definite noun phrases and genitive constructions, and (b) the examples it is needed for are intuitively complex and infrequent in occurrence.

4. Conclusion.

There are many problems of pronouns and reference that I have not touched on. I have not discussed reflexive pronouns, first and second person pronouns, pronouns in modal contexts, the pro-common noun one, anaphoric determiners like same, different, or other, or any of a host of other topics crucial to a fuller account of the role of pronouns in reference. In some cases the problem is just lack of space and time, but in other cases there are still difficult open problems. I hope that some of what I have included is relatively unfamiliar and potentially useful for computational language processing endeavors, and I count on my fellow panelists to fill in some of the holes I have left.

Footnotes

1. There are apparent exceptions to even this weak a statement, but I believe they are best understood as involving elliptical sentences. Consider the following example (from David Kaplan, personal communication):

A: Could a woman become chairman of the Philosophy Department?
B: Yes, if she's qualified.

The she in the second sentence is not a pragmatic pronoun; but I think it is best treated as bound by an unexpressed antecedent within the second sentence, which is not as it stands a complete sentence, rather than as bound by an antecedent in the previous sentence

2. There are exceptions to this statement, too, but they all involve idiomatic pronoun-containing expressions like "shrugged his shoulders" or "lost his cool". Reflexive pronouns are not included in this generalization; they are almost invariably bound variable pronouns, except for certain cases that seem to result from instability in the choice of nominative or accusative form. I will not go into any details about reflexive pronouns here.

3. On the pragmatic pronoun reading, the pronoun he can of course refer to someone other than the prosecutor; in that case the missing verb phrase will always be understood as involving reference to the same third person.

4. There is still an individual/group ambiguity for the subject in this case, but it does not affect the interpretation of the pronoun, so I will ignore it.

5. For simplicity I am ignoring the dialect that allows they with a singular antecedent; in that dialect (12b) is as ambiguous as (13).

6. Not every occurrence of a quantifier phrase with no has this effect, as the following example from Evans (1977) shows:

(i) *John owns no sheep and Harry vaccinates them.

The role of non-linguistic inference in interpreting pragmatic pronouns can be seen from the following linguistically similar examples.

(ii) John owns no sheep because Amherst taxes them.
(iii) John now owns no sheep because Harry poisoned them.

In (ii), them seems to be generic sheep rather than any group of sheep; in (iii) the most plausible interpretation seems to be the sheep that John once owned. Perhaps it would be more accurate to say that no man and n men never serve directly as antecedent to a pragmatic they, but sentences in which they occur do sometimes permit the inference of a suitable referent for a pragmatic they.

7. The group in this case is the group of John and Bill. That group can be put into contextual salience by any mention of John and Bill separately, as in the examples below.

(i) John saw Bill yesterday. They decided to go fishing.
(ii) I invited John, but not Bill. They both came anyway.
(iii) Ask John or Bill. They know where the keys are kept.

8. Montague (1973) treats all noun phrases as devoting property sets and Cooper follows this practice. While that treatment seems essential for a unified account of noun phrases, I have omitted discussion of it here for simplicity.

Bibliography

Bennett, Michael (1974), Some Extensions of a Montague Fragment of English, unpublished Ph.D. dissertation, UCLA. (available through Indiana University Linguistics Club)

Cooper, Robin (forthcoming), "The Interpretation of Pronouns", to appear in Selections from the Third Groningen Round Table, ed. by F. Heny and H. Schnelle, Syntax and Semantics, Vol. 10, Academic Press.

Cooper, Robin and Terence Parsons (1976), "Montague Grammar, Generative Semantics, and Interpretive Semantics", in Partee, ed. (1976).

Evans, Gareth (1977), "Pronouns, Quantifiers, and Relative Clauses (I)", Canadian Journal of Philosophy, 7.3 (467-536).

Geach, Peter T. (1962), Reference and Generality, U. of Cornell Press, Ithaca. (amended ed., 1968)

Jackendoff, Ray (1972), Semantic Interpretation in Generative Grammar, The MIT Press, Cambridge.

Karttunen, Lauri (1969), "Pronouns and Variables" in Binnick et al, eds., Proc. of 5 Chicago Linguistics Society Meeting, 108-116.

Karttunen, Lauri (1976), "Discourse Referents", in J. McCawley, ed., Syntax and Semantics Vol. 7, Academic Press, New York.

Keenan, Edward (1971), "Names, Quantifiers, and a Solution to the Sloppy Identity Problem", in Papers in Linguistics 4.2.

Klima, Edward S. (1964), "Negation in English", in Fodor and Katz, The Structure of Language: Readings in the Philosophy of Language, Prentice-Hall, Englewood Cliffs, N.J., 246-323.

Lakoff, George (1971), "On Generative Semantics" in D. Steinberg and L. Jakobovits, eds., Semantics, Cambridge, 232-296.

Lees, R.B. and E.S. Klima (1963), "Rules for English Pronominalization", Language 39, 17-28.

Montague, Richard (1973), "The Proper Treatment of Quantification in Ordinary English", in K.J.J. Hintikka, J.M.E. Moravcsik, and P Suppes, eds., Approaches to Natural Language: Proceedings of the 1970 Stanford Workshop on Grammar and Semantics, D. Reidel, Dordrecht; reprinted in Montague (1974).

Montague, Richard (1974), Formal Philosophy; Selected Papers of Richard Montague, edited and with an introduction by Richmond Thomason, Yale University Press, New Haven.

Partee, B. (1970), "Opacity, Coreference, and Pronouns", Synthese 21, 359-385.

Partee, B. (1975a), "Montague Grammar and Transformational Grammar", Linguistic Inquiry 6,203-300.

Partee, B. (1975b), "Deletion and Variable Binding", in Edward Klima, ed. (1975) Formal Semantics of Natural Language, Cambridge University Press, Cambridge.

Partee, B, ed. (1976), Montague Grammar, Academic Press, New York.

Quine, W.V.O.(1970), Philosophy of Logic, Prentice-Hall.

Ross, J.R. (1967), Constraints on Variables in Syntax, unpublished doctoral dissertation, MIT, Cambridge, Mass.

Russell, Bertrand (1905), "On Denoting", Mind 14.

Sag, Ivan (1976), Deletion and Logical Form, MIT Ph.D. dissertation, available through Indiana University Linguistics Club.

