

**Ethnoscience 1972**

Oswald Werner

*Annual Review of Anthropology*, Vol. 1 (1972), 271-308.

Stable URL:

<http://links.jstor.org/sici?sici=0084-6570%281972%292%3A1%3C271%3AE1%3E2.0.CO%3B2-7>

*Annual Review of Anthropology* is currently published by Annual Reviews.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/annrevs.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).



## ETHNOSCIENCE 1972

9510

OSWALD WERNER<sup>1</sup>

*Department of Anthropology  
Northwestern University, Evanston, Illinois*

### INTRODUCTION

After the first flurry of successes at discovering folk classifications by relatively rigorous eliciting techniques, few major advances have been made in the direction of a better understanding of lexical/semantic fields.<sup>2</sup> Casagrande & Hale (6) provided a nearly exhaustive list of the major semantic relations, but little work followed in spite of increasing evidence that lexical/semantic relations are more readily recognizable as language universals than are semantic components (Werner 69). A partial explanation is that work in this area is extremely difficult and sociolinguistic approaches (ethnography of communication) seem to offer easier solutions. I think this is an unrealistic estimation since considerations of context, though certainly crucial, are also certainly more complex than so-called "context free" semantics.

Advances have been few because experimentation with large lexical fields exceeds the limitations of procedures by hand. It is simply too difficult to keep track of 1000 or more lexical items and their ramifications. The systematization of lexicography—and ethnoscience can be viewed as an attempt at systematic and nonlinear (nonalphabetic) lexicography—has always been burdened by the bulk of data it had to process. The rather unsystematic variations of definitions in existing dictionaries are a living monument to this problem.

Progress in ethnoscience is slow because the purpose of the exercise was never made entirely clear. Theorizing in terms of "synthetic informants" or in terms of question-answering systems promises to fulfill for the first time Goodenough's dictum that an ethnography should allow one to behave like a native of that culture—at least for the general area of cultural verbal behavior based on cultural knowledge. It cannot be stressed sufficiently that

<sup>1</sup>This research was in part supported by a grant from NIMH MH 10940. I am indebted to the following people for their comments: Roy D'Andrade, Martha Evans, John Farella, Terry Strauss, and especially for the detailed analysis of Paul Friedrich. Any shortcomings that remain are my own.

<sup>2</sup>I simply fail to understand Hymes' (31) optimism or see any evidence for its justification. Compare, for example, Hymes to Roger Keesing's recent (38) opening sentence, "Whatever happened to Ethnoscience?"

cultural knowledge is so vast that work in this area is unrealistic without machine aid. Since anthropologists cannot get inside the informant's head, psychological reality is an empty concept. Mind-like mechanical verbal behavior seems best suited for the validation of our assumptions.

The clarity of purpose of question-answering systems allows the investigator to formulate prerequisites in greater precision and detail than by any other method. The requirement that all of the investigator's ideas be representable as part of a computer program forces him into detailed conceptualization and planning exceeding in rigor by several orders of magnitude any similar requirements in anthropology today.

In order to avoid the fate of machine translation, the emphasis should be not on quick payoffs or early practical applications. If the machine is enlisted in the solution of interesting theoretical problems, applications will inevitably follow.

Answering questions by using a large data base creatively (deductively) is not easy, but no one has yet suggested a better method than attempting it by machine. In this paper I first show that indeed four relatively independent areas of intellectual endeavor (ethnoscience, generative semantics, computerized semantic information processing, and sociolinguistics) do converge. Subsequently I try to enumerate some of the contributions made in each of these fields. Finally, I select a few topics in which I have clarified some points sufficiently for myself to share these insights.

Readers interested in more comprehensive coverage of ethnoscience and its history are referred to the excellent reviews by Hymes (30), Sturtevant (66), Colby (8), Durbin (16), Hymes (31), Friedrich (21), Werner (75) and to Tyler's (67) compendium. With the exception of Werner (75), these papers and most of Tyler's book represent a point of view different from the one I assume in this paper.

#### CONVERGENCES

The convergence of ethnoscience, generative semantics, computer simulation of semantic information processing, and some aspects of sociolinguistics (the ethnography of speaking) can be summarized roughly by the following diagram (Figure 1):

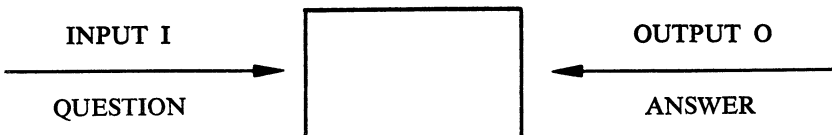


FIGURE 1. Question-answer schema.

The specific inputs and outputs in each of the fields mentioned can be characterized as follows:

*Ethnoscience.*—If a system of this kind is proposed, for example, for a

componential analysis of a kinship terminology, the diagram can be conceived as a device which answers the following questions: "If one person is related to another in a specified manner, what does that person call the other?" or "What is the second person called by the first person?" If the analysis is a taxonomy, the question may be "What 'things' (in this culture, using its language) are considered animals?" or if some processes were part of the analysis, "How do you make a canoe?" or "How does one conduct a funeral?" The questions can be assumed to ask about general topics (e.g. "What are all the things that have been created?" or specific (e.g. "How do you thank the host at the end of a meal?") It seems that a question-and-answer approach is particularly fitting for ethnosciences: if the inputs are culturally appropriate and relevant questions, the outputs will be culturally appropriate and relevant answers.

*Generative semantics.*—One has to distinguish between two cases:

1. The more extreme view where the semantic component of a language is truly generative in the technical sense. By this I mean that the output is a "random generation" of abstract phrase markers. A device of this sort is hardly comparable to any human-like behavior. I shall therefore exclude it from further consideration (see also p. 278).

2. If, however, the generating device is conceived more loosely as one triggered by some thought it is easy to surmise that such thought could have been evoked by a question. Since there seems to be no restrictions on questions that can evoke thought the device is comparable to a general question answerer, analogous to the ethnosciences examples above.

*Semantic information processing.*—In computer experiments of semantic behavior, one possible aim is the simulation of human question-answering behavior. Thus the input is clearly a question and the output is an appropriate answer. If the device is general, that is, able to answer questions potentially in any area of a culture, the device is then analogous to the earlier examples from ethnosciences.

*Sociolinguistics.*—Many problems in the social uses of language can be conceived as answers to questions: e.g. "What does one do when a stranger approaches camp?" "How does one behave in the presence of the president?" or "When is it proper for a young man to speak?" These are questions appropriate to an ethnography of speaking but also clearly questions that have possible answers. They are formulated in the native's language to reduce ethnocentric translation bias and, equally importantly, to avoid imposing one's own categories. Quite likely not all socially significant contexts are explicitly open to description by native speakers. They do misjudge situations on occasion or are unable to state the precise contextual rule. However, since verbalizable contexts are by no means rare or small in numbers there is no reason to exclude them from semantics (following Tyler 67). Contextual attributes can be listed with other attributes. Implicit contexts are considerably more diffi-

cult. Since no minimally adequate metalanguage has been proposed for their description, I exclude them from further consideration, although this exclusion should not be construed as a lack of my appreciation of their importance.

### TYPOLGY

If the discussion thus far is minimally successful in indicating the convergence of these four fields, then why are they separate fields, how do they differ in method and theory, and what constitutes their major emphases? In order to help the reader clarify the destination of ethnoscience, I present a typology which characterizes each field. This typology is also a programmatic attempt at integration of the four fields.<sup>3</sup>

*Ethnoscience.*—Two general approaches in the field require discussion. First is the componential approach. The extension of this method of analysis to the entire vocabulary of a language has been attempted only in spirit. Katz & Postal (33) have sharply criticized anthropologists for this timidity. Nevertheless there are no empirical investigations extant that attempt a componential analysis of all the lexical resources of a language. It has been most successfully applied to small lexical sets, especially kinship terminologies.<sup>4</sup> A subbranch of componential analysis are analyses which can be characterized either as some aspect of decision making (e.g. Gladwins 25) and/or the im-

<sup>3</sup>Howard Maclay (44, p. 180) quotes Victor Yngve's glee in predicting the demise of linguistics as an independent discipline. Although I can empathize with Yngve's joy, I prefer to interpret the demise of "autonomous syntax" [at least in the view of some generative semanticists (e.g. as described by Lakoff 41)] optimistically. It is a positive, promising sign of the integration that Paul Friedrich (personal communication) and I believe to be the best thing that could happen to the fields concerned. I further agree with Friedrich that the most significant advancement of the 1970s will be such a unification.

<sup>4</sup>I have attempted in previous publications (Werner 73, 75) to extend the notion of components to the entire lexicon by the introduction of the 'circle star' (⊙) operation. This operation—if all the components of a language are known (a considerable, or even impossible task)—"automatically" establishes all hierarchies (taxonomic levels) with paradigms on each level of the taxonomy, and enables the investigator to add easily empirical and logical constraints on the occurrence of incompatible component combinations. To the best of my knowledge no one (including myself) has ever seriously tried to apply this model to anything more than my reanalysis of Berlin and co-workers' (1) taxonomy of squashes in Tzeltal. The primary value of my model lies in its illustration of the form a whole language componential analysis might take; especially that there is a requirement for a very large number of empirical and logical constraints that are necessary in order to exclude items not in the speaker's environment, as well as items that are conceptually impossible. The latter idea, not pursued beyond the suggestion that such a logical restriction may be possible, is sometimes raised by generative semanticists.

position of some kind of a temporal order on a decision or recognition process (e.g. Geoghegan 24). It is perhaps too early to tell if such models, whose major overt characteristics are the use of flow charts, are enough of a departure to require separate treatment. Furthermore, it is not clear if the temporal order of the flow chart is an artifact of the analysis, or a reflection of the capabilities of real human beings. Most flow charts can be represented as decision tables which provide a more synchronous impression of the decision processes. More specifically, I was able to reduce the Gladwins' (25) flow chart to two decision tables, one following the other. Although I was unable to reduce it to one table, I am equally unable to decide if the supposed temporal order of the two decision tables has some reality in real time greater or lesser than the stepwise flow charts.

Second is the lexical/semantic relations, semantic field, or lattice approach. It was first introduced in anthropology in the work of Frake (20), although any earlier attempt at drawing folk taxonomies in the strict or extended sense qualifies as well. The early origins of Field Theory are found in the works of Trier, Portzig, and Weisgerber.

By taxonomy I mean: taxonomy in the strict sense which applies only to the test frame "\_\_\_\_\_ is a (kind of) \_\_\_\_\_". By extended sense I mean the more usual interpretation in anthropology which comprises any scheme of classification. For theoretical purposes, as I hope to show and have shown, especially in Werner & Fenton (74), strict taxonomies and other transitive relations (i.e. those that form hierarchies) must be separated. Thus, for example, a taxonomy of most human ethnoanatomies in the extended sense turns out to be an alternation of strict taxonomies ("The thumb is a kind of finger") and the part-to-whole relation ("The finger is part of the hand").

The notion was extended from the beginning to the entire lexicon. The list of available lexical/semantic relations constructed by Casagrande & Hale is amazingly exhaustive (6). The relatively low level of acceptance of this model thus far is apparently due to the unfortunate selection of block diagrams for the representation of taxonomies:

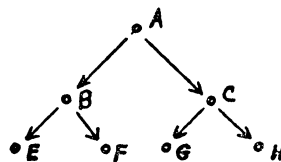
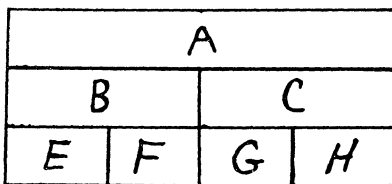


FIGURE 2. Block diagram of a taxonomy. FIGURE 3. Directed graph of a taxonomy.

