II. LEXICAL ANALYSIS RESEARCH PROJECT

1. Introduction

The research of The Ohio State University Lexicology Project is directed toward the general question of the treatment of lexical information in a transformational generative grammar. Workers on the project include Charles J. Fillmore, Sandra S. Annear, Dale E. Elliott, and P. Gregory Lee. D. Terence Langendoen and James T. Heringer, associated primarily with the Mathematical Linguistics Project under the same NSF contract, have closely cooperated with the work of the Lexicology Project, as have several members of the Ohio State University Linguistics staff--George Landon in particular--and a number of graduate students majoring or minoring in Linguistics.

The issues of generative lexicology are formulated differently depending on whether syntax is correctly viewed as generative, with semantics having a purely interpretive role, or whether semantic structures provide input to the syntactic component, syntax being interpretive.

If syntax is generative, the issues are these: Given grammatical descriptions of (deep structure) sentences above the level of words, what can be said about the lexical items that can occur in the various positions in these sentence types, what grammatical and semantic properties do these words contribute to the sentences, and what is the nature of the semantic theory which is capable of interpreting sentences given all of this information.

On the other hand, if syntax is interpretive, questions of lexicology concern the nature of semantic primitives, the apparatus for generating semantic deep structures (if it makes sense to speak of semantics as being generative), the nature of the lexical substitution rules which replace (possibly quite large) segments of the semantic deep structures by lexical items, constraints on the formal relations between deep structures and surface structures, as well as the general question of whether this approach eliminates all of the interpretive role of semantics, or only part.

The results of the project have turned out to be quite compatible with the position that syntax is interpretive rather than generative, but since most of our work has been framed from the interpretive-semantics point of view, that approach is presupposed--for matters of convenience and consistency only--in the descriptions which follow.

The base component of a generative grammar is the component which characterizes the underlying skeletal forms of the grammatical sentences in the languages. The conditions on lexical insertion, and the ways in which they are formulated, obviously depend on the exact nature of the rules of the base.

If the base component does no more than provide configurations of major grammatical categories (i.e., if it contains only branching rules of the type that has come to be called "phrase structure rules"), then lexical insertion will involve (at least intrinsic) ordering and sensitivity to context (in the "transformational" sense). On the other hand, if the base component contains, in addition to the branching rules, devices for introducing lexical features, the process of lexical insertion can be context free and unordered. In this case, the introduction of "complex symbols" (clusters of lexical features linked with each lexical category) will have to be managed by a process that is ordered and transformational. Furthermore, if the rules for constructing complex symbols are designed in such a way as to require the full specification of every lexical feature, then lexical insertion requires nothing more than "feature matching": the features contained in the complex symbol at hand must be matched, one by one, with features associated with the lexical item. If, however, the feature-introducing rules provide only those features relevant to the selection of neighboring lexical items, the process of lexical insertion will have to have associated with it a "distinctness convention": a lexical item may serve as an instance of a complex symbol as long as no feature associated with the lexical item is distinct from (i.e., contradicts) a feature contained in the complex symbol.

Of the two approaches to lexical insertion--depending on the "location" of the ordered and context-sensitive processes--the latter seems to require a great deal of redundant information. It is necessary, in other words, to construct a complex symbol which registers all the relevant environmental information, and it is also necessary to identify this same information as properties of verbs which require insertion in each such environment. The same facts, in other words, are introduced twice, once in a purely automatic way. If this is the only difference between the two main approaches--and that is something we do not yet know--we shall probably choose the approach by which lexical insertion is itself a context-sensitive process.

Once inserted, a lexical item brings along with it, so to speak, certain inherent properties that are relevant to the operation of other grammatical rules, or to the insertion of other lexical items. Nouns classified as "mass" nouns, for example, do not tolerate a as an indefinite determiner, while a singular "non-mass" noun does. (Thus, a bottle is acceptable, but a milk is not.) The verb murder is used appropriately only when both subject and object nouns are capable of denoting human beings. Nouns like boy and uncle are marked as inherently "human," nouns like cat and petunia as inherently "non-human". (Thus, the boy murdered my uncle is acceptable, but the petunia murdered the cat is not.)