Williams, Edwin (1977), "Discourse and Logical Form", Linguistic Inquiry 8.1.

The Use of Focus as a Tool for Disambiguation of Definite Noun Phrases

by

Candace L. Sidner
MIT Artificial Intelligence Laboratory .
Cambridge, Massachusetts

1. Introduction to the Problem

When speakers[1] utter or write sentences, they use certain words in the sentence to refer to people, places, object, times, events and ideas which exist in the real world. When sentences are formed into units of two or more sentences, certain words refer back to other referring expressions in the previous sentences. Among the words which can be used to refer to the real world as well as to refer back (which is called the process of co-reference) are noun phrases containing a definite article, such as *the* (called defnps, hereafter). Several aspects of comprehension of defnps are open problems:

1.) What is the definition of the reference of a defnp? That is, what to we mean by *reference* in computational linguisitics?
2.) How are defnps which are used to co-refer into a discourse distinguished from those which refer to real world objects outside the discourse?
3.) What methods of search will distinguish the referent of a defnp which refers to an entity outside the discourse context?
4.) What different ways can defnps be used to co-refer to other entities in the discourse?
5.) How can co-reference of defnps be detected?
6.) What inferences and data structures will be needed for the detection process? The work of Winograd [1971], Charniak [1972] and Rieger [1973] suggest that inferencing is crucial to the interpretation of co-reference.

This paper presents a viewpoint from which to answer these questions based on the concept of focus, as developed by Grosz [1977] and the author (Bullwinkle [1977]). This paper extends Grosz' work by establishing a framework for communication and a set of rules for use of focus in discourse. The approach taken here represents an alternative to the inference driven schemes of Charniak and Rieger.

1. I will use the term *speaker* to refer to the producer of a spoken or written discourse and *hearer* to refer to the receiver of the discourse.

2. The Communication Process and Focus

The description of the communication process given here contains four simple assumptions which are generally true and will be taken as true in this work. First, the speaker is assumed to be communicating about something. This assumption implies that the speaker is not speaking gibberish, that there are referring expressions and either requests, questions, assertions or acknowledgements being made. The something which the communication is about will be called the focus of the discourse.[2] Second, the hearer is assumed to be able to identify what the focus of the discourse is. The speaker wants to communicate about something, and for the communication to occur, the hearer must be able to distinguish what the speaker is communicating about. Third, the speaker is not trying to confuse or deceive the hearer. The speaker uses referring expressions with the intention of referring to someone or something, or with the intention of describing something or some event. In Gricean[3] terms, the byword is "Be perspicuous." The final assumption claims that the speaker assumes the hearer knows certain knowledge about the real-world which can be referred to during the communication process. Recent research (Cohen [1978]), as well as the well known work of Searle [1969] and Austin [1962], describe models of the speaker's knowledge of what the hearer believes. In this chapter, the weakest form of such a model is assumed: the speaker assumes the hearer has enough real-world knowledge in common with the speaker to know the entities which the speaker refers to, and that knowledge is what the speaker draws on in constructing a message for a hearer. These four assumptions will play an important part in the discussion of co-reference interpretation which follows.

2. I don't want to suggest that only one thing can be communicated in a discourse, for speakers do direct their attention from one thing to another. For the moment, I will speak of the focus as the first center of attention in a speaker's discourse.
3. Grice, H.P. "Logic and Conversation" etc.

This paper makes the claim that the focus acts as an index function for referring expressions. For those referring expressions which are anaphoric, the focus indicates where to look for an antecendent. For those referring expressions which are names or descriptions of things in the world, the focus acts as a generation center for a process that chooses a representation of a real world entity which fits the name or description. However, the focus of a discourse alone is not sufficient to produce the indexing behavior. The focus must be used in conjunction with a hierarchical semantic network of associations. The network will indicate what other concepts are related to the focus. It is a codification of some of the general knowledge speakers and hearers have about the real world. The network is a dynamic structure because the hearer adds to his/her general knowledge in the process of interpreting a piece of discourse. Focus must also act with a third piece of computational machinery, an inferencing mechanism. It is used to infer from general knowledge and some suppositions that a certain proposition is true.

An example will be helpful here. In the discourse below, the focus of discussion is the meeting of D0-1.

D0-1 I want to schedule a meeting with Ira.
2 It should be at 3 p.m.
3 We can get together in his office.
4 Invite John to come, too.

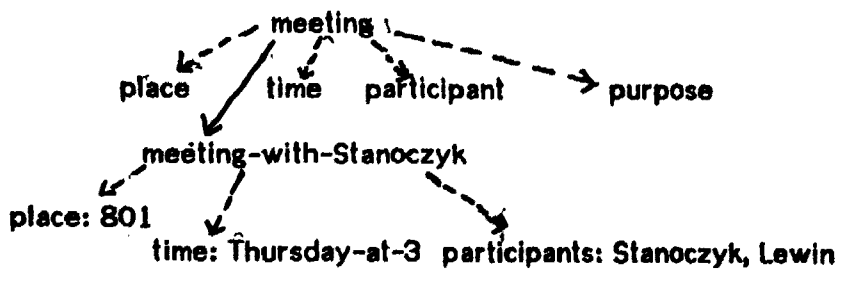
All four sentences give information about the focussed entity. Thus in D0 both sentences 3 and 4 make no direct reference to the meeting of D0-1. As human hearers, we know that these sentences are related to the rest of D0 because they give information about the focus *meeting*. In D0-3 there are three clues which connect this sentence and the rest of the discourse: the use of *get together*, the co-reference of *we* to the participants of the meeting, and *his office* establishing a place for a meeting. D0-4 introduces an additional participant which can be surmised from the use of *invite*, and the fact that the ellipsis of the event that John is invited to is the focus.

A piece of the hierarchical net needed for D0 is given below in figure 1. A prototype meetings has associated places, times, participants, and purposes. The relation between meeting and place is one of occurrence while the relation between meeting and purposes is one of causality. When D0-1 is encountered, the hierarchical net grows a new member: an instance of meeting from D0-1. It inherits the associated entities of meetings and some specific values for the participant entity. D0-2 indicates that something (called *it*) will occur at a particular time. The focus of D0-1 is meeting, so the focus, meeting, is proposed as the antecedent of *it*. To confirm the proposal, the inference mechanism checks to see if meetings occur at times. Indeed they do, so the proposal of meeting as antecedent of *it* is accepted.