While Figures 2 and 3 are for all purposes isomorphic, Figure 2 fails to show clearly that for an interpretation of the formal graph the edges (lines) of the graph need to be labeled (by the relation, here T for strict taxonomy) and directed (by arrows) to imply the asymmetry of the relation, i.e. that

$ATB \neq BTA$ . Almost no one seems to doubt—in sharp contrast to componential analysis—that extension of the relational approach to the entire vocabulary of a language is valid. Among numerous field techniques for the collection of data that have been suggested, the elicitation frame is perhaps foremost. [See compendia of field techniques in Perchonock & Werner (52), Werner & Fenton (74), and Werner et al (75).] It is relevant because a well-chosen question will elicit an explicit statement about some of the field relations that may hold between two or more terms. The ultimate test for the existence of a particular kind of linkage is, with few exceptions, the existence of a linking sentence in the informant's language. Componential analysis can be interpreted as an analysis subsequent to a taxonomic analysis, that is, componential paradigms are further structures imposed on a particular level of a particular taxonomy.

Anyone doubting the foregoing statement should consider that the isolation of the kinship vocabulary which precedes a componential analysis of kinship addresses itself to the question, "Which terms in this language are kin-terms?" or "Is \_\_\_\_\_ a kind of kinsman?" A major shortcoming of many componential analyses is that they do not take into account intermediate taxonomic organization. In the Yankee terminology, terms like "parent," "grandparent," "child," "grandchild," "ancestor," "descendant," or even "blood relative" and "relative by marriage" are surely part of the reckoning. By imposing taxonomic organization first, the paradigms to be analyzed become considerably smaller than those usually found in the literature. Furthermore, the taxonomic structure imposes some degree of order on the components that reduces the number of possible analyses [see Burling's (5) apprehensions]. Furthermore, areas with ambiguities or multiple analyses are often cases where the indeterminacy carries some cultural significance. Surely the fact that all alternate analyses of the Yankee system contrast along one and only one of the dimensions of the analysis (namely the dimension of collateral distance) implies that such a distance measure is at best vague and highly variable, or even idiosyncratic, in view of situations in which American families find themselves today.

Thus paradigmatic structuring is not alien to a lexical/semantic field type description. However, the existence of universal semantic components is debatable (see Werner 69, 76). The unifying principle in both approaches is a common concern with the lexical resources of the languages under investigation. It is therefore possible to characterize this field as lexicographic ethnoscience.

*Summary.*—The scope of the field of ethnoscience vis-à-vis the general phenomenon of language is relatively narrow: the primary focus is the lexicon. Anthropologists have investigated the vocabulary of weddings, cures, illness, religion, anatomy, firewood, how to do things (naming the sequences of behavior), or similarly what the steps are that take one through a day—all in a variety of cultures. Historically most anthropologists did not exclude the

lexicon from language as proponents of autonomous linguistics did nor did they exclude the vocabulary from culture. Almost all statements of linguistic relativity or the Sapir/Whorf hypothesis are statements emphasizing the lexical resources of language.

I have argued (Werner 71) that proponents of an "autonomous linguistics" subscribe to a bias which I called "grammaticalist." In this view everything outside of phonology and grammar—supposedly the only rule-governed parts of language—are outside of linguistics. Perhaps Voegelin and Harris espoused this view most consistently in a series of articles well known to anthropologists of the early 1950s. Most anthropologists (especially Opler as the spokesman of the other side of the Voegelin/Harris position) can be characterized as representing the "lexicologist" bias. To these anthropologists a vocabulary that was not part of language (more precisely linguistics) made no sense (and rightfully, I believe). The lexicographic interests of ethnoscience have their roots in this bias.

In addition, anthropologists were concerned about how lexical items are related to each other. Folk taxonomies emerged early as an important principle of lexical organization. Although implicitly present in many entries of standard monolingual, English dictionaries, for example, they have not yet been systematized to any extent. The list of lexical/semantic relations of Casagrande & Hale, based on some 800 Papago folk definitions, is the first to be almost exhaustive. The authors postulate a relationship of constituency (part-to-whole) but find no examples. Neither do they find the Conklin/Frake relation of "\_\_\_\_\_ is a stage of \_\_\_\_\_," probably because they are dealing with people less essentially agriculturalists than the people whom Conklin and Frake studied.

It is important to mention here the work of Romney & D'Andrade (58) and D'Andrade (12, 13, and especially 14). All of these papers use a statistical approach for the validation of lexical/semantic structure. By this I mean following the typology of Campbell and Fiske, which calls for validation by maximally different procedures and replication by maximally similar procedures.

Eventually it may become possible to simulate the statistical validations as part of the question-answering routines. Similar frequencies for men and for machines will surely be highly significant.

More importantly, D'Andrade's latest paper (14) shows that statistical methods can be helpful in finding some strong lexical/semantic relations between pairs of terms. Some of these relations may not be elicitable with ease by any other technique. It is interesting that D'Andrade's work also seems to imply that significant success lies in the application of modal logic to deductive systems (see p. 275).

Other significant contributions to the notion of lexical/semantic fields in anthropology are Berlin et al (1), Kay (34, 37), Sanday (60), and Werner (75).

The stage is now set for a sophisticated view of the lexicon. In this view



the lexicon is a complex structure, probably some sort of lattice, where large numbers of terms are intimately interrelated. The picture that emerges is close to Bierman's stars. Bierman (2), a logician, envisioned the organization of the vocabulary as a system of relations. Each distinct lexical item occupies a node in a large plane. Lexical items that are related are linked by colored strings. The color represents the type of relations. Since these relations are binary, every ray belongs to two stars, the one from which it emanates and the one that it reaches. Presumably sentences of a language are retrievals from, or paths in, this huge network. Any sequence of nodes, if properly constrained, is a possible sentence. The problem is, however, that anthropologists have found no explicit constraints which restrict in some justifiable manner the choice of possible paths through the lattice.

The unrestricted lattice idea is obviously the major weakness of ethnoscience. In general, a most rudimentary view of syntax prevails. Sentence frames are used to discriminate vocabulary items from each other or help in the assignment of vocabulary items to specific domains. How human beings in any culture, with any language, in spite of (so to speak) huge lexical/semantic fields are able to speak or answer questions has not been broached. It is fair to say that language, or the cultural manifestation of language, is seen from the vantage point of a sophisticated lexicographer. The informant's universe is seen as being predominantly his lexicon.

*Generative semantics.*—Following approximately Postal's (53, p. 261) and Maclay's (44, p. 177) representation GT-3b, generative semantics can be characterized as diagrammed in Figure 4. I am simplifying in order to serve better the purposes of this paper; for a more complex view, see Lakoff (41).

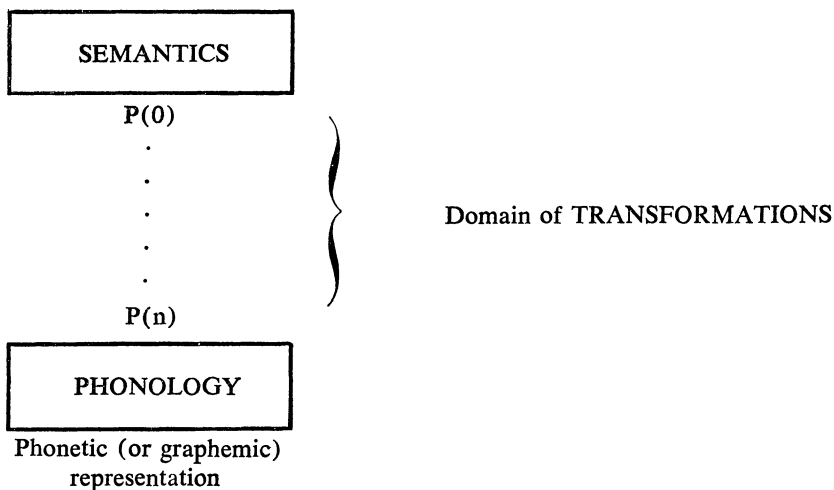


FIGURE 4. Schema of generative semantics.

Roughly, Figure 4 can be explained as follows: SEMANTICS is considered first as truly generative, that is, randomly generating phrase markers  $P(0)$  or tree graphs which serve as structural descriptions of the arrangement of semantic units. I will not utilize this view because I do not see how it could contribute to ethnoscience. I will use the second view throughout—any output produced by the stimulus provided by a question. This view is perhaps closer to the view represented by Binnick (3).

A second interpretation, which assumes that the character of the “device” called SEMANTICS is better known, produces one phrase marker  $P(0)$  at a time. The structure within SEMANTICS is conceived as somehow linked to a question that elicited it. That is, a  $P(0)$  may be produced in response to a question. The structure  $P(0)$  is then taken through a series of TRANSFORMATIONS which (a) insert phonological representations of semantic units and (b) change the structure  $P(0)$  into  $P(n)$ —the surface structure—which is an acceptable input to the PHONOLOGY, which in turn provides the sentence with directions for pronunciation or phonetic representations.

If, according to generative semanticists, there is a boundary left between semantics and syntax, it is the boundary between SEMANTICS and the “transformations.” However, what I identify here as “transformations” must be interpreted more broadly than the transformations in the *Aspects* (Chomsky 7) sense (see Lakoff 42). Furthermore, as Postal (53) points out, some of the transformations must apply before the boundary that separates SEMANTICS from syntax. This is due to the fact that lexical entries themselves possess an internal structure. This structure also has the form of dependency trees. Otherwise it would be difficult to account for the underlying structure of many verbs similar to “kill.” The semantic primitive of this verb is assumed to be “to cause to die.” The lexical representation (for greater detail of the notational convention see Figure 5) of the verb “to kill” is then depicted approximately as in Figure 5.

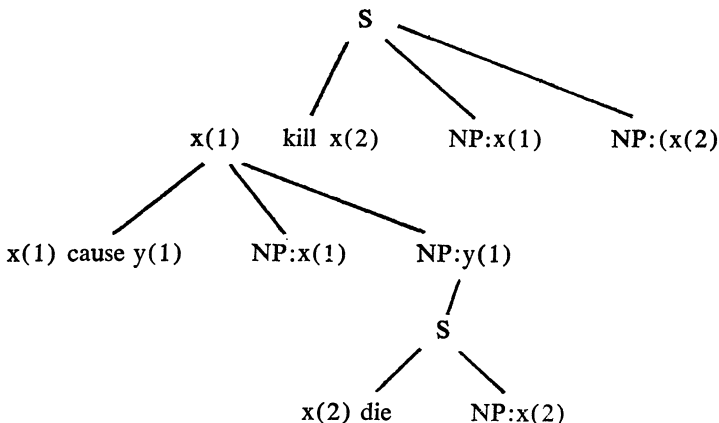


FIGURE 5.

