SUPPES:

MODEL-THEORETIC SEMANTICS FOR NATURAL LANGUAGE*  

I would like to present in summary form my ideas about the use of phrase-structure grammars, in particular context-free grammars, together with an appropriate model-theoretic semantics to provide a semantical apparatus for the generation or analysis of utterances of ordinary language. The machinery described in this paper is certainly not adequate to all the problems that confront us in having a complete theory of natural language, but it does provide, I think, a good basis from which to make additional extensions and modifications, and it fits in historically very nicely with the developments of model-theoretic semantics that began with Frege and received their major impetus from the work of Tarski and his students, and, on the other hand, the much more recent work on generative grammars by linguists. The first part of this paper is drawn from a previous publication (Suppes, 1970a) and the second part on model-theoretic semantics from Suppes (1971).+

A. PROBABILISTIC GRAMMARS

1. Introduction

Although a fully adequate grammar for a substantial portion of any natural language does not exist, a vigorous and controversial discussion of how to choose among several competing grammars has already developed. On occasion, criteria of simplicity have been suggested as systematic scientific criteria for selection. The absence of such systematic criteria of simplicity in other domains of science inevitably raises doubts about the feasibility of such criteria for the selection of a grammar. Although some informal and intuitive discussion of simplicity is often included in the selection of theories or models in physics or in other branches of science, there is no serious systematic literature on problems of measuring simplicity, nor is there any systematic literature in which criteria of simplicity are used in a substantive fashion to select from among several theories. There are many reasons for this, but perhaps the most pressing one is that the use of more obviously objective criteria leaves little room for the addition of further criteria of simplicity. The central thesis of this paper is that objective probabilistic criteria of a standard scientific sort may
be used to select a grammar.
Certainly the general idea of looking at the distribution of linguistic types in a given corpus is not new. Everyone is familiar with the remarkable agreement of Zipf's law with the distribution of word frequencies in almost any substantial sample of a natural language. The empirical agreement of these distributions with Zipf's law is not in dispute, although a large and controversial literature is concerned with the most appropriate assumptions of a qualitative and elementary kind from which to derive the law. While there is, I believe, general agreement about the approximate empirical adequacy of Zipf's law, no one claims that a probabilistic account of the frequency distribution of words in a corpus is anything like an ultimate account of how the words are used or why they are used when they are. In the same sense, in the discussion here of probabilistic grammars, I do not claim that the frequency distribution of grammatical types provides an ultimate account of how the language is used or for what purpose a given utterance is made. Yet, it does seem correct to claim that the generation of the relative frequencies of utterances is a proper requirement to place on a generative grammar for a corpus.

Because of the importance of this last point, let me expand it. It might be claimed that the relative frequencies of grammatical utterances are no more pertinent to grammar than the relative frequency of shapes to geometry. No doubt, in one sense such a claim is correct. If we are concerned, on one hand, simply with the mathematical relation between formal languages and the types of automata that can generate these languages, then there is a full set of mathematical questions for which relative frequencies are not appropriate. In the same way, in standard axiomatizations of geometry, we are concerned only with the representations of the geometry and its invariants, not with questions of actual frequency of distribution of figures in nature. In fact, we all recognize that such questions are foreign to the spirit of either classical or modern geometry. On the other hand, when we deal with the physics of objects in nature there are many aspects of shapes and their frequencies of fundamental importance, ranging from the discussion of the shape of clouds and the reason for their shape to the spatial configuration of large and complex organic molecules like proteins.

From the standpoint of empirical application, one of the more dis-
satisfying aspects of the purely formal theory of grammars is that no distinction is made between utterances of ordinary length and utterances that are arbitrarily long, for example, of more than 10^50 words. One of the most obvious and fundamental features of actual spoken speech or written text is the distribution of length of utterance, and the relatively sharp bounds on the complexity of utterances, because of the highly restricted use of embedding or other recursive devices. Not to take account of these facts of utterance length and the limitations on complexity is to ignore two major aspects of actual speech and writing. As we shall see, one of the virtues of a probabilistic grammar is to deal directly with these central features of language.

Still another way of putting the matter is this. In any application of concepts to a complex empirical domain, there is always a degree of uncertainty as to the level of abstraction we should reach for. In mechanics, for example, we do not take account of the color of objects, and it is not taken as a responsibility of mechanics to predict the color of objects. (I refer here to classical mechanics - it could be taken as a responsibility of quantum mechanics.) But ignoring major features of empirical phenomena is in all cases surely a defect and not a virtue. We ignore major features because it is difficult to account for them, not because they are uninteresting or improper subjects for investigation. In the case of grammars, the features of utterance length and utterance complexity seem central; the distribution of these features is of primary importance in understanding the character of actual language use.

