

THE STRUCTURE OF A SEMANTIC THEORY

JERROLD J. KATZ

JERRY A. FODOR

Massachusetts Institute of Technology

1. Introduction. This paper¹ does not attempt to present a semantic theory of a natural language, but rather to characterize the form of such a theory. A semantic theory of a natural language is part of a linguistic description of that language. Our problem, on the other hand, is part of the general theory of language, fully on a par with the problem of characterizing the structure of grammars of natural languages. A characterization of the abstract form of a semantic theory is given by a metatheory which answers such questions as these: What is the domain of a semantic theory? What are the descriptive and explanatory goals of a semantic theory? What mechanisms are employed in pursuit of these goals? What are the empirical and methodological constraints upon a semantic theory?

The present paper approaches the problem of characterizing the form of semantic theories by describing the structure of a semantic theory of English. There can be little doubt but that the results achieved will apply directly to semantic theories of languages closely related to English. The question of their applicability to semantic theories of more distant languages will be left for subsequent investigations to explore. Nevertheless, the present investigation will provide results that can be applied to semantic theories of languages unrelated to English and suggestions about how to proceed with the construction of such theories.

We may put our problem this way: What form should a semantic theory of a natural language take to accommodate in the most revealing way the facts about the semantic structure of that language supplied by descriptive research? This question is of primary importance at the present stage of the development of semantics because semantics suffers not from a dearth of facts about meanings and meaning relations in natural languages, but rather from the lack of an adequate theory to organize, systematize, and generalize these facts. Facts about the semantics of natural languages have been contributed in abundance by many diverse fields, including philosophy, linguistics, philology, and psychology. Indeed, a compendium of such facts is readily available in any good dictionary. But at present the superabundance of facts obscures a clear view of their interrelations, while such theories as have been proposed to account for the facts have, in general, been either too loosely formulated or too weak in explanatory and descriptive power to succeed.

2. The projection problem. A full synchronic description of a natural language is a grammatical and semantic characterization of that language (where the term 'grammatical' is construed broadly to include phonology, phonemics,

¹ This work was supported in part by the U. S. Army Signal Corps, the Air Force Office of Scientific Research, and the Office of Naval Research; and in part by the National Science Foundation (Grant G-13903).

morphology, and syntax). Hence, a semantic theory must be constructed to have whatever properties are demanded by its role in linguistic description. Since, however, the goals of such description are reasonably well understood and since, in comparison to semantics, the nature of grammar has been clearly articulated, we may expect that by studying the contribution that semantics will be required to make to a synchronic description of a language we can clarify the subject, the form of generalizations, the goals, and the empirical and methodological constraints upon a semantic theory.

A fluent speaker's mastery of his language exhibits itself in his ability to produce and understand the sentences of his language, INCLUDING INDEFINITELY MANY THAT ARE WHOLLY NOVEL TO HIM (i.e. his ability to produce and understand ANY sentence of his language²). The emphasis upon novel sentences is important. The most characteristic feature of language is its ability to make available an infinity of sentences from which the speaker can select appropriate and novel ones to use as the need arises. That is to say, what qualifies one as a fluent speaker is not the ability to imitate previously heard sentences but rather the ability to produce and understand sentences never before encountered. The striking fact about the use of language is the absence of repetition: almost every sentence uttered is uttered for the first time. This can be substantiated by checking texts for the number of times a sentence is repeated. It is exceedingly unlikely that even a single repetition of a sentence of reasonable length will be encountered.

A synchronic description of a natural language seeks to determine what a fluent speaker knows about the structure of his language that enables him to use and understand its sentences. Since a fluent speaker is able to use and understand any sentence drawn from the INFINITE set of sentences of his language, and since, at any time, he has only encountered a FINITE set of sentences, it follows that the speaker's knowledge of his language takes the form of rules which project the finite set of sentences he has fortuitously encountered to the infinite set of sentences of the language. A description of the language which adequately represents the speaker's linguistic knowledge must, accordingly, state these rules. The problem of formulating these rules we shall refer to as the projection problem.