In the view of generative semanticists (especially Lakoff 41, McCawley 48, and other publications),  $P(0)$  markers do not stand alone. They are accompanied by presuppositions. Lakoff (41, p. 235) gives the following example: A sentence such as "Pedro regretted being Norwegian" presupposes that "Pedro is a Norwegian." A semantic representation SF of a  $P(0)$  is conceived as

$$SR = (P(0), PR, Top, F, \dots)$$

". . . where PR is a conjunction of presuppositions, Top is an indication of the 'topic' of the sentence, and F is the indication of the focus of the sentence. We leave open the question of whether there are other elements of semantic representations that need to be accounted for" (Lakoff 41, pp. 234–35). The presuppositions contain the speaker's and the listener's knowledge of the world. They are indistinguishable from standing sentences, that is, sentences with a truth value independent of time. In the above example it is presupposed not only that "Pedro is Norwegian," but also that "Pedro was a Norwegian" and that "Pedro will always be a (native) Norwegian." That the knowledge of all the world is involved emerges from the fact that not only do we know that "Pedro is a Norwegian," but also that "Pedro is a certain man" whom the speaker knows, but the hearer may not. That "Pedro is a man's name," that "Pedro is human, animate and a physical object," that "Pedro is an unusual name for a Norwegian," that "One would expect Pedro to be a Spanish name," that "Einar, Olaf, Knut . . . are more usual Norwegian first names, ". . . etc, etc. All of these presuppositions may come into play at one time or another in a discourse about Pedro and his regrets.

When Lakoff says "Pedro regretted being Norwegian." I am tempted to ask "Who the hell is Pedro?" To which Lakoff would probably reply by providing part of *his* presuppositions about Pedro, e.g. "He is that nut I told you about yesterday who likes Southern California but thinks that he can't immigrate because he doesn't have a valid birth certificate." Lakoff's reminder will serve to refresh my own memory of presuppositions and/or provide new ones as well.

If a sentence contains embedded sentences, each embedded sentence requires its own presuppositions. Lakoff diagrams this as follows:

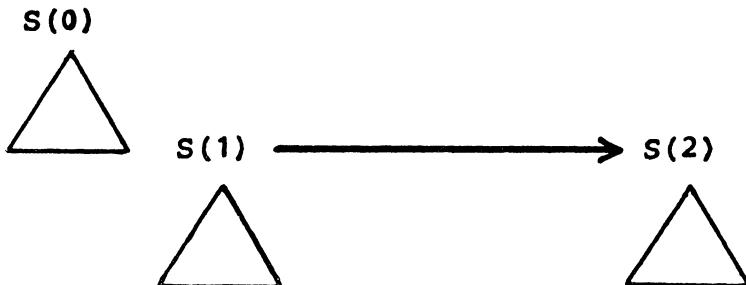


FIGURE 6. Embedded sentence presuppositions.

where the long arrow stands for "sentence S(1) presupposes sentence S(2)."

McCawley (48) further assumes (which is at least implied by Lakoff's diagram) that semantic representations are in the form of phrase markers. Rather than resembling natural language, they resemble the predicate calculus of symbolic logic. Thus Sapir's sentence "The farmer killed the duckling" may be represented as follows:

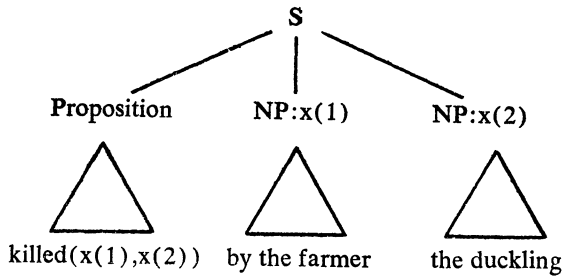


FIGURE 7.

$x(1)$  and  $x(2)$  are the variable arguments of the predicate "killed."

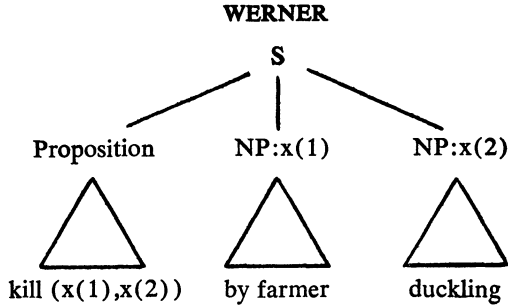
The independence of the noun phrases NP is argued by McCawley on the basis of examples like "John says that Nancy wants to marry a Norwegian."<sup>5</sup>

The sentence is ambiguous among the three senses (i) there is a person who John says Nancy wants to marry and who the speaker identifies as a Norwegian, (ii) John says that Nancy wants to marry a certain person whom John identifies as a Norwegian, and (iii) John says that Nancy wants her future husband (whoever he might be) to be a Norwegian." [McCawley continues] It is difficult to see how these senses could be assigned different 'deep structures' unless those structures allowed Noun Phrases to occur separate from the propositions that they are involved in and to be constituents of sentences in which those propositions are embedded (McCawley 48, p. 225).

The word order in kill ( $x(1),x(2)$ ) is justified by McCawley (48) on the grounds that many transformations in English can be stated considerably more simply if the verb-first order is assumed. The only alternative in English is a verb-second position which complicates matters.

In order to illustrate this point, let us look at a crude approximation of the passive transformation (articles and the introduction of the stative aspect of the verb "to be" in the passive are suppressed):

<sup>5</sup> My apologies to all Norwegians. I am sure that neither Lakoff nor McCawley intended this coincidence, which is purely accidental.



*Active:* Rule: move NP:x(1) to first (focal) position (by deleted in initial position).

*Passive:* Rule: move NP:x(2) to first (focal) position.

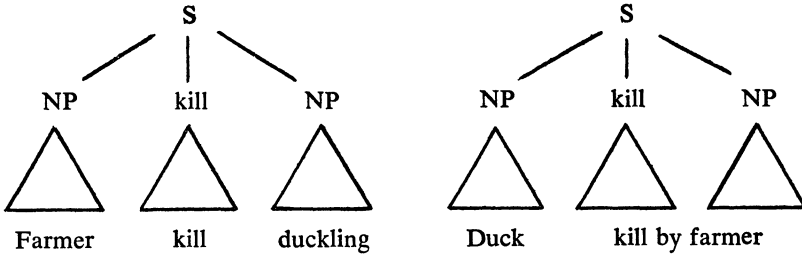


FIGURE 8.

In the verb-second structure both NPs would have to be moved. Thereby an important generalization is lost: move the focal NP into initial position where it becomes the topic of the surface sentence.

Recently it has been argued (e.g. McCawley 47) that the distinction of the traditional parts of speech is a characteristic of relatively shallow structure. Lakoff (39) has shown that adjectives and verbs may be considered as one form class. Furthermore, transitive nouns are formally identical to transitive verbs:

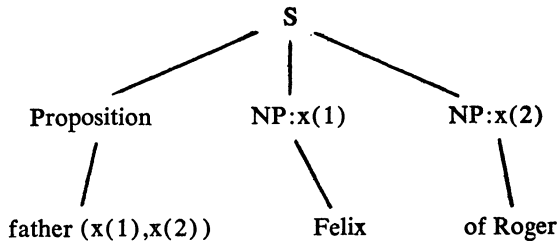


FIGURE 9.

The only notion that remains for SEMANTICS is "predicate of" (more on this later).

The insight that tense and modals and other verbs like "usually" are derivable as higher (further up in the phrase marker) predicates introduce important simplifications.

Other work has concentrated on the nature of quantifiers, especially the similarity of the universal quantifier and the conjunction "and" and the existential quantifier and the disjunction "or." However, there are still many more unsolved problems and summarizing may be premature.

Finally, some as yet rudimentary efforts have been directed toward the inclusion of deductive rules in the theory. Some sentences can be clearly derived by deduction from other sentences and possibly from the set of their presuppositions. Lakoff (40) in particular has made some exploratory forays into applications of modal logic, which introduces predicates of possibility ("may") and necessity ("must") and can be extended to temporal predicates (i.e. "sometimes" and "always"). His insights are at this point largely provisional.

*Summary.*—It is apparent that generative semantics has made significant progress toward the solution of some of the problems that face ethnoscience. One of these is that lexical entries cannot be conceived simply as bundles of components. The notion of presuppositions converges with the notion of a lexical/semantic field. But while anthropologists have pointed out some of the characteristics of the lexical/semantic network, linguists have largely ignored it. Perhaps it can be justly said that the insights of generative semantics are still rooted in the notion of an autonomous syntax, which they rightly attack. Their methods, approaches, and arguments proceed largely from a syntactic base at the expense of the structure of the lexicon and the need for some deductive capacity for natural languages.

*Semantic information processing.*—Although many workers in this field are associated with Minsky (49) at MIT, they do not operate with a paradigm that can be easily characterized. The efforts I want to describe are all concerned with computer-simulated question-answering systems. In recent years these studies have assumed two relatively independent directions differing primarily in their emphasis rather than in overall orientation. The first approach is an attempt to model the structures that are necessary for the representation of large memories. The second concentrates on the deductive capacities of question-answering devices.

*Memory models.*—Among a dozen or so experimenters in artificial intelligence, Quillian (54, 55) is unique. He deals almost exclusively with the simulation of very large memory structures ". . . in which newly input symbolic material would typically be put in relation to large quantities of previously stored information . . ." (54, p. 220). "Actually most simulation pro-

grams . . . have not been primarily concerned with long term memory at all but rather with cognitive [deductive] processing" (54, p. 219). Quillian's semantic memory model can be envisioned as a card file representation of the lexicon. Every dictionary entry has one card. Each card contains one entry name which Quillian calls the Type. Under the entry are paraphrases (definitions) in a special notation. Occurrences of words in these paraphrases are termed Tokens. The linkages of lexical semantic relations are as follows: 1. B names a class of which A is a subclass:  $A \rightarrow B$ : corresponds to the relation of taxonomy; 2. B modifies A:  $A \leftarrow B$  corresponds to the relation of modification; 3. A, B, and C form disjunctive sets  $A \text{ or } B \text{ or } C$ : corresponds to disjunction; 4. A, B, C form a conjunctive set  $A \text{ and } B \text{ and } C$  corresponds to conjunction; 5 and 6. B, a grammatical subject, is related to C, a grammatical object, in the manner specified by the relation A (verb)  $A \xrightarrow{\quad} B \rightarrow C$ ; 7. the associative link is the system of addresses (links) that relates the occurrence of a word to its occurrence in a paraphrase, that is, it connects every entry card to the occurrences of that entry in the definitions of other entries. Thus this system of relations makes Quillian's model a general graph or network of associations.

The experimenter's task is to select any pair of words from a store and to submit them to the program for comparison. He may check the machine's comparison against the comparison made by a native speaker. That is, "he considers whether or not the machine's output is one that might reasonably have been produced by a subject" (Quillian 54, p. 237).

Quillian's second model (55), the TLC (Teachable Language Comprehender), is also an attempt to simulate input and output information in smooth English. As a result the canned phrases and ad hoc syntactic form tests (parsings) of the sample outputs give an extremely and deceptively human-like impression.

The basic parts of the program are (a) the memory network, (b) a processor, and (c) an on-line human monitor. The memory is structured similarly to Quillian's first model but is somewhat more abstract: the dictionary part is simply a list of lexical items with a list of addresses following every entry. The addresses point to nodes in an abstract network. In my terminology one could look at his lexicon as a linear arrangement of items which connect to a giant switching circuit, the abstract network. The network consists of units which have one obligatory element, a pointer to its superset (superordinate taxon). In addition, the unit may have pointers to several properties. These have attributes and values which are also pointers. Roughly, the units are the Types of Quillian's first model and the properties and values are the Tokens. The attributes are verbs and adjectives and the values are nouns that function as grammatical objects. Thus Quillian's second model is essentially Quillian's first with only the relation (1, 5 & 6) retained.