A different kind of objection to considering probabilistic grammars at the present stage of inquiry might be the following. It is agreed on all sides that an adequate grammar, in the sense of simply accounting for the grammatical structure of sentences, does not exist for any substantial portion of any natural language. In view of the absence of even one grammar in terms of this criterion, what is the point of imposing a stricter criterion to also account for the relative frequency of utterances? It might be asserted that until at least one adequate grammar exists, there is no need to be concerned with a probabilistic criterion of choice. My answer to such a claim is this. The probabilistic program described in this paper is meant to be supplementary rather than competitive with traditional investigations of grammatical structure. The large and subtle linguistic literature on important features of natural
language syntax constitutes an important and permanent body of material. To draw an analogy from meteorology, a probabilistic measure of a grammar's adequacy stands to ordinary linguistic analysis of particular features, such as verb nominalization or negative constructions, in the same relation that dynamical meteorology stands to classical observation of the clouds. While dynamical meteorology can predict the macroscopic movement of fronts, it cannot predict the exact shape of fair-weather cumulus or storm-generated cumulonimbus. Put differently, one objective of a probabilistic grammar is to account for a high percentage of a corpus with a relatively simple grammar and to isolate the deviant cases that need additional analysis and explanation. At the present time, the main tendency in linguistics is to look at the deviant cases and to ignore trying to give a quantitative account of that part of a corpus that can be analyzed in relatively simple terms. Another feature of probabilistic grammars worth noting is that such a grammar can permit the generation of grammatical types that do not occur in a given corpus. It is possible to take a tolerant attitude toward utterances that are on the borderline of grammatical acceptability, as long as the relative frequency of such utterances is low. The point is that the objective of the probabilistic model is not just to give an account of the finite corpus of spoken speech or written text used as a basis for estimating the parameters of the model, but to use the finite corpus as a sample to infer parameter values for a larger, potentially infinite 'population' in the standard probabilistic fashion. On occasion, there seems to have been some confusion on this point. It has been seriously suggested more than once that for a finite corpus one could write a grammar by simply having a separate rewrite rule for each terminal sentence. Once a probabilistic grammar is sought, such a proposal is easily ruled out as acceptable. One method of so doing is to apply a standard probabilistic test as to whether genuine probabilities have been observed in a sample. We run a split-half analysis, and it is required that within sampling variation the same estimates be obtained from two randomly selected halves of the corpus. Another point of confusion among some linguists and philosophers with whom I have discussed the methodology of fitting probabilistic grammars to data is this. It is felt that some sort of legerdemain is involved in estimating the parameters of a probabilistic grammar from the data which it is supposed to predict. At a casual glance
it may seem that the predictions should always be good and not too interesting because the parameters are estimated from the very data they are used to predict, but this is to misunderstand the many different ways the game of prediction may be played. Certainly, if the number of parameters equals the number of predictions the results are not very interesting. On the other hand, the more the number of predictions exceeds the number of parameters the greater the interest in the predictions of the theory. To convince one linguist of the wide applicability of techniques of estimating parameters from data they predict and also persuade him that such estimation is not an intellectually dishonest form of science, I pointed out that in studying the motion of the simple mechanical system consisting of the Earth, Moon and Sun, at least nine position parameters and nine velocity or momentum parameters as well as mass parameters must be estimated from the data (the actual situation is much more complicated), and everyone agrees that this is 'honest' science.

It is hardly possible in this paper to enter into a full-scale analysis and defense of the role of probabilistic and statistical methodology in science. What I have said briefly here can easily be expanded; I have tried to deal with some of the issues in a monograph on causality (Suppes, 1970b). My own conviction is that at present the quantitative study of language must almost always be probabilistic in nature. The data simply cannot be handled quantitatively by a deterministic theory. A third confusion of some linguists needs to be mentioned in this connection. The use of a probabilistic grammar in no ways entails a commitment to finite Markovian dependencies in the temporal sequence of spoken speech. Two aspects of such grammars make this clear. First, in general such grammars generate a stochastic process that is a chain of infinite order in the terminal vocabulary, not a finite Markov process. Second, the probabilistic parameters are attached directly to the generation of non-terminal strings of syntactic categories. Both of these observations are easy to check in the more technical details of later sections.

The purpose of this part is to define the framework within which empirical investigations of probabilistic grammars can take place and to sketch how this attack can be made. The full presentation of empirical results will be left to other papers. In the detailed empirical work I have depended on the collaboration of younger
colleagues, especially Elizabeth Gammon and Arlene Moskowitz. I draw on our joint work for examples in subsequent sections of this paper. In the next section I give a simple example, indeed, a simple-minded example, of a probabilistic grammar, to illustrate the methodology without complications. In the third section I indicate how such ideas may be applied to the spoken speech of a young child. In the fourth section I consider briefly the representation problem for probabilistic languages. I emphasize that the results of an empirical sort in this paper are all preliminary in nature. The detailed development of the empirical applications is a complicated and involved affair and goes beyond the scope of the work presented here.

2. A Simple Example

A simple example that illustrates the methodology of constructing and testing probabilistic grammars is described in detail in this section. It is not meant to be complex enough to fit any actual corpus.

The example is a context-free grammar that can easily be rewritten as a regular grammar. The five syntactic or semantic categories are just \( V_1 \), \( V_2 \), Adj, PN and N, where \( V_1 \) is the class of intransitive verbs, \( V_2 \) the class of transitive verbs or two-place predicates, Adj the class of adjectives, PN the class of proper nouns and N the class of common nouns. Additional non-terminal vocabulary consists of the symbols S, NP, VP and AdjP. The set of production rules consists of the following seven rules, plus the rewrite rules for terminal vocabulary that belong to one of the five categories. The probability of using one of the rules is shown on the right. Thus, since Rule 1 is obligatory, the probability of using it is 1. In the generation of any sentence, either Rule 2 or Rule 3 must be used. Thus the probabilities \( \alpha \) and \( 1 - \alpha \), which sum to 1, and so forth for the other rules.

<table>
<thead>
<tr>
<th>Production Rule</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ( S \rightarrow NP + VP )</td>
<td>1</td>
</tr>
<tr>
<td>2. ( VP \rightarrow V_1 )</td>
<td>( 1 - \alpha )</td>
</tr>
<tr>
<td>3. ( VP \rightarrow V_2 + NP )</td>
<td>( \alpha )</td>
</tr>
<tr>
<td>4. ( NP \rightarrow PN )</td>
<td>( 1 - \beta )</td>
</tr>
<tr>
<td>5. ( NP \rightarrow AdjP + N )</td>
<td>( \beta )</td>
</tr>
</tbody>
</table>
This probabilistic grammar has three parameters, $\alpha$, $\beta$, and $\gamma$, and the probability of each grammatical type of sentence can be expressed as a monomial function of the parameters. In particular, if $\text{Adj}^n$ is understood to denote a string of $n$ adjectives, then the possible grammatical types (infinite in number) all fall under one of the corresponding schemes, with the indicated probability.