This problem requires for its solution rules which project the infinite set of sentences in a way which mirrors the way that speakers understand novel sentences. In encountering a novel sentence the speaker is not encountering novel elements but only a novel combination of familiar elements. Since the set of sentences is infinite and each sentence is a different concatenation of morphemes, the fact that a speaker can understand any sentence must mean that the way he understands sentences which he has never previously encountered is compositional: on the basis of his knowledge of the grammatical properties and the meanings of the morphemes of the language, the rules which the speaker knows enable

² There are exceptions, such as sentences with technical words that the speaker does not know and sentences too long for the speaker to scan in his lifetime. But these exceptions are of no systematic importance. Analogously, a person's mastery of an algorithm for propositional calculus can be said to exhibit itself in his ability to decide mechanically whether ANY well-formed formula of propositional calculus is a tautology, even though some well-formed formulae are too long for human processing, etc.

him to determine the meaning of a novel sentence in terms of the manner in which the parts of the sentence are composed to form the whole. Correspondingly, we can expect that a system of rules which solves the projection problem must reflect the compositional character of the speaker's linguistic knowledge.³

3. Synchronic linguistic description minus grammar equals semantics. A description of a natural language is, inter alia, a solution to the projection problem for that language. If we are to discover the goals of semantics by subtracting from the goals of a description of a language whatever the grammar contributes to the solution of the projection problem, we must consider the respect in which a grammar is a solution for the grammatical aspect of the projection problem.

Grammars answer the question: What does the speaker know about the phonological and syntactic structure of his language that enables him to use and understand any of its sentences, including those he has not previously heard? They do so by providing rules which generate the sentences of the speaker's language. In particular, these rules generate infinitely many strings of morphemes which, though they are sentences of the language, have never been uttered by speakers.⁴ Moreover, a grammar generates the sentences which a speaker is, in principle, capable of understanding in such a way that their derivations provide their structural descriptions. Such descriptions specify the elements out of which a sentence is constructed, the grammatical relations between these elements and between the higher constituents of the sentence, the relations between the sentence and other sentences of the language, and the ways the sentence is syntactically ambiguous together with an explanation of why it is ambiguous in these ways. Since it is this information about a novel sentence which the speaker knows and which enables him to understand its syntactic structure if and when he encounters the sentence, an adequate transformational grammar of a language PARTIALLY solves the projection problem for the language.

A semantic theory of a language completes the solution of the projection problem for the language. Thus, semantics takes over the explanation of the

³ A solution to the projection problem is certainly less than a full theory of speech. In particular, it does not provide a theory of speech production (or recognition). The difference between a description of a language and a theory of speech production is the difference between asking for a characterization of the rules of language which a speaker knows and asking for an account of how he actually applies those rules in speaking. Some things left out by the first theory but not by the second are considerations of the psychological parameters of speech production (e.g. limitations of immediate memory, level of motivation), and developmental accounts of the way the child becomes a fluent speaker (by conditioning? by the exploitation of innate mechanisms? by some combination of innate endowment and learning?). Though such problems concerning speech production lie outside the scope of a theory of a language, such a theory is essential to a theory of speech production. It is first necessary to know WHAT is acquired and used before it is sensible to ask HOW it is acquired and used.

⁴ This conception of grammar is due to Chomsky. Cf. *Syntactic structures*² ('s-Gravenhage, 1962); 'Three models for the description of language', *I.R.E. transactions on information theory*: Vol. IT-2, Proceedings of the Symposium on Information Theory (Sept. 1956). For a bibliography on transformational grammar, see Chomsky, 'On the notion "rule of grammar"', *Structure of language and its mathematical aspects*: Proceedings of symposia in applied mathematics 12.16 fn. 24 (1961).

speaker's ability to produce and understand new sentences at the point where grammar leaves off. Since we wish to determine, when we have subtracted the problems in the description of a language properly belonging to grammar, what problems belong to semantics, we must begin by gaining some grasp of how much of the projection problem is left unsolved by an optimal grammar.

One way to appreciate how much of understanding sentences is left unexplained by grammar is to compare the grammatical characterizations of sentences to what we know about their semantic characterizations. If we do this, we notice that the grammar provides identical structural descriptions for sentences that are different in meaning and different structural descriptions for sentences that are identical in meaning. The former will be the case for all morphemically distinct substitution instances of a given sentential type; for example, *The dog bit the man* and *The cat bit the woman*. The latter will be the case for many instances of sentential synonymy; for example, *The dog bit the man* and *The man was bitten by the dog*.⁵

In general, it is obvious that in no sense of meaning does the structural description which the grammar assigns to a sentence specify either the meaning of the sentence or the meaning of its parts. Such considerations must now be made precise in order that we may apply our formula 'linguistic description minus grammar equals semantics' to determine a lower bound on the domain of a semantic theory. Later in this section we will fix an upper bound by determining what problems lie outside the concerns of a complete linguistic description.