" . . . The comprehension procedure of TLC may find a number of properties related to a piece of text and, by using adapted copies of these (i.e. representing them in Quillian's special notation), create a complex inter-

linked structure, in the same format as the memory (i.e. these structures are merged with the existing portion of the memory), representing the particular meaning of the current input string" (55, p. 475).

By relating the meaning of the current input string to already present information, the device is capable of producing one (? all Quillian's examples contain but one) possible paraphrase (see below).

An intermediate model between Quillian's and the models emphasizing the deductive capabilities rather than memory organization is PROTOSYN-THESIS III by Schwartz et al (61).

This structure consists of four parts. First are the ordered triplets of the form X R Y where the R are generally verbs and the X and Y can be simple or complex. Second, there are special relations. These contain the relations of taxonomy and every concept is in a taxonomic relation to at least one other concept. Concepts may also be in the relation of equivalence to each other. The primitives are the most general concepts of the system. Third are the complex relations which form the backbone of the deductive system. These are abbreviations of complex systems (relations of relations) of the special relations and are called inference rules. Finally, there are the semantic event forms whose function is similar to Katz & Postal's selection restrictions. That is, they are called upon to resolve ambiguities (?).

While the deductive capabilities of the system are impressive (it can, for example, deduce the problem of the monkey, a stick, and the bananas above the monkey and out of his reach) it seems in spirit closer to Raphael's and Black's systems of limited deductive capabilities. At no point is there an attempt to correlate a *large* lexical/semantic field. All examples given start with some simple input sentence inserted by hand and deductions based on such very limited input. For lexical/semantic fields the inference rules appear to be unnecessarily complex. For example, if AR1B stands for "A is brother to B" and CR2B for "C is father of B," then the inference rule A[R1 C/P R2]B is constructed, that is, "A is uncle of B" (more precisely, "A is a kind of uncle of B"). It is not clear how this would constitute a simplification over: If (A)T (brother of C) and (C)T(father of B), then (A)T (uncle of B), where T stands for the taxonomic relation of the earlier formulation. This solution requires no new symbols nor new relations.

The program searches a question like "Who lost the battle of Waterloo?" in two steps: (a) "Who lost the battle?" and (b) "Battle of Waterloo." In order for the device to function properly, auxiliary propositions are necessary if it is to solve more complex situations. If the input sentence was "Napoleon was beaten in the Battle of Waterloo," then the above match will not work. Optimally one would expect the device to respond "if 'was beaten in' is equivalent to 'lost' then the answer is 'Napoleon.'" Please confirm first part of proposition." It would be the human monitor's task to make such confirmations. After validation of the equivalence, the two verbs become part of the permanent capability of the system. In this sense Quillian (55) is right in making his device teachable. Teachability, that is, the capacity to ask ques-



tions (or the confirmation of hypotheses as above) in the process of question answering, seems the only economical way to build such devices.

If none of these models represents a "great leap forward," it should be remembered that at present "almost any program able to perform some task previously limited to humans will represent an advance in the psychological theory of that performance" (Quillian 55, p. 459).

*Deductive models.*—Most of the models in this category deal with some form of symbolic logic. Much of the work in the field is brought together by Minsky (49). Later developments are described by Simmons (63).

Instead of going into a detailed exposition of symbolic logic, let me try to indicate by an example roughly what is involved. Part of the following example is from the R2 program (Biss et al 4) which contains in a verb-first logical notation a data base of some 2000 sentences of the Illinois Driver's Manual *Rules of the Road*. The device has thus by far the largest memory on which to base its deduction. Logical operators include "and," "or," "not," and "implies." Quantification may occur over any variables of this artificial language.

Following Bis et al (4), if the system receives the question (conjecture):

(i) Do cars always yield to pedestrians?

the resolution principle is invoked for the solution. Following Slagle (64) this means that ". . . a clause implies its factors . . . [and therefore the solution is found by] . . . working back from the conclusion toward the hypotheses of the theorem to be proved." In the above example this works approximately as follows:

First, the conjecture (question) (i) is negated ('always' assumed for all examples), which yields the 'clause':

(ii) Cars do not yield to pedestrians.

This is compared to the 'knowledge' of the system which states:

(iii) If X yields to Y then Y does not yield to X.

or for this example (insertion of x and y):

(iiib) If a pedestrian yields to a car then a car does not yield to a pedestrian.

Because of the formula  $p \rightarrow q \Leftrightarrow \sim p \vee q$  (p implies q if and only if not p or q). Therefore (iiib) can be restated:

(iv) Pedestrians do not yield to cars or cars do not yield to pedestrians.

It follows from the identity of (ii) and the second half of (iv) that

(v) Pedestrians do not yield to cars

But the 'knowledge' of the system includes the statement:

(vi) Pedestrians yield to cars in crosswalks

Comparison of (v) and (vi) leads to a contradiction and the system replies to question (i): "No."

This problem as presented is almost trivial. However, with a large number of propositions (axioms) to select from, heuristic methods have to be found. Such heuristic programs operate generally as follows: At every step in the proof procedure there are numerous alternatives. The program tries each alternative on the first level (usually there are many levels). Some evaluation measure exists that assigns to the first step of each possible alternative some numerical value. The program selects the alternative with the highest value and checks all the alternatives at this point and so on. In some very sophisticated programs values of past evaluations are stored so that if a lengthy back-track is necessary, the evaluations do not have to be recalculated. The device goes back and simply chooses the alternative with the next highest evaluation. Most chess-playing programs operate this way.

*Summary.*—The workers in the field of automatic answering systems usually use some simplified form of ordinary English. There is usually some kind of representation of propositions in a central memory. The notation systems are usually variations of notations of the first-order predicate calculus. All processes of deduction use variants of the propositional calculus of symbolic logic; some use the full power of the first-order predicate calculus. As soon as the problems reach some degree of complexity exhaustive searches are out of the question and heuristic programs have to be applied. These, stated most simply, use some evaluation procedure to reduce the number of cases which the device has to inspect in its path toward a solution. Both in the complexity of their deductions and the sophistication of their heuristic programs, the workers in this field are exceedingly advanced. However, their notions of the nature of the lexical field, the nature of lexical/semantic relations, and the fact that many of their solutions are simply "to solve the immediate problem" (i.e. ad hoc) rather than motivated by insights into the nature of language weaken their position. Much of the work in this area creates the impression that quick payoffs are the primary goal at the expense of theoretical insights into the nature of language. In this sense, although raising interesting questions, Dreyfus' (15) critique of "artificial intelligence" suffers from his tacit assumption that we have before us the ultimate in sophisticated computers, and more importantly, that our theories of language are anywhere near adequate to the task of answering questions in a humanlike manner.

It is premature to talk seriously about the limitations of computers except as part of the limitation of deductive systems (see Werner 72, in preparation). The relatively slow progress and failure to meet predicted achievements (Dreyfus 15) are poor indicators of the ultimate potential and/or limitations.

For anthropologists the cross-cultural validity of symbolic logic is an important question. Unfortunately, in anthropology the statement "The logic of the ABC" (where ABC is the name of a tribe or nation) is far from clear. It may refer to the fact that the ABC have a different set of basic axioms (propositions) about various parts of their universe than we do (see e.g., Malinowski: 44a), that the ABC use a different style of argumentation and do not accept some of ours. Finally, perhaps their rules of inference are different from ours. However, the evidence to support or refute the latter hypothesis is at best scanty. A very strong commitment to occidental symbolic logic and its extensions to modal logic seems to be the safest approach to the cross-cultural problem. The stronger the commitment and the more rigorous the application of European deductive systems to native systems of knowledge, the sooner will discrepancies surface and demonstrate beyond doubt that the deductive logic in question is truly different.

*Sociolinguistics.*—According to Fishman (19), there are two aspects of sociolinguistics: micro sociolinguistics, which is synonymous with the ethnography of communication, and macro sociolinguistics. In this paper only the former is relevant. Perhaps the most fundamental contribution of sociolinguistics is the emphasis on not merely the speech act but on the context in which it takes place. Although some work in this field assumes context as given or as easily inferable by the observer, a more sophisticated view is closer to our aim. According to Hymes (29, p. 27), "native terms are one guide," and Fishman (19, p. 43) makes "*verification from within the speech community*" (emphasis his) requisite. In other words, context is a series of native language texts which following Ruesch & Bateson (59, p. 276) describe at least the following four contributing factors: "(1) 'Perception of the other's perception,' or the establishment of the unit of communication. (2) The position of each participant and his function as observing reporter. (3) Identification of the rules pertaining to a social situation. (4) Identification of the roles in a social situation."

Except for contexts that are difficult to verbalize, descriptions of context in the native language are propositions about the use of language, hence, an integral part of a system of propositions which form lexical/semantic fields. Tyler's (67, p. 268) notion that "context itself is a part of the semantic system" is therefore correct. However, features of context in a lexical/semantic field are not separate from other semantic facts. An occasional sentence containing the referent "red" (in the literal sense) requires some red object to be present. Similarly, it implies that certain listeners are involved, or that these correlate to each other in some way, or what mode they use to communicate, or their style of delivery and the topic of discourse (based on Hymes 29, p. 216) depend on the presence of these factors. That these are modifiers analogous to "red" can be seen from the observation that the speaker and listener(s) must have past experiences against which to judge these external clues. Such internal representations are probably identical with attributes. It is

not necessary that attributes be simple, especially since the elicitation of contexts may be derived from the elicitation of long texts. But if higher order predicates (like "may") modify an entire sentence, there is no reason to deny the possibility that relatively simple sentences can be modified by several complex interlinked sentences. Since a general question-answering device should be capable of answering all kinds of questions, and since questions about the social, psychological, ecological, spatial, and temporal context are part of the method of getting at cultural rules (including rules for breaking rules, i.e. systems of priorities) there is no rational way to separate contextual variation from other kinds of variation. At least in part (as I show below) it is possible to control variation of the topic of discourse by a partitioning of the vocabulary.

*Summary.*—Micro sociolinguistics or the ethnography of communication is a young field, although Malinowski ought to be considered an important ancestor. It does not have an explicit metalanguage for the description of its major concern, namely "context." The demonstration of the importance of context is a major contribution to our understanding of some of the variations of meaning. Nevertheless, the field is at this point primarily descriptive rather than highly theoretical, and although it can as yet contribute little to question-answering systems its lesson, that context is semantic and needs to be represented, or at least be representable, is of the utmost importance.

#### PROBLEMS AND COMMENTS

In this section I select a few problem areas raised in my exposition and elaborate on them. The selection is nearly random and reflects mostly my own interests among those which may some day lead to a working question-answering device:

*Predicate calculus.*—According to Reichenbach (56), the notational convention of  $f(x)$  for intransitive sentences ( $f$  = predicate,  $x$  = argument) and  $f(x,y)$  ( $y$  = second argument) for transitive sentences rests on the fact that although both can be written  $\alpha [f,x]$ , or  $\alpha [f,x,y]$  respectively, the function  $\alpha$  is a constant. I will demonstrate, however, that in fact  $\alpha$  does assume different values.

In many languages the relations of taxonomy, synonymy, and attribution (modification) are all expressed by the same surface syntax. For example, "A doctor is a professional" (taxonomy); "A doctor is a physician" (synonymy); "A doctor is well educated" (attribution). This can be restated briefly as follows: all three relations are cases of attribution. The only explanation that remains is what kind of a relation of attribution is the relation of taxonomy or why it should be considered as one. First, it is well known that the taxonomic relation is the result of taking a superordinate taxon and modifying it in some way. Thus, according to Denisson's *Loose Leaf Dictionary*, "A lion is a wild animal," or (lion)T( (animal)M (wild) ) where T is the

relation of taxonomy and  $M$  is the relation of modification. By adding attributes (modifiers) a term is created that refers to a subclass of the superordinate taxon. It can then be argued that the superordinate taxon's function is also an attribute of the taxon it governs. That is, lion not only possesses the attributes of "wildness," it also has the attributes of "animalness."