<table>
<thead>
<tr>
<th>Grammatical Type</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. $\text{PN} + V_1$</td>
<td>$(1 - \alpha)(1 - \beta)$</td>
</tr>
<tr>
<td>2. $\text{PN} + V_2 + \text{PN}$</td>
<td>$\alpha(1 - \beta)^2$</td>
</tr>
<tr>
<td>3. $\text{Adj}^n + N + V_1$</td>
<td>$(1 - \alpha)\beta(1 - \gamma)^{n-1}\gamma$</td>
</tr>
<tr>
<td>4. $\text{PN} + V_2 + \text{Adj}^n + N$</td>
<td>$\alpha\beta(1 - \beta)(1 - \gamma)^{n-1}\gamma$</td>
</tr>
<tr>
<td>5. $\text{Adj}^n + N + V_2 + \text{PN}$</td>
<td>$\alpha\beta(1 - \beta)(1 - \gamma)^{n-1}\gamma$</td>
</tr>
<tr>
<td>6. $\text{Adj}^m + N + V_2 + \text{Adj}^n + N$</td>
<td>$\alpha\beta^2(1 - \gamma)^{m+n-2}\gamma^2$</td>
</tr>
</tbody>
</table>

On the hypothesis that this grammar is adequate for the corpus we are studying, each utterance will exemplify one of the grammatical types falling under the six schemes. The empirical relative frequency of each type in the corpus can be used to find a maximum-likelihood estimate of each of the three parameters. Let $x_1, \ldots, x_n$ be the finite sequence of actual utterances. The likelihood function $L(x_1, \ldots, x_n; \alpha, \beta, \gamma)$ is the function that has as its value the probability of obtaining or generating sequence $x_1, \ldots, x_n$ of utterances given parameters $\alpha$, $\beta$, $\gamma$. The computation of $L$ assumes the correctness of the probabilistic grammar, and this implies among other things the statistical independence of the grammatical type of utterances, an assumption that is violated in any actual corpus, but probably not too excessively. The maximum-likelihood estimates of $\alpha$, $\beta$, and $\gamma$ are just those values $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ that maximize the probability of the observed or generated sequence $x_1, \ldots, x_n$. Let $y_1$ be the number of occurrences of grammatical type 1, i.e., $\text{PN} + V_1$, as given in the above table, let $y_2$ be the number of occurrences of type 2, i.e., $\text{PN} + V_2 + \text{PN}$, let $y_3, n$ be the number of occurrences of type 3 with a string of $n$ adjectives, and let
similar definitions apply for \( y_{4,n} \), \( y_{5,n} \) and \( y_{6,m,n} \). Then on the assumption of statistical independence, the likelihood function can be expressed as:

\[
L(x_1, \ldots, x_n; \alpha, \beta, \gamma) = \left[ (1 - \alpha)(1 - \beta) \right]^N \left[ \alpha(1 - \beta)^2 \right]^P \prod_{n=1}^{N} \prod_{m=1}^{P} \left[ \alpha \beta^2 (1 - \gamma)^{m+n-2} \gamma^2 \right]^{y_{6,m,n}}.
\]

Of course, in any finite corpus the infinite products will always have only a finite number of terms not equal to one. To find \( \alpha \), \( \beta \) and \( \gamma \) as functions of the observed frequencies \( y_1, \ldots, y_6, m, n \), the standard approach is to take the logarithm of both sides of (1), in order to convert products into sums, and then to take partial derivatives with respect to \( \alpha \), \( \beta \) and \( \gamma \) to find the values that maximize \( L \). The maximum is not changed by taking the log of \( L \), because log is a strictly monotonic increasing function. Letting

\[
L = \log L, \quad y_3 = \sum y_{3,n}, \quad y_4 = \sum y_{4,n}, \quad y_5 = \sum y_{5,n}, \quad \text{and} \quad y_6 = \sum y_{6,m,n},
\]

we have

\[
\frac{\Delta L}{\Delta \alpha} = \frac{y_1 + y_3 + y_2 + y_4 + y_5 + y_6}{1 - \alpha} = 0;
\]

\[
\frac{\Delta L}{\Delta \beta} = \frac{y_1}{1 - \beta} - \frac{2y_2}{1 - \beta} + \frac{y_3}{\beta} + \frac{y_4}{\beta} - \frac{y_5}{1 - \beta} + \frac{2y_6}{\beta} = 0;
\]

\[
\frac{\Delta L}{\Delta \gamma} = \frac{y_3 + y_4 + y_5 + y_6}{\gamma} - \frac{y_{3,2} + y_{4,2} + y_{5,2}}{1 - \gamma} + \frac{(y_{3,n} + y_{4,n} + y_{5,n})}{1 - \gamma} + \ldots - \frac{y_{b,1,1}}{1 - \gamma} + \frac{(m - n - 2)y_{b,m,n}}{1 - \gamma} + \ldots = 0.
\]
If we let
\[
Z_{6,n} = \sum_{m' + n' = n+1} \sum_{m', n'} y_{b, m', n'}
\]
then after solving the above three equations we have as maximum-likelihood estimates:

\[
\hat{a} = \frac{y_2 + y_4 + y_5 + y_b}{y_1 + y_2 + y_3 + y_4 + y_5 + y_b}
\]

\[
\hat{b} = \frac{y_3 + y_4 + y_5 + 2y_b}{y_1 + 2y_2 + y_3 + 2y_4 + 2y_5 + 2y_b}
\]

\[
\hat{\gamma} = \frac{y_3 + y_4 + y_5 + z_6}{\sum n(y_{3,n} + y_{4,n} + y_{5,n} + z_{6,n})}
\]

As would be expected from the role of \( \gamma \) as a stopping parameter for the auditon of adjectives, the maximum-likelihood estimate of \( \gamma \) is just the standard one for the mean of a geometrical distribution. Having estimated \( a \), \( b \) and \( \gamma \) from utterance frequency data, we can then test the goodness of fit of the probabilistic grammar in some standard statistical fashion, using a chi-square or some comparable statistical test. Some numerical results of such tests are reported later in the paper. The criterion for acceptance of the grammar is then just a standard statistical one. To say this is not to imply that standard statistical methods or criteria of testing are without their own conceptual problems. Rather the intention is to emphasize that the selection of a grammar can follow a standard scientific methodology of great power and wide applicability, and methodological arguments meant to be special to linguistics - like the discussion of simplicity - can be dispensed with.