The explanation about the use of focus is not really so simple because the focus of a discourse changes. The interpretation of focus requires a means of recognizing that the focus has changed to some other entity. In D0' the focus begins on meeting, but the *it* in D0'-3 has *my office* as its co-referent, not the meeting. Detecting this co-reference requires a means of noticing a shift of focus and using the inferencing mechanism to confirm the choice of co-referent. Focus shift detection will not be discussed here; the reader is referred to Bullwinkle [1977] for a discussion of focus shift where the term "sub-topic shift" is used.

D0'-1 I want to schedule a meeting with George, Jim, Steve and Mike.
2 We can meet in my office.
3 It's kind of small, but the meeting won't last very long anyway.

Fig. 1. Instances of a General Meeting Concept



3. Reference in Computational Terms

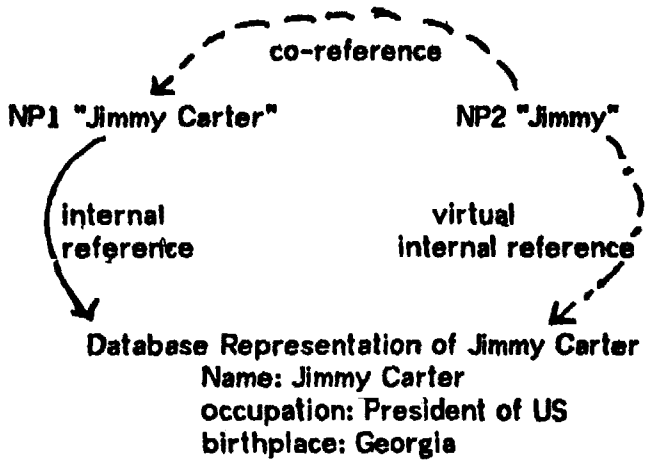
The theory presented here distinguishes two kinds of referring. The first is an internal reference between a noun phrase and some pre-existing database object. That database object represents a real world entity. In Figure 2 below internal reference links the noun phrase NP1 "Jimmy Carter" to a representation of Jimmy Carter (who is described as president of the US, etc.). How the noun phrase and the database object refer to the real world is the classical semantic problem of reference (cf. Kripke [1972] among others) and is beyond the scope of this work. The other kind of referring is co-reference. Co-reference links a noun phrase to another noun phrase. The two noun phrases are said to co-refer, and both internally refer to the same database object, both refer to the real world object. In Figure 1, the dashed link from NP2 "Jimmy" to NP1 is a co-reference link. The dot-dash link from NP2 to the database object is a virtual internal reference link which results from the co-reference link from NP2 to NP1 and from the internal reference link from NP1 to the database object. Internal reference and co-reference links are distinguished because co-reference links can be established more easily using discourse context, which will be discussed in detail later in this paper. In the remainder of this paper when I speak of internal reference, I will drop the phrase "internal" and use only "reference."

A computational theory of co-reference comprehension must answer the following questions about the use of referential terms in natural language:

(1) Does the expression refer to someone or something?
(2) What conceptual entity in the memory or the database of the hearer's knowledge, if any, is denoted by the referring expression?
(3) When does a given expression refer to the same entity as another referring expression?

The expression *Julius Caesar* is used to refer, and can refer to the person represented in the hearer's knowledge as Julius

Fig. 2. Reference Links Between Noun Phrases



Caesar. To answer the first question above, the hearer must decide that names are referring expressions. To answer the second question, the hearer must decide 1) whether *Julius Caesar* refers uniquely and 2) what conceptual entity in the hearer's memory represents the hearer's real-world referent. These two decisions together with the initial assumptions appear to make necessary and sufficient conditions for comprehension since by deciding that *Julius Caesar* refers uniquely and choosing a conceptual entity, the hearer has decoded what entity the speaker was referring to.

There are, however, situations where the hearer's choices to the above decisions and the speaker's intended referent do not coincide. Suppose the hearer decides that *Julius Caesar* refers uniquely and refers to Julius Caesar, who was a Roman emperor. The speaker may also have intended it to refer uniquely, but to the author's deceased cat, whose name was Julius Caesar. Now there are three possibilities: either the hearer knew about Julius Caesar the cat, but decided the expression referred to Julius the emperor, or the hearer only knew about the emperor, or the hearer didn't know of either. In the last case, the hearer "found" a referent by a chance from randomly linking up the name and some memory representation. The last possibility does not fit a description of reference comprehension of any kind. Randomly hooking up information from one's memory to what appears to be a referring expression may be a cognitive act, but intuitively no one would call it reference comprehension.

In the case where the hearer only knew about the emperor, it seems safe to conclude that the reference may have been comprehended, but *incompletely.* As we shall see, there are many other clues in communication about the referent of terms than those given by referring expressions in isolation. Without these, reference comprehension is incomplete because the hearer has no means of knowing whether s/he may have the wrong referent. Even with the best set of clues, the hearer may still choose Julius the emperor. Here we will say that comprehension has taken place, completely but incorrectly, because the hearer has used all the relevant communication knowledge to decode the speaker's message. What can be concluded is that the speaker's rules for reference generation and/or the speaker's knowledge of the hearer is faulty (thereby contradicting the speaker's assumption above).

In the case where the hearer knows of both possibilities and chooses the incorrect one, the hearer may have erred due to failure to follow other communication clues or again because the speaker's rules and knowledge were lacking. In conclusion, a referring expression is *comprehended as intended,* if and only if the same referent as that intended by the speaker is chosen from the entities in memory. The expression is otherwise just *comprehended* when the hearer chooses an entity from memory which is denoted by the referring expression using all the available communication clues but does not choose the same entity as intended by the speaker. An expression will be considered *incompletely comprehended* if the hearer fails to use all the communication cues available at the time the communication occurs.

So far I have not considered the possibility of error on the part of the hearer because of the hearer's beliefs. Suppose, for example, that the hearer believes the speaker hates to even speak of cats. Then the hearer may conclude that *Julius Caesar* is most likely a reference to the emperor of Rome. I am not going to consider this possibility in the forthcoming discussion; instead I will restrict the discussion to cues from the communication process. Hearer beliefs raise a separate set of philosophical as well as computational problems and entends the scope of this study too broadly. However, the issues are significant in the total picture of reference and co-reference comprehension.