There is one more consideration before summarizing. In componential semantic analyses one encounters another kind of attributive (modification) relation. Note that in the more usual lexical modification the order is crucial especially when nouns are involved: "A dog house is not a house dog." However, in componential analysis or modification by adjectives there is little or no difference. For most purposes a "big red ball" is equivalent to "red big ball."<sup>6</sup>

The above can be restated as follows. There are basically at least four relations of predication (attribution) that differ from each other in the following manner:

1. Attribution  $\alpha(1)$ , symbolized by  $M(1)$  is
  - (1a) Reflexive:  $(x) M(1) (x)$
  - (1b) Symmetric: If  $(x) M(1) (y)$  then  $(y) M(1) (x)$ .
  - (1c) Intransitive: It is not true that if  $(x) M(1) (y)$  and  $(y) M(1) (z)$  then  $(x)M(1)(z)$ .
2. Attribution  $\alpha(2)$ , symbolized by  $M(2)$  is
  - (2a) Reflexive: (see above, 1a).
  - (2b) Asymmetric: If  $(x) M(2) (y)$  then  $(y) M(2) (x)$  may or may not be true.
  - (2c) Intransitive: (see above, 1c).
3. Synonymy  $\alpha(3)$  symbolized by  $S$  is
  - (3a) Reflexive: (see above, 1a)
  - (3b) Symmetric: (see above, 1b)
  - (3c) Transitive: If  $(x)S(y)$  and  $(y)S(z)$  then  $(x)S(z)$ .
4. Taxonomy  $\alpha(4)$  symbolized by  $T$  is
  - (4a) Reflexive (see above)
  - (4b) Asymmetric: (see above)
  - (4c) Transitive: (see above, 3c)

Many properties of lexical/semantic fields can be characterized by these four relations.

Applying this notation to McCawley's verb-first propositional form (p. 300 fn) and using  $M$  for  $M(2)$  and Polish parentheses free notation:

<sup>6</sup> Although I do not want to ignore semantic subtleties that may point to differences of the two phrases, these are at present beyond the range of semantic sophistication available in anthropology or elsewhere.

Note that the order of the arguments is changed. The first M links the verb to its object, the second links the resulting verb phrase to the subject.<sup>7</sup>

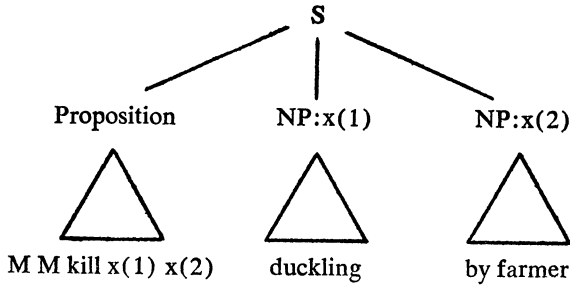


FIGURE 10.

*Compound relations.*—There are at least two reasons for discussing what I call here compound relations: (a) there is a need to show how all of Casagrande & Hale’s relations can be explained by taxonomy, synonymy and modification; and (b) some notational convention is needed to account for the speaker’s intuition that such compound relations are perceived very similarly to the primitive ones, or in other words, that lexical/semantic fields can be constructed from simple as well as from complex relations.

The procedure I want to follow is to show the nature of complex relations and a proposed notational convention first on one example and then to generalize it to others that are possible.

*Part-to-whole.*—This relation is based on a sentence frame in English roughly as follows:

“       is part of       .”  
           1                          2

There are two ways to arrive at the same result: (1) By making the above sentence into a taxonomic statement in the strict sense: “       is a (kind of)       -part.” Thus by extension the first element is also a kind of a part,  
           2

<sup>7</sup> The simplest justification of this arrangement of the relation of modification is obtained by nominalizing the above sentences three ways:

- The *killing* of the duckling by the farmer . . .
- The *duckling* killed by the farmer . . .
- The *farmer* who killed the duckling . . .

Farmer is deletable in the first two instances where it is not focal. There seems to be a general rule that a verb can be focal only in the nominalized form of a sentence, while only noun phrases can be focal in the declarative mode.

that is, the second element modifies 'part' and the first element is taxonomically (transitively) related to the second element plus part; therefore T M (part) ( $\frac{\text{---}}{2}$ ) ( $\frac{\text{---}}{1}$ ). (2) Treating 'part' as a transitive predicate leads

to the same conclusion: the second element is analogous to the object, the first to the subject; therefore, by considering transitivity, T M (part) ( $\frac{\text{---}}{2}$ ) ( $\frac{\text{---}}{1}$ ). A very similar argument can be made for most Casagrande & Hale relations (6).

Some of the Casagrande & Hale (6) relations translate into the three basic relations T, S, M (for M(2)) as follows:

*Class inclusion:* ( $\frac{\text{---}}{1}$ )T( $\frac{\text{---}}{2}$ ) or T( $\frac{\text{---}}{2}$ ) ( $\frac{\text{---}}{1}$ ) is equivalent to T M(kind) ( $\frac{\text{---}}{2}$ ) ( $\frac{\text{---}}{1}$ ).

*Spatial:* (usually of the form) " $\frac{\text{---}}{1}$  Preposition  $\frac{\text{---}}{2}$ " or M M(Preposition) ( $\frac{\text{---}}{2}$ ) ( $\frac{\text{---}}{1}$ ). Some of these 'prepositions' require dual or plural objects in position 2, e.g. M M(between) ( $\frac{\text{---}}{2}$  &  $\frac{\text{---}}{1}$ ) ( $\frac{\text{---}}{1}$ ).

*Attributive:* ( $\frac{\text{---}}{1}$ )M( $\frac{\text{---}}{2}$ ), or M( $\frac{\text{---}}{2}$ ) ( $\frac{\text{---}}{1}$ ).

*Function and Operation:* Although the precise meaning of these designations is not clear, I have argued in (74) that the function relation is that between subject and verb, therefore M; and similarly, the relation of operation between verb and object, also M. Thus the sentence "Brewers make beer" becomes M M (make) (beer) (by brewers).

*Comparison:* " $\frac{\text{---}}{1}$  is like  $\frac{\text{---}}{2}$ ," or M M(like) ( $\frac{\text{---}}{2}$ ) ( $\frac{\text{---}}{1}$ ).

*Exemplification:* " $\frac{\text{---}}{1}$  is an example of  $\frac{\text{---}}{2}$ ," or " $\frac{\text{---}}{1}$  is exemplified by  $\frac{\text{---}}{2}$ ." The first is equivalent to the taxonomic relation T, and since it is transitive is also equivalent to T M(example) ( $\frac{\text{---}}{2}$ ) ( $\frac{\text{---}}{1}$ ). (The transitivity needs amplification, i.e. "Grass is an example of green" is probably incomplete and should read fully "Grass is an example of a green object" and "Green is an example of color" is extended to "A green object is an example of a colored object." The second is the inverse of the taxonomic relation (or T).

*Grading:* This is a relation not reducible to the basic T S M relation. It involves temporal and spatial order (the symbol for this relation is Q for Queuing):

	Left to Right	Right to Left
Proximal neighborhood (immediate successor)	Q(1)	Q(2)
Distal neighborhood (eventual successor)	Q(3)	Q(4)

FIGURE 11. Queuing relation.

Where Q(1) is the inverse of Q(2) and Q(3) of Q(4) and Q(1) implies Q(3) and Q(2) implies Q(4) but not vice versa. However, before further commitment more work is needed in the explication of this relation.

*Provenience:* “\_\_\_\_\_ comes from \_\_\_\_\_,” “\_\_\_\_\_ is made of \_\_\_\_\_,” and possibly “\_\_\_\_\_ is a stage of \_\_\_\_\_” (or “matures into \_\_\_\_\_”). This looks, at this point, like variations of the queuing relation: stage and provenience are related to the distal Q(2). Apparently some more or less complex process intervenes between the first and second term. Again more work is necessary.

*Synonymy:* (\_\_\_\_\_)S(\_\_\_\_\_) or S(\_\_\_\_\_) (\_\_\_\_\_) is reserved for identity. It is possible that some other relation may be necessary for functional

identity: for example, Lounsbury type expansion/reduction rules. In the Yankee kinship system M M(\_\_\_\_\_) (uncle) (vocative) is in direct address functionally equivalent to ‘uncle.’ In other words, any modifier of uncle is dropped in direct address.

*Antonymy:* Lyons (43) lists three types: 1. Complementarity, either “\_\_\_\_\_ implies not \_\_\_\_\_,” or “\_\_\_\_\_ is equivalent to not \_\_\_\_\_” (or if not = N, unary relation, and implies = F, a binary one) F N(\_\_\_\_\_) (\_\_\_\_\_) or possibly S N (\_\_\_\_\_) (\_\_\_\_\_) . Whether the two cases are equivalent remains to be seen. 2. Antonymy par excellence, a relation not well understood at present but somehow linked to queuing: big Q(1) bigger Q(1) biggest, implies small Q(2) smaller Q(2) smallest. (3) Converseness, also somehow connected to queuing: “A sold B to C” implies “(A owned B)” Q(1) “(C owns B)” and “C bought B from A” implies “(C owns B)” Q(2) “(A owned B)” Paul Kay’s notion of ‘contrast’ (37) also sheds some light on the problem. However, much more needs to be done.

*Partitioning of the vocabulary.*—Partitioning of the vocabulary for ques-



tion-answering machines is extremely important. It accounts for the notion of context of discourse, and possibly can be extended to account for sociolinguistic partitionings and the reduction of searches to manageable proportions both in space and in time. The details will become apparent with the example in Figure 12 of a taxonomic representation: a hypothetical three-level taxonomy, highly regular and with four branches on each level in a tree representation (drawn only partially).

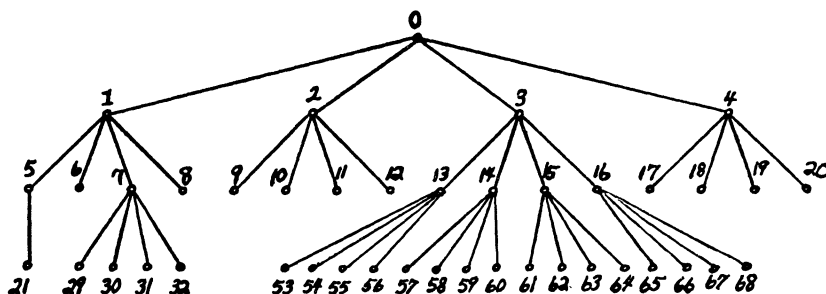


FIGURE 12. Hypothetical three-level taxonomy.

This taxonomy can be coded into a matrix. A three-level taxonomy was selected because it can be represented in three dimensions. Obviously for matrices of greater depth  $n$ -dimensional matrices will be necessary. These are simple generalizations of the three-dimensional case. The three-dimensional matrix is represented as a cube.

The rationale for this representation is to construct a space in such a way that every sibling and especially every ancestor node is an immediate neighbor (i.e. of one link distance). This is a condition of the transitivity of the relation of taxonomy.