3. Grammar For Adam I

Because of the relative syntactic simplicity and brevity of the spoken utterances of very young children, it is natural to begin attempts to write probabilistic grammars by examining such speech.
This section presents some preliminary results for Adam I, a well-known corpus collected by Roger Brown and his associates at Harvard. Adam was a young boy of about 26 months at the time the speech was recorded. The corpus analyzed by Arlene Moskowitz and me consists of eight hours of recordings extending over a period of some weeks. Our work has been based on the written transcript of the tapes made at Harvard. Accepting for the most part the word and utterance boundaries established in the Harvard transcript, we found that the corpus consists of 6109 word occurrences with a vocabulary of 673 different words and 3497 utterances.

Even though the mean utterance length of Adam I is somewhat less than 2.0, there are difficulties in writing a completely adequate probabilistic grammar for the full corpus. An example is considered below.

To provide, however, a sample of what can be done on a more restricted basis, and in a framework that is fairly close to the simple artificial example considered in the preceding section, I restrict my attention to the noun phrases of Adam I. Noun phrases dominate Adam I, if for no other reason than because the most common single utterance is the single noun. Of the 3497 utterances, we have classified 956 as single occurrences of nouns. Another 192 are occurrences of two nouns in sequence, 147 adjective followed by noun, and 138 adjectives alone. In a number of other cases, the whole utterance is a simple noun phrase preceded or followed by a one-word rejoinder, vocative or locative.

The following phrase-structure grammar was written for noun phrases of Adam I. The seven production rules are given below with the corresponding probabilities shown on the right. This particular probabilistic model has five free parameters; the sum of the $a_i$'s is one, so the $a_i$'s contribute four parameters to be fitted to the data, and in the case of the $b_i$'s there is just one free parameter.

<table>
<thead>
<tr>
<th>Production Rule</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. NP $\rightarrow$ N</td>
<td>$a_1$</td>
</tr>
<tr>
<td>2. NP $\rightarrow$ AdjP</td>
<td>$a_2$</td>
</tr>
<tr>
<td>3. NP $\rightarrow$ AdjP + N</td>
<td>$a_3$</td>
</tr>
<tr>
<td>4. NP $\rightarrow$ Pro</td>
<td>$a_4$</td>
</tr>
<tr>
<td>5. NP $\rightarrow$ NP + NP</td>
<td>$a_5$</td>
</tr>
</tbody>
</table>
b. AdjP + AdjP + Adj

7. AdjP + Adj

What is pleasing about these rules, and perhaps surprising, is that six of them are completely standard. (The one new symbol introduced here is Pro for pronoun; inflection of pronouns has been ignored in the present grammar.) The only slightly non-standard rule is Rule 5. The main application of this rule is in the production of the noun phrases consisting of a noun followed by a noun, with the first noun being an uninflected possessive modifying the second noun. Examples from the corpus are Adam horn, Adam hat, Daddy racket and Doctor Dan circus.

To give a better approximation to statistical independence in the occurrences of utterances, I deleted successive occurrences of the same noun phrase in the frequency count, and only first occurrences in a run of occurrences were considered in analyzing the data. The maximum-likelihood estimates of the parameters were obtained from the resulting 2434 occurrences of noun phrases in the corpus.

Estimated Parameter Values

\[ a_1 = 0.6391 \]
\[ a_2 = 0.0529 \]
\[ a_3 = 0.0497 \]
\[ a_4 = 0.1439 \]
\[ a_5 = 0.1144 \]
\[ b_1 = 0.0581 \]
\[ b_2 = 0.9419 \]

On the basis of remarks already made, the high value of \( a_1 \) is not surprising because of the high frequency of occurrences of single nouns in the corpus. It should be noted that the value of \( a_1 \) is even higher than the relative frequency of single occurrences of nouns, because the noun-phrase grammar has been written to fit all noun phrases, including those occurring in full sentence context or in conjunction with verbs, etc. Thus in a count of single nouns as noun phrases every occurrence of a single noun as a noun phrase was counted, and as can be seen from Table I, there are 1445 such single nouns without immediate repetition. The high value of \( b_2 \) indicates that there are very few occurrences of successive adjectives, and therefore in almost all cases the adjective phrase was rewritten simply as an adjective (Rule 7).

Comparison of the theoretical frequencies of the probabilistic
grammar with the observed frequencies is given in Table I.