Grammars seek to describe the structure of a sentence IN ISOLATION FROM ITS POSSIBLE SETTINGS IN LINGUISTIC DISCOURSE (WRITTEN OR VERBAL) OR IN NON-LINGUISTIC CONTEXTS (SOCIAL OR PHYSICAL). The justification which permits the grammarian to study sentences in abstraction from the settings in which they have occurred or might occur is simply that the fluent speaker is able to construct and recognize syntactically well-formed sentences without recourse to information about settings, and this ability is what a grammar undertakes to reconstruct. Every facet of the fluent speaker's linguistic ability which a grammar reconstructs can be exercised independently of information about settings: this is true not only of the ability to produce and recognize sentences but also of the ability to determine syntactic relations between sentence types, to implicitly analyze the syntactic structure of sentences, and to detect grammatical ambiguities. Since, then, the knowledge that a fluent speaker has of his language enables him to determine the grammatical structure of any sentence without reference to information about setting, grammar correspondingly forms an independent theory of this independent knowledge.

We may generalize to arrive at a sufficient condition for determining when an ability of speakers is the proper subject matter of a synchronic theory in linguistics. The generalization is this: IF SPEAKERS POSSESS AN ABILITY THAT ENABLES THEM TO APPREHEND THE STRUCTURE OF ANY SENTENCE IN THE INFINITE

⁵ Moreover, sentences that receive the same structural description may differ in that one is semantically ambiguous or anomalous but the other is not. Compare *The bill is large*, *The paint is silent*, and *The street is wide*, all of which receive the same structural description from the grammar.

SET OF SENTENCES OF A LANGUAGE WITHOUT REFERENCE TO INFORMATION ABOUT SETTINGS AND WITHOUT SIGNIFICANT VARIATION FROM SPEAKER TO SPEAKER, THEN THAT ABILITY IS PROPERLY THE SUBJECT MATTER OF A SYNCHRONIC THEORY IN LINGUISTICS.

The first question in determining the subject matter of a semantic theory is: Can we find an ability which satisfies the antecedent of this generalization, which is beyond the range of grammatical description, and which is semantic in some reasonable sense? If we can, then that ability falls within the domain of a semantic theory.

In order to find such an ability, let us consider a communication situation so constructed that no information about setting can contribute to a speaker's understanding of a sentence encountered in that situation. Any extragrammatical ability that a speaker can employ to understand the meaning of a sentence in such a situation will ipso facto be considered to require semantic explanation.

The type of communication situation we shall consider is the following. A number of English-speakers receive an anonymous letter containing only the English sentence S. We are interested in the difference between this type of situation and one in which the same anonymous letter is received by persons who do not speak English but are equipped with a completely adequate grammar of English. To investigate what the first group can do by way of comprehending the meaning of S that the second group cannot is to factor out the contribution of grammar to the understanding of sentences. We will only investigate aspects of linguistic ability which are invariant from individual to individual within each group. We thus make sure that the abilities under investigation are a function not of idiosyncrasies of a speaker's personal history but only of his knowledge of his language.

Suppose S is the sentence *The bill is large*. Speakers of English will agree that this sentence is ambiguous, i.e. that it has at least two readings. According to one it means that some document demanding a sum of money to discharge a debt exceeds in size most such documents; according to the other it means that the beak of a certain bird exceeds in bulk those of most similar birds. However, the fact that this sentence is ambiguous between these readings cannot be attributed to its syntactic structure, since, syntactically, its structure on both readings is as shown in Figure 1. That is, the group who do not speak English but are equipped with a grammar can say no more about *The bill is large* than what is represented in Fig. 1. Thus, this sentence, which is marked as unambiguous by

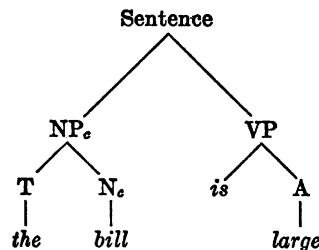


FIG. 1

the grammar, will be understood as ambiguous by a fluent speaker. From this difference between the performances of the two groups, it follows that one facet of the speaker's ability that a semantic theory will have to reconstruct is that he can detect nonsyntactic ambiguities and characterize the content of each reading of a sentence.

Now suppose S is the sentence *The bill is large but need not be paid*. Speakers of English will understand this sentence only on readings in which *bill* means an order to pay a sum of money to discharge a debt. This shows that a speaker can disambiguate parts of a sentence in terms of other parts and thereby determine the number of readings of a sentence. Thus, another facet of the speaker's semantic ability is that of determining the number of readings that a sentence has by exploiting semantic relations in the sentence to eliminate potential ambiguities.