Since node 0 dominates all subordinate nodes, it occupies an entire plane in the space (note the 0 plane in Figure 13—two rows have been excised in order to reveal part of the internal structure). The next level of the taxonomy is perpendicular to this plane and consists of the rows of 1, 2, 3, and 4, i.e. 1 T 0 (or 1 is a kind of 0), 2 T 0, 3 T 0, and 4 T 0.

The next level of the taxonomy is again perpendicular to these rows. These are the single cells—for example, 5, 6, 7, and 8 with 5 T 1, 6 T 1, 7 T 1, and 8 T 1—but since each of these is also dominated by 0 they are also perpendicular to the 0 plane. The illustration shows the terminal elements of the taxonomy being perpendicular to the element 7, i.e. 29 T 7, 30 T 7, 31 T 7, and 32 T 7. Note that the terminal nodes are also perpendicular to the 0 plane and the 1 row, as well as to element 7.

The major advantages of this representation are:

1. The transitivity of the taxonomic relation is a characteristic of the spatial arrangement of elements. In other words, syllogism, the simplest deductive capacity of the device, is part of the spatial arrangement.

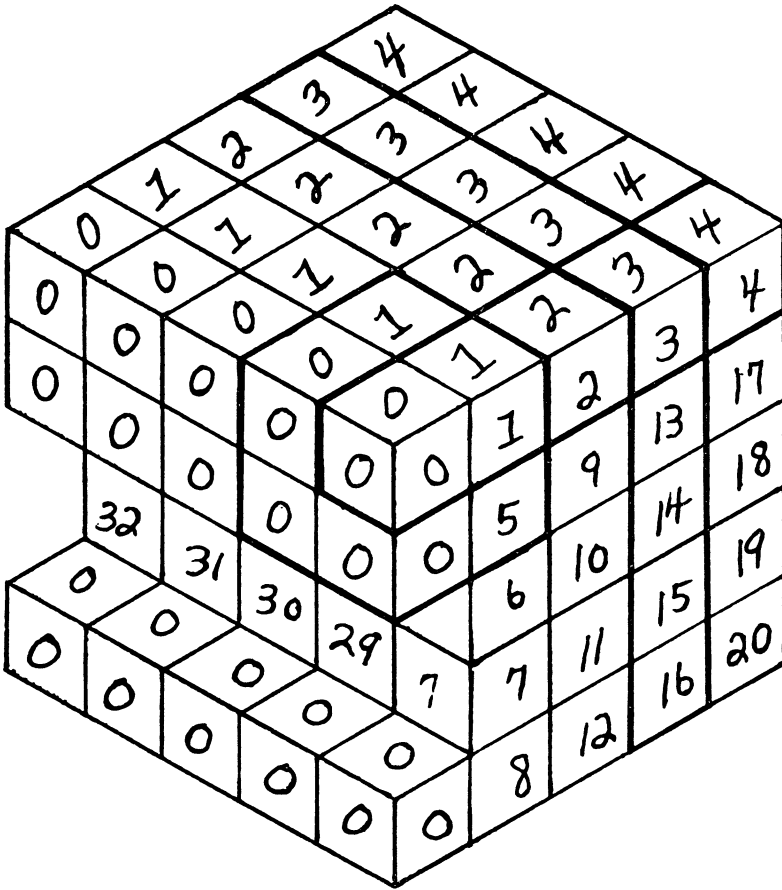


FIGURE 13. Three-dimensional matrix of three-level taxonomy.

2. Perhaps more importantly, partitioning of a large taxonomy, that is, the application of different search strategies, is simply a partitioning of the  $n$ -dimensional taxonomic space (three-dimensional in the example).

A *depth-first* search is, in the three-dimensional example, any proper sub-cube. For example, the following illustration shows the cube which represents the right-most branch of the tree in Figure 12. Searches that are depth first with a fanning out on the lower levels are sub-oblongs of the  $5 \times 5 \times 5$  taxonomic cube (see dotted line in Figure 14).

A *breadth-first* search is the row shown in Figure 15. It includes only the first level of the tree. A breadth-first subtree of two levels is the plane represented in Figure 16.

A *subtree search* on a subtree of  $n-1$  depth is a hyperplane (true plane in

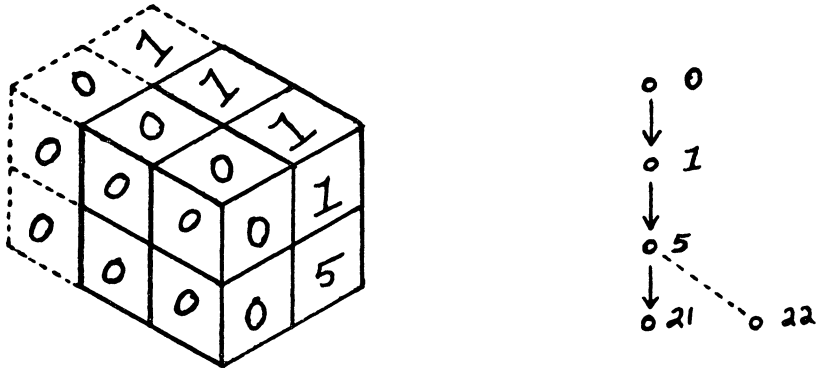


FIGURE 14. Three-dimensional *depth-first* search.

the three-dimensional illustration) parallel to the hyperplane of the top node. Figure 17 illustrates a subtree of Figure 13.

Subtrees can be found in any search as subfigures of larger figures (submatrices of larger matrices). Thus Figure 18a is a subtree of Figure 16 and Figure 18b is one of Figure 15.

The need for partitioning taxonomies can be explained in detail as follows: It is a well-known fact about human information retrieval that it is fast rather than accurate; it is not as exhaustive as computer retrievals tend to be. Frequently used items tend to be more in focal awareness and hence more easily accessible than little-used items. Items in the device are moved up to a more focal position every time they are used for retrieval or in deductions. The corner facing the viewer in Figure 13 represents such a more focal position. Thus if the branch of node 2 should become more used or of greater interest, it could be moved to the left of the branch of node 1 (that is, branch of 1 and branch of 2 would change places, and so on). This moving to the left procedure explains what happens if discourse is domain limited. For example, if branch 1 is dealing with chemistry and if the discussion is about chemistry, the term "radical" would be interpreted by its chemical sense. If the branch of node 2 deals with politics, then an exchange of the branches of

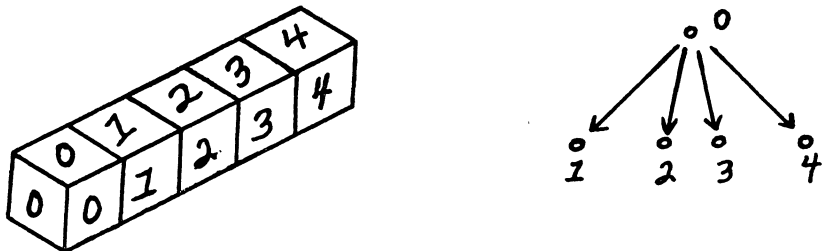


FIGURE 15. *Breadth-first* search.

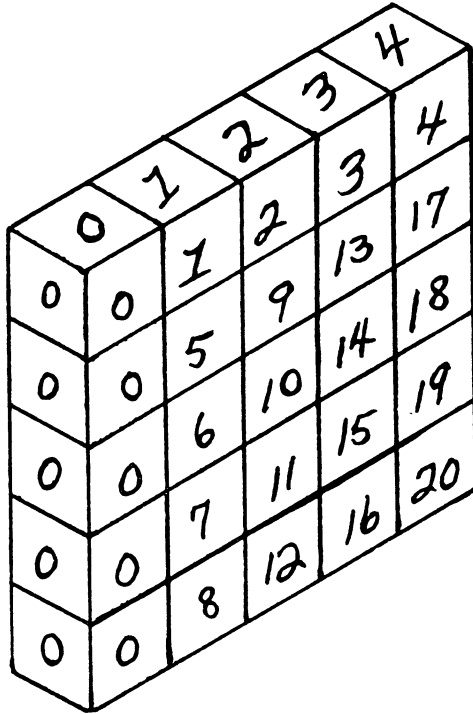
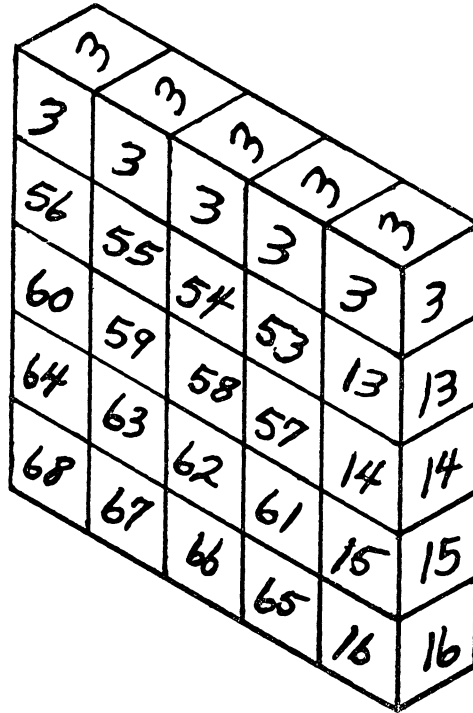


FIGURE 16. Two-level *breadth-first* search.

node 1 with the branch of node 2 brings politics into focus and gives an appropriate interpretation to the term "radical."

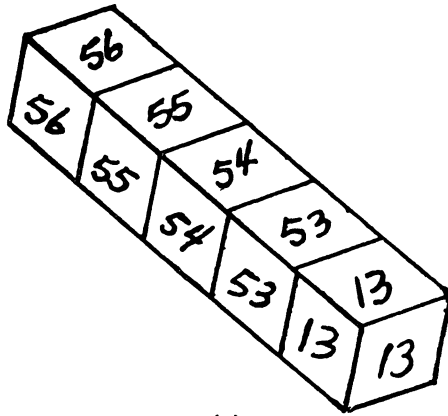
Thus partitioning of the vocabulary accounts for use of speech in contexts bound to particular isolable cultural domains. Similarly, partitioning can be extended to include sociolinguistic contexts, e.g. intimate discourse vs formal discourse. It is conceivable that one could partition along two parameters at the same time: e.g. intimate talk about the domain of sex versus formal medical talk about the same domain.

In addition, the partitioning schema accounts for the indeterminacy of the length of paraphrases in the following way: the length of a paraphrase to a particular question is determined by the maximum amount of available information, i.e. the maximum extent of the subtree dealing with a particular topic plus its associations (i.e. the rest of the lattice) and the available time to answer the question. The available time is gauged by some perceived implicit sociolinguistic criteria. The computer device, however, would have to request clarification by an explicit "impatience factor" or other such indicators. The response of the computer to the answer of the above request for a time limit is also representable as a partition of the response matrix to the most commonly used pathways, i.e. some sub-solid of the cube in Figure 13.

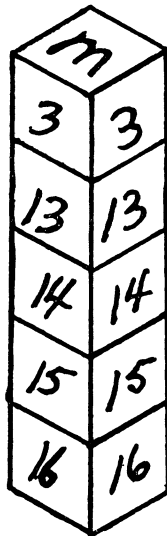
FIGURE 17. *Subtree search.*

Since the mechanisms of such subpartitioning are not very well understood, the  $n$ -dimensional hypercube representation of a taxonomy seems particularly well suited for the investigation of this phenomenon. By varying different parameters of partitioning, comparisons can be made with the behavior of human beings.