TABLE I
Probabilistic NounPhrase Grammar for Adam I

<table>
<thead>
<tr>
<th>Noun phrase</th>
<th>Observed frequency</th>
<th>Theoretical frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1445</td>
<td>1555.6</td>
</tr>
<tr>
<td>P</td>
<td>388</td>
<td>350.1</td>
</tr>
<tr>
<td>NN</td>
<td>231</td>
<td>113.7</td>
</tr>
<tr>
<td>AN</td>
<td>135</td>
<td>114.0</td>
</tr>
<tr>
<td>A</td>
<td>114</td>
<td>121.3</td>
</tr>
<tr>
<td>PN</td>
<td>31</td>
<td>25.6</td>
</tr>
<tr>
<td>NA</td>
<td>19</td>
<td>8.9</td>
</tr>
<tr>
<td>NNN</td>
<td>12</td>
<td>8.3</td>
</tr>
<tr>
<td>AA</td>
<td>10</td>
<td>7.1</td>
</tr>
<tr>
<td>NAN</td>
<td>8</td>
<td>8.3</td>
</tr>
<tr>
<td>AP</td>
<td>6</td>
<td>2.0</td>
</tr>
<tr>
<td>PPN</td>
<td>0</td>
<td>.4</td>
</tr>
<tr>
<td>ANN</td>
<td>5</td>
<td>8.3</td>
</tr>
<tr>
<td>AAN</td>
<td>4</td>
<td>6.0</td>
</tr>
<tr>
<td>PA</td>
<td>4</td>
<td>2.0</td>
</tr>
<tr>
<td>ANA</td>
<td>3</td>
<td>.7</td>
</tr>
<tr>
<td>APN</td>
<td>3</td>
<td>.1</td>
</tr>
<tr>
<td>AAA</td>
<td>2</td>
<td>.4</td>
</tr>
<tr>
<td>APA</td>
<td>2</td>
<td>.0</td>
</tr>
<tr>
<td>NPP</td>
<td>2</td>
<td>.4</td>
</tr>
<tr>
<td>PAA</td>
<td>2</td>
<td>.1</td>
</tr>
<tr>
<td>PAN</td>
<td>2</td>
<td>1.9</td>
</tr>
</tbody>
</table>

Some fairly transparent abbreviations are used in the table to reduce its size; as before, N stands for noun, A for adjective, and P for pronoun. From the standpoint of a statistical goodness-of-fit test, the chi-square is still enormous; its value is 309.4 and there are only seven net degrees of freedom. Thus by ordinary statistical standards we must reject the fit of the model, but at this stage of the investigation the qualitative comparison of the observed and theoretical frequencies is encouraging. The rank order of the theoretical frequencies for the more frequent types of noun phrases closely matches that of the observed frequencies. The only really
serious discrepancy is in the case of the phrases consisting of two nouns, for which the theoretical frequency is substantially less than the observed frequency. It is possible that a different way of generating the possessives that dominate the occurrences of these two nouns in sequence would improve the prediction. Summation of the observed and theoretical frequencies will show a discrepancy between the two columns. I explicitly note this. It is expected, because the column of theoretical frequencies should also include the classes that were not observed actually occurring in the corpus. The prediction of the sum of these unobserved classes is that they should have a frequency of 100.0, which is slightly less than 5% of the total observed frequency of 2434.

Note that the derivation of the probabilities for each grammatical type of noun phrase used the simplest derivation. For example, in the case of Adj + N, the theoretical probability was computed from successive application of Rule 3, followed by Rule 6, followed by Rule 7. It is also apparent that a quite different derivation of this noun phrase can be obtained by using Rule 5. Because of the rather special character of Rule 5, all derivations avoided Rule 5 when possible and only the simplest derivation was used in computing the probabilities. In other words, no account was taken of the ambiguity of the noun phrases. A more exact and sensitive analysis would require a more thorough investigation of this point. It is probable that there would be no substantial improvement in theoretical predictions in the present case, if these matters were taken account of. The reader may also have observed that the theoretical frequencies reflect certain symmetries in the predictions that do not exist in the observed frequencies. For example, the type Pro + Pro + N has an observed frequency of six, and the permutation N + Pro + Pro has an observed frequency of two. This discrepancy could easily be attributed to sampling. The symmetries imposed by the theoretical grammar generated from Rules 1 to 7 are considerable, but they do not introduce symmetries in any strongly disturbing way. Again let me emphasize that the symmetries that are somewhat questionable are almost entirely introduced by means of Rule 5. Finally, note that I have omitted from the list of noun phrases the occurrence of two pronouns in sequence because all cases consisted of the question Who that? or What that?, and it seemed inappropriate to classify these occurrences as single noun phrases. I hasten to add that remarks of a similar sort can be made about
some of the other classifications.

It is important for the reader to keep in mind the various qualifications that have been made here. I have no intention of conveying the impression that a definitive result has been obtained. I present the results of Table I as a preliminary indication of what can be achieved by the methods introduced in this paper. Appropriate qualifications and refinements will undoubtedly lead to better and more substantial findings.

4. Representation Problem For Probabilistic Languages

From what has already been said it should be clear enough that the imposition of a probabilistic generative structure is an additional constraint on a grammar. It is natural to ask if a probabilistic grammar can always be found for a language known merely to have a grammar. Put in this intuitive fashion, it is not clear exactly what question is being asked.

As a preliminary to a precise formulation of the question, an explicit formal characterization of probabilistic grammars is needed. In a fashion familiar from the literature we may define a grammar as a quadruple $(V_N, V_T, R, S)$, where $V_N$, $V_T$ and $R$ are finite sets, $S$ is a member of $V_N$, $V_N$ and $V_T$ are disjoint, and $R$ is a set of ordered pairs, whose first members are in $V^+$, and whose second members are in $V^*$, where $V = V_N \cup V_T$, $V^*$ is the set of all finite sequences whose terms are elements of $V$, and $V^+$ is $V^*$ minus the empty sequence. As usual, it is intended that $V_N$ be the non-terminal and $V_T$ the terminal vocabulary, $R$ the set of productions and $S$ the start symbol. The language $L$ generated by $G$ is defined in the standard manner and will be omitted here.