Now let S be the sentence *The paint is silent*. English speakers will at once recognize that this sentence is anomalous in some way. For example, they will distinguish it from such sentences as *The paint is wet* and *The paint is yellow* by applying to it such epithets as 'odd', 'peculiar', 'paradoxical', and 'bizarre'. Though it is clear that the speaker does not have the explicit conceptual machinery to correctly characterize the difference between these sentences, his consistent use of such rough labels shows that he is aware of some sort of linguistic anomaly. But the group who do not speak English and are equipped only with a grammar will regard all these sentences as fully regular, since there is no grammatical basis for distinguishing between them. Hence, another facet of the semantic ability of the speaker is that of detecting semantic anomalies. Correspondingly, a semantic theory will be needed to mark the distinction between semantically anomalous sentences and semantically regular sentences, so far as this distinction is not co-extensive with the distinction the grammar makes between ungrammatical and grammatical strings of morphemes.

Finally, whatever sentence the anonymous letter contains, as a rule, speakers of English can easily decide what sentences are paraphrases of it and what are not, in the sense that they can answer the questions *What does the letter say? Does the letter say such-and-such? How can what the letter says be rephrased?* This facet of the speaker's ability cannot be referred to his mastery of grammar either, for a person who is equipped with a grammar but who does not speak English will be unable to tell whether or not a sentence is a paraphrase of S. The reasons are simply that there need be no definite grammatical relation between a sentence and its paraphrases, e.g. between *Two chairs are in the room* and *There are at least two things in the room and each is a chair*, and that where a definite grammatical relation obtains between a pair of sentences, neither need be a paraphrase of the other, e.g. *The ball was hit by the man* and *The ball was hit*, *The man hit the ball* and *The man did not hit the ball*.⁶ Thus, still another facet of the speaker's semantic ability which must fall within the domain of a semantic theory is his paraphrasing skill.

We can now tentatively characterize the lower bound on the domain of a

⁶ Cf. *Syntactic structures*, Appendix II, for the transformations which relate these sentences.

semantic theory, since we have found an ability of speakers which cannot be accounted for by grammar, which is semantic in a reasonable sense, and which enables speakers to apprehend the semantic structure of an infinite number of sentences without information about setting and independent of individual differences between speakers. We thus take the goals of a semantic theory to include at least the explication of each facet of this ability and of the interrelations between them.

The speaker's exercise of this ability, which henceforth we shall refer to as THE ABILITY TO INTERPRET SENTENCES, provides empirical data for the construction of a semantic theory, just as the construction of a grammar draws upon empirical data supplied by the exercise of the speaker's ability to distinguish well-formed sentences from ungrammatical strings, to recognize syntactic ambiguity, and to appreciate relations between sentence types. A semantic theory describes and explains the interpretative ability of speakers by accounting for their performance in determining the number and content of the readings of a sentence, by detecting semantic anomalies, by deciding on paraphrase relations between sentences, and by marking every other semantic property or relation that plays a role in this ability.

4. What is beyond the descriptive scope of a semantic theory. Having fixed a lower bound on the domain of a semantic theory, our next step must be to fix an upper bound, thus uniquely determining the set of problems forming the domain of a semantic theory of a natural language.

Previous conceptions of semantics have usually defined the goals of a semantic description of a natural language in such a way that to achieve them a semantic theory would have to account for the manner in which settings determine how an utterance is understood. We shall now show that to set the goals of a semantic theory this high is to set them too high. Once we have shown that a semantic theory cannot be expected to account for the way settings determine how an utterance is understood, we will have fixed an upper bound on the domain of semantic theories. That is, we will have shown that a semantic theory is a theory of the speaker's ability to interpret the sentences of his language.

The form of a theory of how settings control the understanding of utterances of sentences is as follows. Such a theory is a function F whose arguments are a sentence S ; GS , a grammatical description of S ; IS , a semantic interpretation of S (where IS is the set of possible readings of S); and C , an abstract characterization of a setting. $F(S, GS, IS, C)$ is

- (1) the particular reading in IS that speakers of the language give to S in settings of the type C , or
- (2) An n -tuple ($n \geq 2$) of the readings from IS that speakers of the language give to S if S is ambiguous n -ways in settings of type C , or
- (3) The null element if speakers of the language give to S none of the readings in IS when S occurs in settings of type C .

The value of $F(S, GS, IS, C)$ is (1) just in case C fully disambiguates S , i.e. C determines a unique reading from the one or more in IS ; it is (2) just in case C