Experimental work in psychology (e.g. Mandler 45) seems to indicate that the size of the taxonomies is somewhere of the order of  $5 \pm 2$  items on each level. Assuming the maximum number, i.e. 7, the requirement is an  $8^n$  matrix. For a maximal depth of  $n = 7$   $8^7 = 2,097,152$ , which is ample for extremely large vocabularies [even considering that 262,144 (or  $8^6$ ) cells are needed for the most general term—however, what seems wasted in space is gained in speed]. A reduction of storage requirements is achievable by breaking the large taxonomies into a series of interlinked smaller ones. This could account for the fact born out by observation that genus terms and species terms are more easily accessible than intermediate and very much more general terms. In Perchonock & Werner (52) we showed that the term “food” and specifically named foods were easiest to elicit. This argues for a relatively independent subtaxonomy of “food” loosely linked to the taxonomic aspect of the Navajo universe (Werner & Begishe 70). Such an arrangement



(a)



(b)

FIGURE 18 a, b. *Subtrees.*

makes the retrieval of items in subtaxonomies like food, plants, or animals very efficient.

Attributes can be incorporated into the present schema in the following fashion: First I expand the previously given "definition" of (lion) in the following way:

"A lion is an African or Asian wild animal and a big cat."

Let us mark modifiers with a suffix M and as an edge pointing *to* the term they modify. Further we assume that every modifier with the modified repre-

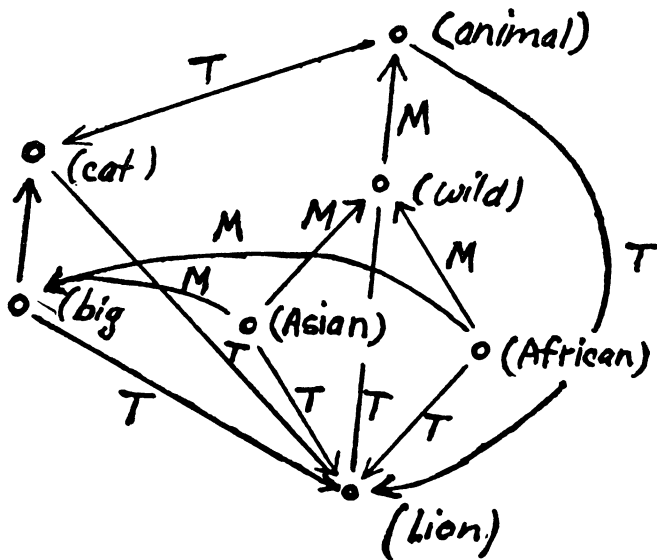


FIGURE 19. Graph of attribution relations.

sents a new taxonomic “double” node. The above sentence can be first represented as follows:

(lion) T (((animal) M (wild)) M ((African) D (Asian))) and (this could be expressed by a conjunction but that would complicate matters)  
 (lion) T ((cat) M (big)) (The truth of (cat) T (animal) assumed.)

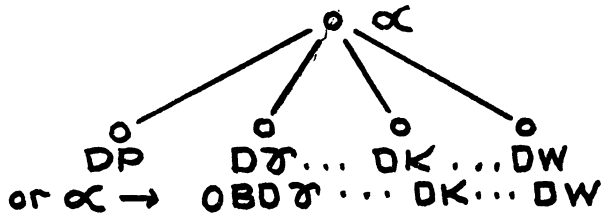
Graphically this is representable as in Figure 19.<sup>8</sup>

In order to represent the above graph as a matrix visually it must be reduced in such a way that it can be represented in three dimensions. This is accomplished by substituting (A) for (animal), (Bo) for (wild African),

<sup>8</sup> The McCawley type Polish and verb-first representation for the formulas is as follows:

M D(African)D(Asian)T M(wild) (Animal) (lion)  
 T M(big) (cat) (lion)

In the first formula D is introduced following McCawley (47) as an n-ary relation:



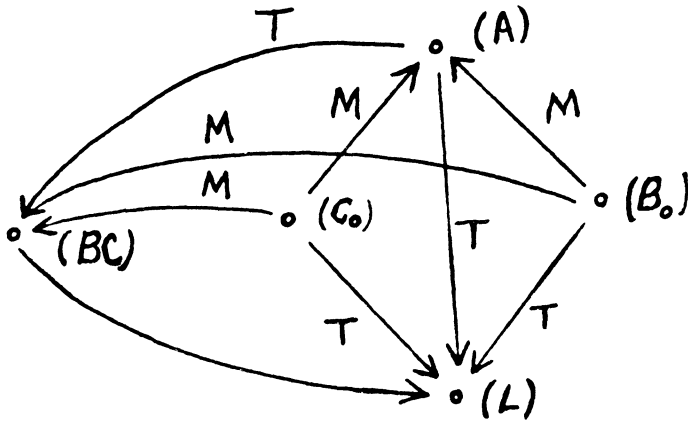


FIGURE 20. Simplified graph of attribution relations.

(Co) for (wild Asian), (BC) for (big cat) and (L) for (lion). The matrix representation of the graph in Figure 20 is shown in Figure 21.

Both (lion) and (big cat) occupy two cells. Note that following the graph, (lion) is adjacent to (animal), to (wild African animal), to (wild

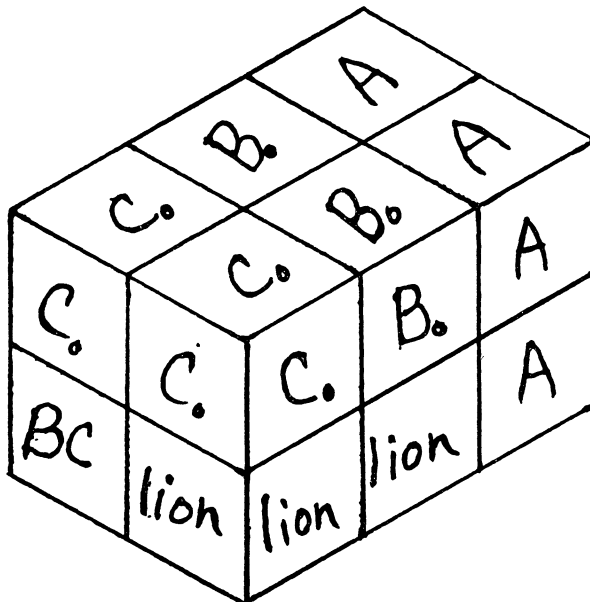


FIGURE 21. Matrix of attribution relations.



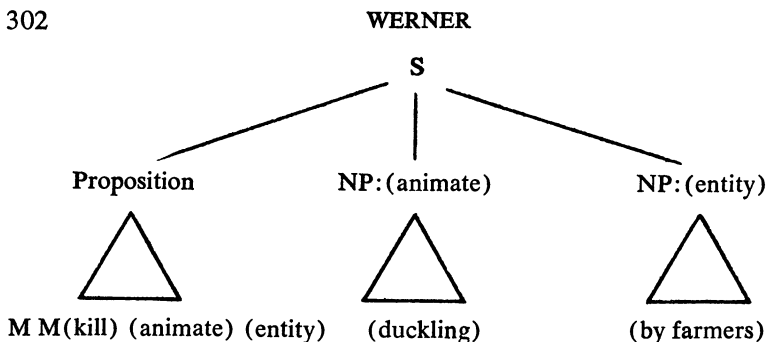


FIGURE 22.

Asian animal), to (big cat), and to (animal), etc ( o(animal) o(Bo) should be read "wild African animal."

Whether it will always be possible in large taxonomies (multidimensional matrices) to move the end points (terminal nodes of different paths) into adjacent positions remains to be seen. Possibly polysemy will become definable by the impossibility of moving two homophonous items adjacent to each other.

The above model is not only plausible but very promising for the simulation of a large number of so far unstudied properties of lexical/semantic fields.

*Variables.*—A final point concerns the use of variables. In the preceding pages we have used *x* and *y* to denote the use of variable (usually) noun phrases. I do not think that in any simulation of human lexical/semantic behavior such use is justified or necessary. It is possible to utilize the natural characteristics of taxonomies, namely that every superordinate concept can be viewed as a pro-concept (in the sense of pro-noun for pronoun). All variables in sentence frames have ranges which extend over various subtaxonomies which are dominated by some concept. It is therefore natural to consider a convention which would view all superordinate nodes as variables for any concept under its domination. Thus McCawley's example may be represented as in Figure 22.

In fact, there is no need to represent the entry for the verb "kill" in this complex manner. The noun phrases for discourse on certain topics will come from the topic itself, and as long as this topic fulfills the requirement of the superordinate pro-concepts, insertion will be possible. Thus all we need is something like *M M (kill) (animate) (entity)*. This is the full entry of the word "kill" in the context of this discussion. Lakoff (42) has argued that the above formula (or some variant of it) is not sufficient. Pronominalizations need to take into account the entire noun phrase, not just the head noun. Thus while "man" is insertable in a slot marked (animate), "dead man" cannot be. One solution is to treat (*M*) (dead) (man) as if (dead) was an operator that overrides the animateness of (man).

## SUMMARY

My emphases in this paper are the analytical aspects of ethnoscience. This is an important theoretical emphasis since no praxis is adequate without sound theoretical foundations. At present practical ethnoscience leads to better lexicography and better translation. However, our sophistication has not progressed much beyond the atomic definitions of Casagrande & Hale (6). Whether for analysis and critique of dictionary entries or for synthesis of new entries there is nothing better. The next steps appear to be much more difficult.

Relatively little has been done with the problem of individual variation. Here D'Andrade's work (11, 13) stands out. Although his approach is averaging or statistical, the relationship of his methods to the usual lexicographic concentration on the knowledge of experts is poorly understood (however, see D'Andrade 14).

Computer exploration and experimentation with large lexical/semantic fields holds hope for the illumination of both: the problem of individual variation and the problem of practical applications.

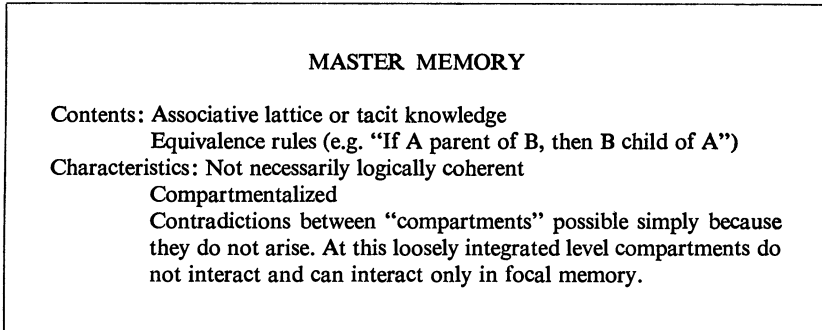
The notion of context must be further explored. Partitioning of the vocabulary was one possible technique. Better methods and better techniques for eliciting subtle contextual variations must be developed. Neither this nor further advances in ethnosemantics through the use of computerized models of human memory are conceivable without a more thorough knowledge of the psychological literature. It is certain that the problem of computer simulation of lexical/semantic fields requires the integration of more knowledge than just the four fields discussed in this paper. For example, psychological experiments seem to establish the existence of a short-term memory. However, its utility for simulation is not clear and has consequently not been discussed in any of the proposed question-answering systems to date.

A general schema of the nature of language as described in this paper can be summarized as appears on the next page.