In the sense of the earlier sections of this paper, a probabilistic grammar is obtained by adding a conditional distribution on the set $R$ of productions. Formally we have:

**DEFINITION:** A quintuple $G = (V_N, V_T, R, S, p)$ is a probabilistic grammar if and only if $G = (V_N, V_T, R, S)$ is a grammar, and $p$ is a real-valued function defined on $R$ such that

1. For each $(\sigma_i, \sigma_j)$ in $R$, $p(\sigma_i, \sigma_j) \geq 0$,
2. For each $\sigma_i$ in the domain of $R$
   $$\sum_{\sigma_j} p(\sigma_i; \sigma_j) = 1,$$

where the summation is over the range of $R$. 
Various generalizations of this definition are easily given; for example, it is natural in some contexts to replace the fixed start symbol $S$ by a probability distribution over $V_N$. But such generalizations will not really affect the essential character of the representation problem as formulated here. For explicitness, we also need the concept of a probabilistic language, which is just a pair $(L, p)$, where $L$ is a language and $p$ is a probability density defined on $L$, i.e., for each $x$ in $L$, $p(x) > 0$ and 

$$\sum_{x \in L} p(x) = 1.$$ 

The first formulation of the representation problem is then this. Let $L$ be a language of type $i$ $(i = 0, 1, 2, 3)$, with probability density $p$. Does there always exist a probabilistic grammar $G$ (of type $i$) that generates $(L, p)$?

What is meant by generation is apparent. If $x \in L$, $p(x)$ must be the sum of the probabilities of all the derivations of $x$ in $G$. Ellis (1969) answered this formulation of the representation problem in the negative for type 2 and type 3 grammars. His example is easy to describe. Let $V_T = \{a\}$, and let $L = \{a^n | n \geq 1\}$. Let $p(a^{n+1}) = 1/\sqrt{n}$, $n > 0$, where $t_1 = 4$, and $t_i = \text{smallest prime such that } t_i^i > \text{max}(t_{i-1}, 2^{2i})$ for $i \geq 1$. In addition, set

$$p(a) = 1 - \sum_{n=1}^\infty p(a^{n+1}).$$

The argument depends upon showing that the probabilities assigned to the strings of $L$ by the above characterization cannot all lie in the extensions of the field of rational numbers generated by the finite set of conditional probabilities attached to the finite set of production rules of any context-free grammar.

From the empirically-oriented standpoint of this paper, Ellis' example, while perfectly correct mathematically, is conceptually unsatisfactory, because any finite sample of $L$ drawn according to the density $p$ could be described also by a density taking only rational values. Put another way, algebraic examples of Ellis' sort do not settle the representation problem when it is given a clearly statistical formulation. Here is one such formulation. (As a matter of notation, if $p$ is a density on $L$, $p_s$ is the sample density of a finite random sample drawn from $(L, p)$.)
Let $L$ be a language of type $i$ with probability density $p$. Does there always exist a probabilistic grammar $G$ (of type $i$) that generates a density $p'$ on $L$ such that for every sample $s$ of $L$ of size less than $N$ and with density $p_s$ the null hypothesis that $s$ is drawn from $(L, p')$ would not be rejected?

I have deliberately imposed a limit $N$ on the size of the sample in order directly to block asymptotic arguments that yield negative results. In referring to the null hypothesis' not being rejected I have in mind using some standard test such as Kolmogorov's and some standard level of significance. The details on this point do not matter here, although a precise solution must be explicit on these matters and also on problems of repeated sampling, fixing the power of the test, etc. My own conjecture is that the statistical formulation of the problem has an affirmative solution for every $N$, but the positive solutions will often not be conceptually interesting.

A final remark about the density $p$ on $L$ is perhaps needed. Some may be concerned about the single occurrence of many individual utterances even in a large corpus. The entire discussion of the representation problem is easily shifted to the category descriptions of terminal strings as exemplified in earlier sections of this paper, and at this level, certainly many grammatical types occur repeatedly.\textsuperscript{2}
1. Introduction

The search for a rigorous and explicit semantics of any significant portion of a natural language is now intensive and far-flung—in the sense that wide varieties of approaches are being taken. Yet almost everyone agrees that at the present time the semantics of natural languages are less satisfactorily formulated than the grammars, even though a complete grammar for any significant fragment of natural language is yet to be written.

A line of thought especially popular in the last couple of years is that the semantics of a natural language can be reduced to the semantics of first-order logic. One way of fitting this scheme into the general approach of generative grammars is to think of the deep structure as being essentially identical with the structure of first-order logic. The central difficulty with this approach is that now as before how the semantics of the surface grammar is to be formulated is still unclear. In other words, how can explicit formal relations be established between first-order logic and the structure of natural languages? Without the outlines of a formal theory, this line of approach has moved no further than the classical stance of introductory teaching in logic, which for many years has concentrated on the translation of English sentences into first-order logical notation. The method of translation, of course, is left at an intuitive and ill-defined level.

The strength of the first-order logic approach is that it represents essentially the only semantical theory with any systematic or deep development, namely, model-theoretic semantics as developed in mathematical logic since the early 1950's, especially since the appearance of Tarski (1935). The semantical approaches developed by linguists or others whose viewpoint is that of generative grammar have been lacking in the formal precision and depth of model-theoretic semantics. Indeed, some of the most important and significant results in the foundations of mathematics belong to the general theory of models. I shall not attempt to review the approaches to semantics that start from a generative-grammar viewpoint, but I have in mind the work of Fodor, Katz, Lakoff, Mc Cawley, and others.
My objective is to combine the viewpoint of model-theoretic semantics and generative grammar, to define semantics for context-free languages and to apply the results to some fragments of natural language. The ideas contained in this paper were developed while I was working with Hélène Bestougeff on the semantical theory of question-answering systems. Later I came across some earlier similar work by Knuth (1968). My developments are rather different from those of Knuth, especially because my objective is to provide tools for the analysis of fragments of natural languages, whereas Knuth was concerned with programming languages.

Although on the surface the viewpoint seems different, I also benefited from a study of Montague's interesting and important work (1970) on the analysis of English as a formal language. My purely extensional line of attack is simpler than Montague's. I adopted it for reasons of expediency, not correctness. I wanted an apparatus that could be applied in a fairly direct way to empirical analysis of a corpus. As in part A. on probabilistic grammars, I began with the speech of a young child, but without doubt, many of the semantical problems that are the center of Montague's concern must be dealt with in analyzing slightly more complex speech. Indeed, some of these problems already arise in the corpus studied here. As in the case of my earlier work on probabilistic grammars, I have found a full-scale analytic attack on a corpus of speech a humbling and bedeviling experience. The results reported here hopefully chart one possible course; in no sense are they more than preliminary.