In the remainder of this paper I will consider co-reference comprehension just from the hearer's point of view. Thus in discussing referential and co-referential expressions, I will be concerned with a model of how the hearer disambiguates these expressions used in discourse. By symmetry, one might suppose that the generation of referential expressions by a speaker could make use of a similar model. Such a supposition will remain untested in this paper and is to be verified by later work. Furthermore, I will not be concerned with comprehension as intended since this process requires the additional information of what the hearer believes that the speaker knows about. Instead I will point out at various times how the theory under discussion would need modification if hearer's beliefs were included.

4. Problems with Definite Noun Phrases

Definite noun phrases can be used to refer to entities in the real world. Russell [1905] says of the expression *the author of Waverly* that it denotes Sir Walter Scott, and that when it is strictly used, a defnp denotes uniquely.[4] Thus by using a definite article, a speaker is saying in effect "there is one object in the world denoted by the phrase that follows and I mean that one." Of course a defnp may be *used* to denote someone without *actually* denoting anyone, as is the case with *the woman who wrote Waverly.* This defnp is used to refer to someone, but there is no conceptual representation in the hearer's (or for that matter, the speaker's) memory which corresponds to a real world

[4]. By *strictly used* Russell means used without ambiguity.

entity assuming the normal case.[5] Nothing in the syntactic or semantic form of the expression itself suggests that the expression has no denotation. How can the hearer determine whether the defnp refers to someone or not? Of course, if there exists a memory entity the author of Waverly, which is attributed as male, the hearer can decide that the expression does not refer to anyone on the basis of a contradiction. But if no memory entity exists, the hearer cannot decide whether *the woman who wrote Waverly* refers to anyone. This way of looking at defnps, however, fails to account for all the phenomena of defnps because it involves an assumption which is not true.

The Russellian analysis has difficulties because defnps are not always used to refer! The problem is not only whether a particular defnp actually denotes a real world object; it is also a question of whether the defnps is intended to refer at all. Even more surprising, a defnp may be used to refer, but the speaker may not intend for the hearer to know the referent of the defnp; the defnp form is used to indicate that the referent is knowable, but possibly not significant for the communication at hand. Donnellan [1977] points out that some defnps are used attributively. If we happen upon Smith who lies dead with foul wounds, one can say "Smith's murderer is insane." Used attributively, *Smith's murderer* does not refer to anyone, and the phrase does not describe a particular person. It is as if to say, Smith was murdered and the murderer, whoever that may be, is insane. Thus the speaker using an attributive defnp does not assume that someone fits the description, whereas with a referential defnp the speaker expects the hearer to realize who is being pointed to.

The other distinction a speaker can make is to use a defnp to indicate that the referent is knowable. Thus if one says:

(1) Larry read a lot of linguistics in the hospital.
(2) Larry read a lot of linguistics in a hospital.

the (2) usage is not the same as the (1). While the hearer does not know which hospital *the hospital* refers to, it is clear it refers to some particular one. Comprehension of the referential term does not involve finding a memory entity which represents the real world entity that the expression refers to. For reference comprehension, this concern is considerable since *the Chinese government* in (3) does not demand reference disambiguation, while (4) does; the disambiguation is difficult because the expression can refer to more than one thing.

(3) John got help from the Chinese government in adopting an Oriental child.
(4) Get a visa for your trip from the Chinese government.

Another difficulty with defnps is that sometimes they are used not to refer to or to describe specific individuals or objects, but to characterize a class of entities with the properties of the head noun phrase and any of its modifiers. Thus (e1) used in this way does not refer to an individual. It characterizes a member of the class of individuals who are men and book writers. (e1) is similar to attribution except that the description applies to a class.

(e1) the man who writes books

So far, then, the following classes of defnps can be stated. A defnp that is used to refer uniquely to one entity, whether or not such an entity exists in the real world, is a specific defnp. A defnp that characterizes a class of entities by means of an individual whose properties are delineated by the properties of the head noun phrase and its modifiers is a generic defnp. A defnp is attributive if it describes an entity without referring. A defnp can be ambiguous in use (u-ambiguous hereafter) if its use as a specific, attributive or generic is not identifiable, while a defnp is ambiguous in reference (r-ambiguous hereafter) if it is used specifically and there is more than one entity fitting the description of the defnp.

5. General Role of Context in Disambiguation

Little of language, if any at all, is said without some surrounding contexts of information. For example, most conversations happen in a location where there are other objects present. Most stories have at least the context of there being a story teller, a hearer and the story being told. There are contexts with more presumed common knowledge, such as what the hearer knows of the speaker's own identity or some shared additional information between them (e.g. they have children or parents in common).

Contexts are needed to determine what a defnp refers to. If I say (5), when I am standing in my kitchen with a friend, the defnp, if specific, must refer to some unique object in the world.

(5) Get me the hot dish holder.

There may be lots of hot dish holders denoted by conceptual entities in my friend's mind, but I am referring to a specific one. Since nothing in (5) distinguishes the one I mean from the whole collection, either I have misused the language, or there is a context which contains only one such hot dish holder, and my friend is aware of that context at the time of my saying (5). In this case, the necessary context is the kitchen, and the referent is probably an item in the kitchen. Reference made to an object external to the conversion in called extra-sentential reference. It is discussed here to exemplify the role of one class of contexts used in reference determination. I call contexts of reference which exist in additional to the one created by the discourse implicit contexts. In this paper I will show how use of implicit contexts can avoid the problem of searching a general database for the entities denoted by defnps.

Other defnps make use of different implicit contexts. Instead of an implicit context consisting of objects near the speaker, the implicit context may be events that the speaker believes are common to the hearer. The speaker who opens a dialogue with (6) below is assuming some previous context (a discussion with the hearer or some other situation) where the reference of *the A.I. Lab Language Group* was first established. In (7), the speaker is again assuming a pre-established referent, but since the hearer may know of several different

---

5. Possible world semantics will not be discussed here. Issues of transworld identity and designation by definite descriptions may require more machinery than is considered here.

dogs, some specific context must be chosen that will distinguish a single dog. Later in this paper some heurisitics for choosing a context will be discussed.

(6) The A.I. Lab Language Group wants to meet next week.
(7) The dog is sick again.

Contextual information of yet another kind appears in story telling. At the beginning of a story, the hearer expects characters to be introduced. Sometimes this is done with indefinite noun phrases, which are a way all discourses introduce new items, but often a story-teller uses names or defnps as (8) below shows.