Other inherent lexical properties of words are the idiosyncratic properties, in particular those that relate to the applicability of syntactic rules. There is a rule in English which converts (the structures underlying) active transitive sentences into their passive counterparts. This rule can be stated most generally in terms of the placement of noun-phrases, modification of the verbal expression, etc. It happens, however, that some sentences which satisfy the conditions for the passive transformation in their categorial form, nevertheless cannot be made passive; and that the structures underlying certain other sentences must be expressed only as passive sentences. Constraints on the applicability

of the passive rule seem to be most clearly associated with the main verb. Hence, in some way it must be made explicit that verbs like have and resemble do not permit passivization, while some other verbs (in certain constructions) must be marked as appearing passively only. (Thus, while Martha resembles a horse and John has three dollars are acceptable, *a horse is resembled by Martha and *three dollars are had by John are not; on the other hand, while the Egyptians are said to have worshiped the cat is acceptable, its expected active counterpart *someone says the Egyptians to have worshiped the cat is not.) The theoretical issue that must be faced in dealing with these facts is whether these properties of lexical items should be stated in a way which makes explicit reference to specific rules (the rules they require, tolerate, or disallow), or whether the rules should themselves be formulated in such a way as not to work, or to work obligatorily, when the lexical environment is of one kind or another.

The alternative which allows lexical features to refer to specific rules requires the elaboration of a full theory of irregularity of the kind suggested by Lakoff.[1] Not to allow this

[1] George Lakoff, "On the Nature of Syntactic Irregularity", Math. Linguistics and Automatic Translation, Computation Laboratory of Harvard University Report No. NSF-16, Dec., 1965.

is either to list exceptions with each rule (a procedure which is effectively the same as the former, but one which does not require--or rather, which does not admit the need for--a theory of exceptions); or to assume that a careful statement of the conditions of application for each rule can be accomplished which will in fact have the correct results but which refers only to independently justified properties of words. This last alternative might be based on the assumption that the speakers of a language do not learn, with the rules that they learn, idiosyncratic facts about the words that govern them, nor do they learn, with the words that they learn, idiosyncratic facts about the rules that these govern. Speakers

of a language, under this view, learn general conditions on the application of rules, conditions which refer only to properties of lexical items that are justified and discoverable on independent grounds.

If this third position is taken as a philosophy of research, one may still list exceptions to rules, but such a list is to be viewed as evidence that the full range of facts about the matter at hand is not yet known, and that a more careful analysis will permit a reduction in the list of exceptions; or that, just in case no further improvements are in fact possible, language is at least partly unsystematic after all. This position is illustrated by by Chomsky's (factually incorrect) claim that the passive transformation can be limited to those verbs capable of taking manner adverbs.[2]

---

[2]Noam Chomsky, Aspects of the Theory of Syntax, M.I.T. Press, 1965, pp. 103-106.

Lexical information that is relevant to the semantic interpretation of sentences involves many issues. We may wish to be able to distinguish, for example, those aspects of the semantic structure of a word which represents what the word directly "asserts" (when used as a predicate) from those which specify the conditions of appropriateness for its use. We may further wish to discover principles for distinguishing between the "meaning" of a word and the properties of the real-world object which the word "names": we need to consider the difference, in short, between a dictionary and an encyclopedia. The formal nature of semantic "features" needs to be determined, in some way which makes it possible to escape the one-term predicate form implicit in some of the earlier work in semantic theory.

Concerning the difference between what a predicate asserts and what it presupposes, we might with to separate out the component 'unmarried' as the asserted meaning of bachelor and state that the components 'male, adult, human' identify what is "presupposed" whenever the word is used appropriately. The sentence Harry is a

bachelor informs its hearers that Harry is unmarried, but it is only used appropriately if Harry is a human male adult.