This part is organized in the following fashion. In Section 2, I describe a simple artificial example to illustrate how a semantic Valuation function is added to the generative mechanisms of a context-free grammar. The relevant formal definitions are given in Section 3. The reader who wants a quick survey of what can be done with the methods, but who is not really interested in formal matters, may skip ahead to Section 4, which contains a part of the detailed empirical results. On the other hand, it will probably be somewhat difficult to comprehend fully the machinery used in the empirical analysis without some perusal of Section 3, unless the reader is already quite familiar with model-theoretic semantics.
2. A Simple Example

To illustrate the semantic methods described formally below, I use as an example the same single language I used in part A. As remarked there, this example is not meant to be complex enough to fit any actual corpus.

Production Rule                      Semantic Function
1. S → NP + VP                      Truth-function
2. VP → V₁                           Identity
3. VP → V₂ + NP                      Image under the converse relation
4. NP → PN                           Identity
5. NP → AdjP + N                     Intersection
6. AdjP → AdjF + Adj                  Intersection
7. AdjP → Adj                           Identity

The grammatical types are the same as in A.2.

What needs explaining are the semantic functions to the right of each production rule. For this purpose it is desirable to look at an example of a sentence generated by this grammar. The intuitive idea is that we define a valuation function v over the terminal vocabulary, and as is standard in model-theoretic semantics, v takes values in some relational structure.

Suppose a speaker wants to say 'John hit Mary'. The valuation function needs to be defined for the three terminal words 'John', 'hit', and 'Mary', we then recursively define the denotation of each labeled node of the derivation tree of the sentence. In this example, I number the nodes so that the denotation function ψ is defined for pairs (n,a), where n is a node of the tree and a is a word in the vocabulary. The tree looks like this.
Let I be the identity function, $\tilde{A}$ the converse of A, i.e.,

$$\tilde{A} = \{<x,y>:<y,x> \in A\},$$

and $f^*A$ the image of A under f, i.e., the range of f restricted to the domain A, and let $T$ be truth and $F$ falsity. Then the denotation of each labeled node of the tree is found by working from the bottom up:

\[
\begin{align*}
\psi(10, \text{Mary}) &= \nu(\text{Mary}) \\
\psi(9, \text{PN}) &= I(\nu(\text{Mary})) \\
\psi(8, \text{hit}) &= \nu(\text{hit}) \\
\psi(7, \text{John}) &= \nu(\text{John}) \\
\psi(6, \text{NP}) &= \Pi(\nu(\text{Mary})) \\
\psi(5, \text{V$_2$}) &= I(\nu(\text{hit})) \\
\psi(4, \text{PN}) &= I(\nu(\text{John})) \\
\psi(3, \text{VP}) &= \Pi(\nu(\text{hit})) = \Pi(\nu(\text{Mary})) \\
\psi(2, \text{NP}) &= \Pi(\nu(\text{John})) \\
\psi(1, S) &= f(\psi(2, \text{NP}), \psi(3, \text{VP})) = \begin{cases} T & \text{if } \psi(2, \text{NP}) \subseteq \psi(3, \text{VP}) \\ F & \text{otherwise} \end{cases}
\end{align*}
\]

Clearly, the functions used above are just the semantic functions associated with the productions. In particular, the production rules for the direct descendants of nodes 2, 4, 5, 6, and 9 all have the identity function as their semantic function.

One point should be emphasized. I do not claim that the set-theoretical semantic functions of actual speech are as simple as those associated with the production rules given in this section. Consider Rule 5, for instance. Intersection is fine for old dictators, but not for alleged dictators. One standard mathematical approach to this kind of difficulty is to generalize the semantic function to cover the meaning of both sorts of cases. In the present case of
adjectives, we could require that the semantic function be one that maps sets of objects into sets of objects. In this vein, Rule 5 would now be represented by

$$\psi(n_1, \text{NP}) = \psi(n_2, \text{AdjP})\psi(n_3, \text{N}).$$

Fortunately, generalizations that rule out the familiar simple functions as semantic functions do not often occur early in children's speech. Some tentative empirical evidence on this point is presented in Section 4.

3. Denoting Grammars

I turn now to formal developments. Some standard grammatical concepts are defined in the interest of completeness. First, if V is a set, $$V^*$$ is the set of all finite sequences whose elements are members of V. I shall often refer to these finite sequences as strings. The empty sequence, $$\epsilon$$, is in $$V^*$$; we define $$V^+ = V^* - \{\epsilon\}$$.

A structure $$G = \langle V, V_N, P, S \rangle$$ is a phrase-structure grammar if and only if V and P are finite, nonempty sets, $$V_N$$ is a subset of V, S is in $$V_N$$ and $$P \subseteq V_N \times V^+$$. Following the usual terminology, $$V_N$$ is the nonterminal vocabulary and $$V_T = V - V_N$$ the terminal vocabulary. S is the start symbol of the single axiom from which we derive strings or words in the language generated by G. The set P is the set of production or rewrite rules. If $$\langle a, \epsilon \rangle \in P$$, we write $$a \rightarrow \epsilon$$, which we read: from a we may produce or derive $$\epsilon$$ (immediately).

A phrase-structure grammar $$G = \langle V, V_N, P, S \rangle$$ is context-free if and only if $$P \subseteq V_N \times V^+$$, i.e., if $$a \rightarrow \gamma$$ is in P then $$\gamma \in V_N$$ and $$\gamma \epsilon V^+$$.