(8) The heiress lived the life of a recluse. She died under mysterious circumstances, but the murderer was never found.

(8) is not a case of cataphoric referencing (referring forward in a text) since the phrase *the heiress* can fully specify an object itself. However, hearers of (8) do not have to search their memories for a referent to *the heiress* in (8). They use the context of story beginning to guide them in reference disambiguation.

6. Distinguishing Generics

Defnps must be disambiguated as generic or non-generic. As will be shown later, generics in the midst of a discourse can be easily disambiguated, but in an initial sentence only implicit contexts may exist in which a co-referent can be found to use in disambiguation. Implicit contexts may be helpful in some cases, but in general they are not sufficient to indicate the interpretation. However certain rules can be postulated based on observable sentence data. This data indicates that there are several levels of sentential and phrasal information used for disambiguation. The rules which will be summarized[6] here give preferences for generic and non-generic readings.

Some rules govern whether the defnp itself is preferred as a generic or non-generic reading. A small collection of phrases like *the sun, the moon* and *the president* default to specific, well defined entities. Which entities are defaulted to depends on the presence or absence of an implicit context to which the phrase may co-refer. For other defnps, a "yes" answer to (1) of the following questions indicates a specific reading preference, while a "yes" to (2) and (3) indicates a generic.

1.) Is there a specific individual so described?
2.) Is there a class use acceptable for this pair?
3) Are there many individuals described but none outstanding?

"No" answers to all causes a preferred specific reading. Question (2) is necessary because generic readings are difficult to obtain for certain classes of entities. Thus color words like *black, red, yellow* and *white* applied to *man* describe a class specification while the other color words indicate a non-generic description.

In addition to phrasal preferences, predicate argument relations for certain verbs may indicate a preferred non-generic reading. In these cases, a u-ambiguous defnp will be taken as a non-generic, and a generic defnp will cause an odd sentence.

(9) The black man was moving towards the window.
(10) The woman who reads *Total Woman* is coming to dinner on Saturday.

Some classes of speech acts are also distinctly generic. The *is-a* sentence below is always generic; this reading may result from the use of *is-a* to indicate further characteristics of the subject. The announcement speech act in (12) is generic as long as an implicit context does not exist which contains an acceptable co-reference for the defnp.

(11) The elephant is a large mammal.
(12) I want to tell you about the orangetang.

Speaker-hearer assumptions about perspicuity can force a reading to be generic or non-generic. The defnp in (13) is forced to be read generically because a specific reading would be r-ambiguous and therefore not perspicuous. On the other hand, (14) is odd since *invite* requires a non-generic object. However, because the speaker is assumed to be maintaining perspicuity, the hearer may attempt to read the defnp in (14) as a non-generic.

(13) Bill considers the black man to be the source of Boston's social unrest.
(14) Invite the man who reads *The Grapes of Wrath* to dinner.

I want to emphasize that the noun phrase, verb phrase and sentence level rules are only preferences for readings. In the worst case, as (15) shows, an initial sentence may contain a u-ambiguous phrase which, while preferred as generic, can be used either way.

(15) The robot is replacing the car.

7 . The Explicit Backwards Co-reference Rule for Defnp Disambiguation

Using the concept of focus, rules governing the co-referential use of defnps in discourse can be stated. The rules for defnp co-reference which follow depend upon the ability of the hearer to identify focus. This process is a complex one and will not be discussed here. The reader is referred to Sidner [forthcoming] for full details. In brief, the focus of a sentence depends upon predicate argument relations and in some cases, special syntactic forms, such as clefts and pseudoclefts.

In the simplest formulation, the rules for defnp co-reference states: the discourse focus provides a reference point for the co-references of defnps. As I will show below, the rule contains several subparts which must be stated separately. In this paper I will refer to cases of a defnp used anaphorically as explicit backwards co-referencing (EBC). The EBC rule states that a defnp with the same noun phrase head as the focus, and which appears in a sentence following the sentence with the focus, is co-referential to the focus. The more common forms of explicit backwards co-referencing are found in D1 and D2 below:

6. See Sidner [forthcoming] for a full specification of the interpretation rules.

D1-1 I want to have a big party$_i$ with lots of guests.
2 The party$_i$ ought to be on Saturday so everyone can come.

D2-1 I'm going to tell you about the elephant$_i$.
2 The elephant$_i$ is the largest of the jungle mammals.
3 He$_i$ weighs over 3000 pounds.
4 At one point in its$_i$ existence, the elephant$_i$ had to protect itself from the lion,
5 but now its$_i$ herds are so large, that most lions won't even venture near.

What the reader will notice about D2 is not only the co-referenciality of focus for the second and third uses of *the elephant*, but also the co-referenciality implies that these uses are generic. Where defnps in isolation are often ambiguous on the generic-specific classification, in discourse context, this rarely occurs since the focus provides the class type for the defnp. As stated, the EBC rule makes a true prediction about u-ambiguous defnps which occur in sentences following the focus: they are co-referential with the focus, and hence disambiguated as non-generic. D3-2 below contains a defnp which is u-ambiguous in isolation, but in the discourse context, it refers to George's elephant, the reference of the focus.

D3-1 I sent George an elephant last year for a birthday present.
2 The elephant likes potatoes for breakfast.

The EBC rule is inaccurate when applied to strictly generic defnps, and where it fails, the role of phrasal and sentential level processing in co-reference comprehension is indicated. D4 is an indication of the problem. D4-2 is generic in isolation. Even in the context of D4, where the focus is Mary's ferret, hearers interpret the underlined defnp as generic.

D4-1 Mary got a ferret for Christmas last year.
2 The ferret is a very rare animal.

The context cues of discourse are not strong enough to reverse a strongly generic reading of a defnp. In order for this to be so, sentential level processing must have occurred without consideration of the demands of the context. Since the EBC rule as stated predicts co-reference in cases like D4, it must be revised: specific and u-ambiguous defnps which contain the same noun phrase head as the focus, and which follow the focus in the discourse, co-refer with the focus.

A further refinement on the EBC rule is needed. Consider *the fairy tale book* in D5-2. The EBC rules predicts it will be co-referential with the focus of book in D5-1. In fact, English speakers find D5-2 an odd sentence in the discourse because it is not clear what *the fairy tale book* has to do with the rest of D5.