The difficulty in distinguishing lexical from encyclopedic information may be seen in connection with the definition of a word like frog. The question is whether it is possible to distinguish between statements about what the word means and statements about what frogs are like. Traditional notions of "definition" seem to presuppose that the discovery of the criterial aspect of the meaning of a word is identical with the discovery of either some diagnostic criterion for identifying the associated object, or the "essential characteristics" of that object or object type. (A diagnostic definition of man is exemplified by the traditional pair of criteria, featherless biped. The definition fails, as is well-known, because of plucked chickens, kangaroos, and amputees. An "essential characteristic" definition of man is exemplified by the various attempts to associate man with tool and language using, etc.) For words that denote real world objects, it is not at all clear whether there is a distinction to be made between lexical and encyclopedic information, and that is another way of saying that it is unclear whether one should speak of a "definition" of a word like frog at all. What is clear is that lexicologists are not zoölogists.

Semantic features associated with words may be analyzed according to their predicate type. One can go fairly far in applying concepts from the calculus of relations to ordinary language vocabulary. Thus if we assert that the English word brother (rather the term is a brother to) is transitive, irreflexive and meso-symmetric, we know that conclusions that one can draw from a relation that bears these particular properties apply equally well to formal and natural language terms. Concerning the relational nature of semantic properties of words, we may wish to show how certain words in natural language represent "derelativized" versions of inherently relational terms. The concept "unmarried," for example, although it appears in the guise of a property term, is

in reality a relational term with one of the variables unexpressed. (To say that A is unmarried is to say that there is no B such that A is married to B.) The analysis of ordinary language vocabulary requires, in fact, a rather rich and detailed description of the use of unexpressed variables: to illustrate this with concepts involving marriage, we may note that brother-in-law means male sibling of a spouse or male spouse of a sibling, where expressions of the type A is B's brother-in-law do not mention the intervening spouse or sibling; or adjectives like divorced or betrothed do not make explicit the (past or future) time or the other party to the marriage.

The nature of a semantic theory capable of interpreting sentences on the basis of the grammatical organization of its components and the lexical information associated with each of its words is a matter that must await the solution of some of these more essential problems. We know for sure that the treatment of semantic properties as one-term predicates, nurtured by the use of pluses and minuses in current notations for semantic features, is quite wrong. We know that the rule for projecting semantic characterizations of lexical items into the semantic reading of a sentence requires the elaboration of much more subtly arranged relationships among the parts of a sentence than those provided by the small number of definable syntactic relations formulatable within the current received version of transformational grammar. And we know furthermore that there are a great many semantic facts that cannot be related either to the properties of individual words (taken one at a time) or to well-understood syntactic constructions. These facts include the system of presuppositions of sentences that are most directly related to what appear to be properties of clauses, such as that of the "counterfactual" clause in a conditional sentence.

2. Activities Connected With the Project

The work of the project began with the part-time employment of Linguistics graduate student Sandra S. Annear in the winter quarter of 1967. Her research at that time was concerned with the

general role of the lexicon in a generative grammar. She was joined in the spring quarter by Dale E. Elliott, P. Gregory Lee, and James T. Heringer.* Charles J. Fillmore joined the project

*primarily associated with the Mathematical Linguistics project

on a full-time basis in the summer. During the ten weeks of the summer quarter, the members of the project and several other members of the staff and student body met every Wednesday and Thursday morning for reports and discussions on a large number of topics relating to semantic theory and lexicology. A partial list of the topics discussed, with the name of the reporter given in parentheses after each title, follows.

The general problem of lexicography in generative grammar (Annear)

Meaning vs. entailment vs. presupposition (Fillmore)

Semantic vs. encyclopedic information, meaning vs. "use" conditions (Lee)

Bendix's application of Weinreichian semantics to verbs semantically related to 'have' (Fillmore)

Reichenbach's application of concepts from the calculus of relations to natural language vocabulary (Heringer)

Bierwisch's treatment of spatial extent adjectives (John Diskin)

Bach's treatment of nouns as predicates (Langendoen)

The conjunction source of relative clauses (Annear)

Semantic selection vs. semantic projection (Langendoen)

The possible relevance of general semantics to lexicography (Elliott)

The Indiana University project on the classification of verbs (Landon)

Componential analysis in the analysis of kinship systems (Landon)

Syntactic irregularity and Lakoff's theory of exceptions (Heringer)

be, with, and have in absolute constructions in English (Lee)

Participants' papers completed at the time of this report follow.

Charles J. Fillmore

December 1967