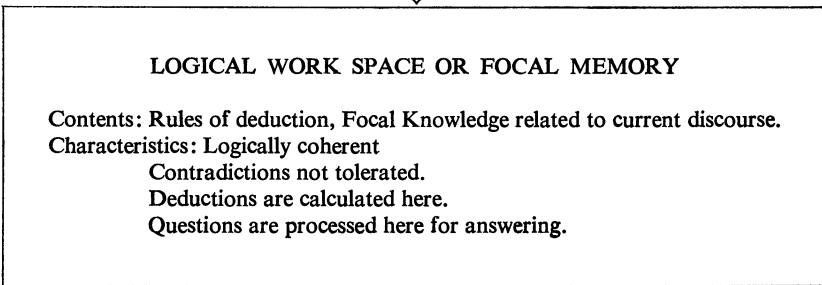
I would like to close this summary with a challenge: Kay (36, p. 30) has stated: ". . . cognitive models do not predict overt behavior. But when the cognitive model is supplied with the *information it specifies as necessary for reaching a decision* (emphasis added), it can predict overt behavior accurately." Perhaps Kay's statement is too strong. Surely not all behavior is predictable nor is all necessary information elicitable. But the basic question remains. If cognitive models can predict behavior, then where is the boundary between ethnoscience and the study of social systems?

## CONCLUSION

About 10 years ago the late Uriel Weinreich (68) said hopefully, "[In semantics] almost everything remains to be done." Some progress has been made but basically the situation remains unchanged. Eighteen months ago I (Werner 73) predicted the turn of the century as a possible target date for a working general question-answering device. I envisioned the day of a three-

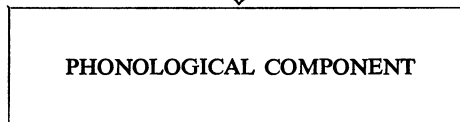


Movement of subsystems from Master Memory to Workspace by partitioning; which may be dynamic, i.e. varying as speech act or discourse proceeds. Movement makes tacit knowledge focal.



P(0)  
·  
·  
·  
·  
·  
P(n)

Domain of TRANSFORMATIONS



**ACTUAL SENTENCES**

way interaction at a computer console between an informant, an anthropologist, and a machine that tries to mimic the informant's verbal behavior. Such interaction will be instructive to the anthropologist as well as the informant, and will allow for improvement of the internal workings of the machine and hence anthropological theory. For the first time it will make falsification of anthropological (or ethnoscience) hypotheses possible. That will revolutionize anthropology. However, I am less optimistic today than I was 18 months ago. Almost everything still remains to be done.

## LITERATURE CITED

1. Berlin, B., Breedlove, D. E., Raven, P. H. 1968. Covert categories and folk taxonomies. *Am. Anthropol.* 70:290-99
2. Bierman, A. K. 1964. *Logic: A Dialogue*. Holden Day
3. Binnick, R. I. 1969. *An Application of an External Generative Semantic Model of Language to Man-machine Interaction*. Presented at Int. Conf. Comput. Ling., Sanga-Saby, Sweden
4. Biss, K., Chien, R., Stahl, F. 1971. R2—A Natural Language Question Answering System. *AFIPS Joint Spring Computer Conf.* 1971:303-8
5. Burling, R. 1964. Cognition and componential analysis: God's truth or hocus pocus? *Am. Anthropol.* 1966:20-28. See also Ref. 67, 419-28
6. Casagrande, J. B., Hale, K. L. 1967. Semantic relations in Pappo folk definitions. See Ref. 32, 165-96
7. Chomsky, N. 1965. *Aspects of the Theory of Syntax*. MIT Press
8. Colby, B. N. 1966. Ethnographic semantics: A preliminary survey. *Curr. Anthropol.* 7:3-32
9. Cowan, J. L. 1970. *Studies in Thought and Language*. Univ. Arizona Press
10. Dalton, G., Ed. 1971. *Studies in Economic Anthropology*. Am. Anthropol. Assoc.
11. D'Andrade, R. G. 1970. *Culture Shared and Unique*. Presented at 69th Ann. Meet. Am. Anthropol. Assoc., San Diego
12. D'Andrade, R. G. 1965. Trait psychology and componential analysis. *Am. Anthropol.* 67:215-28
13. D'Andrade, R. G. 1971. Symbols, Memory and Behavior. In manuscript
14. D'Andrade, R. G. 1972. A Propositional Analysis of U.S. American Beliefs About Illness. In manuscript
15. Dreyfus, H. L. 1972. *What Computers Can't Do*. Harper & Row
16. Durbin, M. 1970. Cognitive Anthropology. In press
17. Filmore, D. J., Langendoen, D. T. 1971. *Studies in Linguistic Semantics*. Holt, Rinehart & Winston
18. Fischer, A. 1970. Current directions in anthropology. *Am. Anthropol. Bull.* 3(2) No. 3
19. Fishman, J. A. 1971. *Sociolinguistics: A Brief Introduction*. Newbury House
20. Frake, C. O. 1964. Notes and queries in ethnography. *Am. Anthropol.* 66(2) No. 3:132-45. See also Ref. 67, 123-36
21. Friedrich, P. 1971. *Anthropological Linguistics: Recent Research and Immediate Prospects. Proc. 22nd Round Table Conf. Lang. Ling.* Georgetown Univ. Press
22. Garvin, P. L., Ed. 1970. *Cognition: A Multiple View*. Spartan Books
23. Garvin, P. L., Ed. 1971. *Method and Theory in Linguistics*. Mouton
24. Geoghegan, W. H. 1971. Information processing systems in culture. See Ref. 35, 3-35
25. Gladwin, H., Gladwin, C. 1971. Estimating market conditions and profit expectations of fish sellers at Cape Coast, Ghana. See Ref. 10
26. Gladwin, T., Sturtevant, W. C. 1962. *Anthropology and Human Behavior*. Washington, D.C.: Anthropol. Soc.
27. Greenberg, J. H., Ed. 1963. *Universals of Language*. MIT Press
28. Hammel, E. A. 1965. Formal semantic analysis. *Am. Anthropol.* 67(2) No. 5
29. Hymes, D. 1962. The ethnography of speaking. See Ref. 26, 13-53
30. Hymes, D. 1964. Directions in (Ethno.) Linguistic Theory. *Am. Anthropol.* 66(2) No. 3:6-56
31. Hymes, D. 1971. Linguistic method in ethnography: Its development in the United States. See Ref. 23, 249-325
32. Hymes, D., Bittle, W. E., Eds. 1967. *Studies in Southwestern Ethnolinguistics*. Mouton
33. Katz, J., Postal, P. M. 1964. *An Integrated Theory of Linguistic Description*. MIT Press
34. Kay, P. 1966. Comment on Colby. *Curr. Anthropol.* 7:21
35. Kay, P., Ed. 1971. *Explorations in Mathematical Anthropology*. MIT Press
36. Kay, P. 1970. Some theoretical implications of ethnographic semantics. See Ref. 18, 19-31

37. Kay, P. 1971. Taxonomy and semantic contrast. *Language* 47: 866-87
38. Keesing, R. M. 1972. Paradigms Lost: The New Ethnography and the New Linguistics. In manuscript
39. Lakoff, G. 1967. *Stative Adjectives and Verbs in English*. In Harvard Comput. Lab. Rep. No. 17:1-1-1-16
40. Lakoff, G. 1971. Introduction to linguistics and natural logic. In *Handbook of Cognitive Psychology*, ed. J. Mehler. In press
41. Lakoff, G. 1971. On generative semantics. See Ref. 65, 232-96
42. Ibid. Presuppositions and relative well-formedness, 329-40
43. Lyons, J. 1968. *Introduction to Theoretical Linguistics*. Univ. Cambridge Press
44. Maclay, H. 1971. Linguistics: Overview. See Ref. 65, 157-82
- 44a. Malinowski, B. 1965. *The Language of Magic and Gardening*. Indiana Univ. Press
45. Mandler, G. 1970. Words, lists, and categories: An experimental view of organized memory. See Ref. 9, 100-31
46. McCawley, J. D. 1970. *Syntactic and Logical Arguments for Semantic Structures*. Proc. 5th Int. Seminar Theoret. Ling. Tokyo: TEC
47. McCawley, J. D. 1971. A Programme For Logic. In manuscript
48. McCawley, J. D. 1971. Where do nounphrases come from? See Ref. 65, 217-31
49. Minsky, M. 1968. *Semantic Information Processing*. MIT Press
50. Naroll, R., Ed. 1972. *Directions in Anthropological Theory*. In preparation
51. Naroll, R., Cohen, R., Eds. 1970. *Handbook of Method in Cultural Anthropology*. Natural History Press
52. Perchonock, N., Werner, O. 1968. Navajo systems of classification: The domains of foods. *Ethnology* 8:229-42
53. Postal, P. M. 1971. On the surface verb "remind." See Ref. 17, 180-272
54. Quillian, M. R. 1968. Semantic memory. See Ref. 49, 216-70
55. Quillian, M. R. 1969. The teachable language comprehender: Simulation program and theory of language. *Commun. ACM* 12: 459-76
56. Reichenbach, H. 1947. *Elements of Symbolic Logic*. Macmillan
57. Romney, A. K., D'Andrade, R. G. 1964. Transcultural studies in cognition. *Am. Anthropol.* 66(2) No. 3
58. Ibid. Cognitive aspects of English kin terms, 146-70
59. Ruesh, J., Bateson, G. 1951. *Communication: The Social Matrix of Psychiatry*. Norton
60. Sanday, P. R. 1968. The 'psychological reality' of American-English kinship terms: An information-processing approach. *Am. Anthropol.* 70:508-23
61. Schwartz, R. M., Burger, J. F., Simmons, R. F. 1970. A deductive question-answering system for natural language inference. *Commun. ACM* 13:167-83
62. Sebeok, T. A., Ramsay, A. 1972. *Trends in Linguistics*, Vol. 12. In press
63. Simmons, R. F. 1970. Natural language: Question-answering systems: 1969. *Commun. ACM*: 15-30
64. Slagle, J. R. 1971. *Artificial Intelligence: The Heuristic Programming Approach*. McGraw Hill
65. Steinberg, D. D., Jakobovits, L. A., Eds. 1971. *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*. Cambridge Univ. Press.
66. Sturtevant, W. C. 1964. Studies in ethnoscience. *Am. Anthropol.* 66(2) No. 3:99-131
67. Tyler, S. A., Ed. 1969. *Cognitive Anthropology*. Holt, Rinehart & Winston
68. Weinreich, U. 1963. On the semantic structure of language. See Ref. 27, 114-71
69. Werner, O. 1969. *On the Universality of Lexical/Semantic relations*. Presented at 68th Meet. Am. Anthropol. Assoc.
70. Werner, O., Begishe, K. Y. 1970. *The Taxonomic Aspect of the Navajo Universe*. Proc. 39th Congr. Int. Americanistas, Lima, Peru. In press
71. Werner, O. 1970. Cultural knowledge, language and world view. See Ref. 22, 155-76
72. Werner, O. 1972. *On the Limits of Social Science Theory in Anthropology and Linguistics: Essays in*

- honor of C. F. Voegelin*. In preparation
73. Werner, O. 1972. Structure, anthropology, Lévi-Strauss and ethnoscience. See Ref. 50
74. Werner, O., Fenton, J. 1970. Method and theory in ethnoscience or ethnoepistemology. See Ref. 51, 537-80
75. Werner, O., Hagedorn, W., Roth, G., Schepers, E., Uriarte, L. 1972. New developments in ethnoscience. See Ref. 62
76. Werner, O. 1972. *On the Structure of Ethnoscience*. Presented at Conf. Methods Struct. Anal., Univ. North Carolina, Chapel Hill, 1972