D5-1 I bought a book today.
2 The fairy tale book is by the Brothers Grimm.
3 It is really well illustrated.

It seems that defnps which co-refer with the focus canno contain anymore information than is known about the focus. Thus one could say following D5-1. "The book I bought is a fairy tale book by the Brothers Grimm" (since D5-1 states that the speaker bought the book), but one cannot say D5-2. Why can't a defnp that contains more information than the focus co-refer to the focus? Returning to the discussion of focus-shift earlier, a referring expression following the focus is either co-referential to the focus or introduces an entity which is the potential new focus of the discourse. The difficulty with phrases like *the fairy tale book* is that one cannot tell if it is intended to co-refer, or because it is somewhat different from the focus, intended to be used as a potential new focus. The EBC rule must be revised to state: specific and u-ambiguous defnps which contain the same noun phrase head as the focus, which follow the focus in the discourse, and which do not contain more information than is known about the focus co-refer with the focus.

The EBC rule explains why a negative existential cannot be referred to using a defnp. A sample case, from Karttunen [1968], is given in D6. D6-2 is generally regarded as an unacceptable sentence following D6-1. The sentence is certainly grammatical, so the assumption by Karttunen is that the referential term *the car* is being used in some inappropriate manner.

D6-1 I don't have a car.
2 * The car is black.

The EBC rule predicts that *the car* co-refers with the focus in D6-1. But *a car* in that sentence does not have a referent (because the speaker has just said so). Thus the use of the defnp in D6-2 causes the hearer to expect a reference when in fact there is no referential entity.

A similar case, (16), also from Karttunen, does not involve negative existentials, but entities within modal contexts:

(16) * Mary expected a present from John although the present was expensive.
(17) Mary expected a present from John although the present wasn't the thing that worried her.

The defnp in (16) according to the EBC rule must co-refer with the focus. What is significant is that the co-reference is acceptable, as (17) shows. What is odd about the second clause of (16) is the predication. This paper cannot give an account of such semantics, but intuitively, it seems odd to predicate the property of being expensive to something one expects. Thus as long as there is a co-referen. entity specified by the focus, a defnp may be used, but the predication about the defnp must be semantically meaningful.

Another form of explicit backwards co-referencing is slightly different than the previous examples. It involves the use of lexical generalization of the focus. Grosz [1977] first categorized the relation of focus and defnps with a more general noun in the noun phrase head. In D7, *the poor old beast* is a lexical generalization of *the dog*, that is, its head noun is a term which is a class generalization of the focus.[7]

7. This term comes from the observation of Halliday and Hasan [1976] that lexical cohesion includes the use of reiteration of four types: same word; synonym, superordinate, and general word.

Determining the class generalization of the focus is possible when the focus is represented in the way that is assumed in this paper: as an association network with an is-a hierarchical structure. Using that hierarchy, it is possible to determine whether a phrase like *beast* is hierarchically related to *Salamut*.

D7-1 Harold took his Salamut$_i$ to the vet yesterday.
2 The poor old beast$_i$ was quite lame.

One might expect that some constraint on the amount of information in the lexical generalization of the focus is needed. This is the case, since the underlined defnp in D7-2' is unacceptable following D7-1 as a co-referent with the focus.

D7-2' The beast who is old was quite lame.
-2" The mangy, snarling, unfriendly beast was quite lame.

It appears from all the cases I can find that any post-nominal modifiers on a noun which is a lexical generalization of the focus force the defnp to be non-coreferential with the focus, while pre-nominal modifiers, no matter how complex, preserve co-referentiality. It is unclear why pre-nominal modifiers and post-nominal modifiers have these different behaviors.

8. Implicit Backwards Co-reference

Many definite noun phrases which occur in discourse are not cases of backwards co-reference to the focus. Grosz [1977] suggested that the focus brings other items implicitly into focus as well, by means of association. Such defnps are related to the focus in one of several ways. Since the focus is well specified, these relationships can be easily determined. The focus acts as an anchor point for finding references for such defnps. In D8, the defnp *the time* refers to the time of the discourse focus, the meeting. This defnp use I will call implicit backwards co-reference. Such cases are to be distinguished from explicit backwards co-referencing because the defnp is co-referential with an entity that is closely associated with the focus rather than to the focus itself. The phenomenon of association between two noun phrases has been cited by Norman and Rumelhart [1975].

D8-1 The pa group wants to have a meeting.
2 The time will be 3 p.m. on Tuesday.

Implicit backwards co-referencing is constrained by the association network surrounding the focus. Any entity closely associated with the entity which represents the focus can be mentioned using a simple defnp. Thus in D9, sentences with acceptable defnps as well as ones with unacceptable defnps are given.

D9-1 I went to a new restaurant with Sam.
2 The waittress was nasty.
3 The food was great.
4 The soup was salty, but the wine was good.
5 * The rug was ugly.

Non-simple defnps have infinitely more variety because the modifiers can specify the relation of the defnp to the focus at hand as in D9-6. Non-simple defnps which do not suggest some connection are less acceptable, but hearers, in reliance on the perspicuity maxim, may attempt some connection. Thus if D9 included D9-7 below, some hearers might attempt to connect the defnp with focus.

D9-6 I like the band that plays there.
-7 The elephant with the green tutu danced an impressive jig.

Another use of focus is as an inference point for inferred co-references. Inferred co-references, like *the murderer* in (8), presented here as D10, are not mentioned explicitly in the previous discourse nor can they be considered closely associated to the focus on general principles. Their use reflects an inference about the focus on the part of the speaker.

D10-1 The heiress lived the life of a recluse.
2 She died under mysterious circumstances, but the murderer was never found.

In D6, *the murderer* represents an inference that the heiress death was due to a specific type of circumstance, a murder. Such an inference is possible given a Fahlman [1977] type net with two inference points like heiress and murderer (and the information associated to heiress from the context thus far); from the net, the relation of the murderer to the heiress can be inferred. Such an inference does not produce a real-world entity to which *the murderer* refers. Instead the inferred relation of murderer and heiress provides sufficient information to produce the entity if it exists in the database. When a denotation does not exist in the database, the inference between the murderer and the heiress sugests that the speaker is attributing of some individual that s/he is a murderer.

A concrete example will illustrate my point. Suppose the hearer knows that the heiress was killed by Jones. Then on hearing D10, the hearer not only concludes that *the murderer* refers to the murderer of the heiress, but also based on that conclusion, the hearer decides that Jones is denoted by the referring expression. However, another hearer upon hearing D10 and not knowing what the first hearer knew, could only conclude that *murderer* is attributed of a person who is assumed to have murdered the heiress. The referent is not known to the second hearer, but if someone were to tell him/her that Jones murdered the heiress, the hearer could conclude who *the murderer* refers to. In effect, the defnp used in this way points out the attributional use of expressions which Donnellen has observed. The argument presented here is not only about the nature of focus; it is a statement of what information is sufficient to make up a description which can denote a unique entity. Viewing inferred co-reference defnps as attributions has an implication for a computational model which disambiguates such defnps. This model must be able to use an expression without knowing its referent and be able to link up the denotation to the referring expression if some knowledge makes that denotation available at a later point.

Implicit and inferred co-reference at first glance appear to be one in the same thing. The discourse below, from Karttunen [1968], will indicate just how the two differ.

D11-1 I was driving on the freeway the other day.
2 Suddenly the engine began to make a funny noise.

3 I stopped the car.
4 When I opened the hood, I saw that the radiator was boiling over.

With focus of *freeway* in D11-1, the relation of the engine can be found since vehicles are driven on freeways and vehicles have engines. The association chain here suggests that the connection between D11-1 and *the engine* involves a few inferences. These inferences are part of a hearer's general knowledge and true of the world. They are part of the knowledge in the association network. With D10, however, the inference about the murderer involves a supposition which is not necessarily true, since dying under mysterious circumstances does not necessarily imply murder. The distinction between implicit and inferred co-reference can thus be stated: implicit co-reference involves inferences which are true about the world, while inferred co-reference involves a supposition which the speaker has made which is not necessarily true.

Another kind of implicit focus-defnp relation exists in D12. I call this relation the set-element relation since *the clown with a unicycle* is an element of the set of clowns which the focus denotes.

D12-1a I went downtown today,
1b and there were clowns performing in the square.
2 The clown with a unicycle did this really fantastic stunt.

As with inferred references, the focus does not make it possible to identify a specific denotation with the referring expression. Instead the focus is the set of which the referent of that phrase is a member. These cases are easier to distinguish than those of inferred reference because the head noun is the singular of the noun phrase represented in the focus Unlike defnps using the EBC rule, set-element co-reference demands a modifier that distinguishes it from the focus. Without the modifier, there is no means of determining which member of the set is being discussed. Grosz [1977] says of cases like the set-element relation that an inference is needed to establish additional properties of an object in focus. However, because the head noun phrase is the same as the focus, while the modifier is different the relation can be established without the need for inference.

Another kind of focus relation, which I call computed reference, can be seen in D13. Here *the last meeting* does not refer to the meeting mentioned in the previous sentence, but that meeting can be used as a point for determining a last meeting, if one is known in the database, else it is a description of the entity required, as with inferred references and set-element references. Several modifiers - *first, last, next, second* and the other ordinals - are used in this way.

D13-1 Aunt Het's Sewing Bee wants to have a meeting this week.
2 The meeting should be on Tuesday.
3 The last meeting, which was at 5, was too late, so schedule this one earlier.

From these examples, the nature of focus in discourse can be re-emphasized. It is the focus which connects sentences of the discourse. In the process of determining the focus relation between a defnp and the focus, the link in the discourse is created. From these examples, one can predict that there ought to be cases of defnps which bear none of these relations to the focus and which result in unacceptable discourses. This prediction is accurate as D12' shows.

D12'-1a I went downtown today,
1b and there were clowns performing in the square.
2 * I saw the chair

The difficulty with D12' is that a chair is not associated with clowns, and the discourse does not suggest any suppositions that would link chairs and clowns inferentially

There are, however, cases of defnps which do not bear any relation to the focus and which are perfectly acceptable in discourse. Consider *the chairman of the math department* in D14 below.

D14-1 George wants to have a seminar to discuss representation in frame-like languages.
2 He wants to invite the chairman of the math department.

The focus in D14 is the seminar of D14-1. The math department chairman is not directly related to the seminar. However, the focus does provide an important piece of information; it is the source of the ellided event to which the chairman is invited. Hence D14 is different from D9-6 or D12' where there is no link between the focus and the sentence in question. What can be concluded is that the focus is not a useful reference point for the referent of the chairman of the math department. The sentence is not odd because of focus links. Thus something outside of the discourse context must contain the needed denotation. This conclusion points to the limitation of the focus: it captures only those reference relations which are internal to the discourse. In a sense, the focus is a summary of the discourse context and what the hearer knows about it.

How can the denotation of *the chairman of the math department* be found? Since the denotation of *the chairman of the math department* lies outside the discourse context, a more global context such as that of the speaker's situation in time and space must be used to determine a context of possible referents. This context must be limited because there are potentially many math department chairmen in the speaker's and hearer's memories. I do not intend to describe just what such a context will look like, but I do want to indicate that it may be "grown" from a search through the associative network to other entities which are related to any of the objects in the discourse, including the speaker and hearer. The association net includes not only abstract representations of general classes of real world entities, it also contains representations of real world objects. The associations between real world objects can be gathered by a search method which collects associations close to the focus and then extends for other associations until one is found that matches the defnp in question. Of course, it is possible that no entity will be found: such a circumstance is yet another example of the hearer knowing that a defnp refers without being able to tell who the

speaker had intended as the reference. The implications of extra-discourse contexts for computational models is clear: models are needed of what the speaker assumes his/her hearer knows about, so that the speaker does not produce referring expressions which the hearer cannot disambiguate, and models are needed of what the hearer assumes the speaker has beliefs about so that the hearer can decide what to do with referring expressions which s/he cannot disambiguate.

Earlier in this paper I mentioned the use of such defnps as *the sun, the moon,* and *the Earth.* These defnps have default referents in initial sentences of a discourse. Inside a discourse, they can act in one of two ways: related to the focus as either a backwards co-reference, implicit co-reference, inferred reference, set-element reference or computed reference; or they may act as references to entities outside the context. The EBC rule predicts that such defnps will be taken as co-referring when the focus has the same defnp head. Thus a defnp like *the sun* or *the moon* will refer to its default value only if the focus does not predict a referent based on one of the five co-reference relations discussed here. These kinds of defnps are especially noteworthy because they are a clear example of a defnp that can be used in either role.

The four related co-reference relations specify ways in which a defnp can be predicted as co-reference to an entity associated with the focus. Other defnps either refer to objects outside the discourse context or the defnp fails to refer as intended by the speaker. The former case is marked by the presence of discourse links elsewhere in the sentence to the focus or by the use of default reference. In those cases where the reference as intended fails, the hearer may attempt to create a connection to the focus, as was suggested with D10-7, and thereby maintain the perspicuity contract. Alternatively, the hearer may fail to understand the referent.

Focus can be used for disambiguation of generic defnps in a manner similar to the cases presented above. As discussed earlier, the EBC rule predicts that a defnp which is u-ambiguous will be generic or not based on the focus, and that a strictly generic defnp is independent of the focus. The strictly generic defnp case, as shown in D15, presents one means of shifting focus in a discourse with defnps.

D15-1 I got a new ASR 33 this week.
2 *The ASR 33 is an old but reliable output device*

If D15-3 were "It was available long before the newer electronic consoles," the focus would have shifted from the ASR33 which the speaker had gotten to the generic entity of ASR33 teletypes.

The strictly generic defnp used after a non-generic focus is just one case of implicit backwards co-reference using associations. Other associations occur as well. However, implicitly related defnps are not distinguishable as generic unless a full modifying noun phrase is attached, as D16 shows. *The time* as a simple defnp can be used only as an implicit co-reference to the focus of party. Only the complex noun phrase has the syntactic and semantic distinctions which reflect the generic usage. The simple defnp used implicitly takes its generic/specific classification from the focus.

D16-1 I want to have a party.
1a The time of a party is hard to decide on.
2b The time is hard to decide on.

Set-element implicit co-reference exists for generic foci as well as specific foci. A significant difference is that the foci for the generic case can be a singular defnp, or they can be a plural noun phrase with either a definite article or no article. The set membership is indicated by a distinguishing modifier, just as with specific set-element implicit co-reference. An example of generic set-element co-reference is given below with both a singular defnp focus and a plural unspecified focus.

D17-1 The Austrailian aborigine represents an almost extinct hunter-scavenger social group.
1' Australian aborigines represent an almost extinct hunter-scavenger social group.
2 The aborigine in the southern sections of Austrailia sometimes gathers food, but the other aborigines do not.

Inferred generic co-references also occur. In D18, *the owner of a motorcycle* is a generic defnp:

D18-1 Alronso was in an accident with a motorcycle last week.
2 I think the owner of a motorcycle ought to be required to take driving lessons.

The owner of a motorcycle is generically related to the first sentence by an inference of what happened to the agent. The same kind of machinery that is used for specific inferred co-references can be used for making these inferences as well. How can generic inferred co-references be distinguished from specific inferred co-reference? A strictly generic defnp as in D18-2 remains generic. Those defnps which are u-ambiguous at the sentential level, as with D10-2, can be disambiguated as specific because of the relation to the focus.

The use of a semantic network with a focus and inference mechanism results in a computational theory of co-reference which makes use of representation properties such as prototype of entities, hierarchical connections and associative links between entities in the representation. The use of focus for co-reference rules such as the EBC rules, implicit backwords co-reference rules and inferred co-reference relies on this net representation. An inference mechanism is necessary both to verify co-reference predictions and to test suppositions used in inferential co-reference. With the net and a focus, predictions about acceptable co-reference for noun phrases has been shown and verified with linguistic evidence. Psychological predictions, such as implicit co-reference requiring more time than explicit co-reference, can also be tested although the related psychological literature has not been discussed in this paper. The limits of focus as a co-reference mechanism suggest that focus is central for co-reference of noun phrases related to previous discourse. For noun phrases that refer outside the discourse, focus may also be used to generate a context of entities from which a co-referent may be chosen. Further research can extend the focus mechanism to rules involving

other types of noun phrases and the personal pronoun anaphors.

9. Acknowledgements

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under the Office of Naval Research under Contract Number N00014-75-C-0643.

10. References

Austin,J.L. [1962] *How to Do Things With Words*, Oxford University Press.

Bullwinkle, C. [1977] *Levels of Compexity in Discourse for Anaphora Disambiguation and Speech Act Interpretation*, Proceedings of the Fifth International Joint Conference in Artificial Intelligence, August 1977; also M.I.T. A.I. Lab Memo 413.

Charniak, E [1972] *Toward a Model Of Children's Story Comprehension*, M.I.T. A.I. Lab TR-266.

Cohen, Philip R. [1978] *On Knowing What to Say: Planning Speech Acts*, University of Toronto Department of Computer Science, T.R. No. 118.

Donnellan, Keith S. [1977] *Reference and Definite Descriptions* in Naming, Necessity, and Natural Kinds, S.P. Schwartz (ed.), Cornell University Press, Ithaca.

Fahlman, Scott E. [1977] *A System for Representing and Using Real-World Knowledge*, M.I.T. A.I. Laboratory AI-TR-450, December.

Grosz, Barbara [1977] *The Representation and Use of Focus in Dialogue . Understanding*. Stanford Research Institute Technical Note 151, Menlo Park, California

Halliday, M.A. and Ruqaiya Hasan [1976] *Cohesion in English* London, Longman press.

Karttunen, Lauri [1968] *What Makes Definite Noun Phrases Definite?* Paper p-3871, The RAND corporation, Santa Moncia, California.

Kripke, Saul A. [1972] *Naming and Necessity*. in Semantics of Natural Language, Davidson and Harman (eds.) Reidel Publishing Co Boston.

Norman, Donald A. and David E. Rumelhart [1975] *Explorations in Cognition* W.H Freeman, San Francisco.

Rieger, Charles J. [1974] *Conceptual Memory: A Theory and Computer Program for Processing the Meaning Content of Natural Language Utterances*: Stanford Artificial Intelligence Lab Memo AIM-233.

Russell, Bertrand [1905] *On Denoting*, reprinted in Logic and Knowledge, R.W. March (ed.). Allen and Unwin London, 1956.

Searle, J.R. [1969] *Speech Acts: An Essay in the Philoshpy of Language*, Cambridge University Press.

Sidner, C. [forthcoming]. *A Computational Model of Co-reference Comprehension in English*. Ph.D. dissertation, M.I.T.

Winograd, Terry [1971] *Procedures as a Representation for Data in a Computer Program for Understanding Natural Language*. M.I.T. dissertation.