These ideas may be illustrated by considering the simple language of the previous section. Although it is intended that N, PN, Adj, $$V_1$$, and $$V_2$$ be nonterminals in any application, we can treat them as terminals for purposes of illustration, for they do not occur on the left of any of the seven production rules. With this understanding

$$V_N = \{S, \text{NP}, \text{VP}, \text{AdjP}\}$$

$$V_T = \{N, \text{PN}, \text{Adj}, V_1, V_2\}$$

and P is defined by the production rules already given. It is obvious from looking at the production rules that the grammar is context-free, for only elements of $$V_N$$ appear on the left-hand side of any of the seven production rules.
The standard definition of derivations is as follows. Let \( G = \langle V, VN, P, S \rangle \) be a phrase-structure grammar. First, if \( \alpha \to \beta \) is a production of \( P \), and \( \gamma \) and \( \delta \) are strings in \( V^* \), then \( \gamma \alpha \delta \Rightarrow \gamma \beta \delta \). We say that \( \beta \) is derivable from \( \alpha \) in \( G \), in symbols, \( \alpha \Rightarrow^* \beta \) if there are strings \( \alpha_1, \ldots, \alpha_n \) in \( V^* \) such that \( \alpha = \alpha_1 \ldots \alpha_n = \beta \). The sequence \( \Delta = \langle \alpha_1, \ldots, \alpha_n \rangle \) is a derivation in \( G \). The language \( L(G) \) generated by \( G \) is \( \{ \alpha : \alpha \Rightarrow^* V_T \cup S \} \). In other words, \( L(G) \) is the set of all strings made up of terminal vocabulary and derived from \( S \).

The semantic concepts developed also require use of the concept of a derivation tree of a grammar. The relevant notions are set forth in a series of definitions. Certain familiar set-theoretical notions about relations are also needed. To begin with, a binary structure is an ordered pair <\( T, R \)> such that \( T \) is a nonempty set and \( R \) is a binary relation on \( T \), i.e., \( R \subseteq T \times T \). \( R \) is a partial ordering of \( T \) if and only if \( R \) is reflexive, antisymmetric and transitive on \( T \). \( R \) is a strict simple ordering of \( T \) if and only if \( R \) is asymmetric, transitive, and connected on \( T \). We also need the concept of \( R \)-immediate predecessor. For \( x \) and \( y \) in \( T \), \( x \prec y \) if and only if \( xRy \), not \( yRx \) and for every \( z \) if \( z \neq y \) and \( zRy \) then \( zRx \). In the language of formal grammars, we say that if \( x \prec y \), then \( x \) directly dominates \( y \), or \( y \) is the direct descendant of \( x \).

Using these notions, we define in succession tree, ordered tree, and labeled ordered tree. A binary structure <\( T, R \)> is a tree if and only if (i) \( T \) is finite, (ii) \( R \) is a partial ordering of \( T \), (iii) there is an \( R \)-first element of \( T \), i.e., there is an \( x \) such that for every \( y \), \( xRy \), and (iv) if \( x \prec y \) and \( y \prec z \), then \( x \prec z \). If \( xRy \) in a tree, we say that \( y \) is a descendant of \( x \). Also the \( R \)-first element of a tree is called the root of the tree, and an element of \( T \) that has no descendants is called a leaf. We call any element of \( T \) a node, and we shall sometimes refer to leaves as terminal nodes.

A ternary structure <\( T, R, L \)> is an ordered tree if and only if (i) \( L \) is a binary relation on \( T \), (ii) <\( T, R \)> is a tree, (iii) for each \( x \) in \( T \), \( L \) is a strict simple ordering of \{ \( y : x \prec y \) \}, (iv) if \( xLy \) and \( yRz \) then \( xLz \), and (v) if \( xLy \) and \( xHz \) then \( zLy \). It is customary to read \( xLy \) as "\( x \) is to the left of \( y \)." having this ordering is
fundamental to generating terminal strings and not just sets of terminal words. The terminal string of an ordered labeled tree is just the sequence of labels \( \langle f(x_1), \ldots, f(x_n) \rangle \) of the leaves of the tree as ordered by \( L \). Formally, a quinary structure \( \langle T, V, R, L, f \rangle \) is a labeled ordered tree if and only if (i) \( V \) is a nonempty set, (ii) \( \langle T, R, L \rangle \) is an ordered tree, and (iii) \( f \) is a function from \( T \) into \( V \). The function \( f \) is the labeling function and \( f(x) \) is the label of node \( x \).

The definition of a derivation tree is relative to a given context-free grammar.

Definition 1. Let \( G = \langle V, V_N, P, S \rangle \) be a context-free grammar and let \( \tau = \langle T, V, R, L, f \rangle \) be a labeled ordered tree. \( \tau \) is a derivation tree of \( G \) if and only if

1. If \( x \) is the root of \( \tau \), then \( f(x) = S \);
2. If \( xRy \) and \( x \neq y \), then \( f(x) \) is in \( V_N \);
3. If \( y_1, \ldots, y_n \) are all the direct descendants of \( x \), then
   \[ \bigcup_{i=1}^{n} \{ y_i \} \neq \emptyset, \text{ and } y_i \neq y_j \text{ if } i < j, \]
   \[ \langle f(x), f(y_1), \ldots, f(y_n) \rangle \]

is a production in \( P \).

We now turn to semantics proper by introducing the set of set-theoretical functions. We shall let the domains of these functions be n-tuples of any sets (with some appropriate restriction understood to avoid set-theoretical paradoxes).

Definition 2. Let \( \langle V, V_N, P, S \rangle \) be a context-free grammar. Let \( \phi \) be a function defined on \( P \) which assigns to each production \( p \) in \( P \) a finite, possibly empty set of set-theoretical functions subject to the restriction that if the right member of production \( p \) has \( n \) terms of \( V \), then any function of \( \phi(p) \) has \( n \) arguments. Then \( G = \langle V, V_N, P, S, \phi \rangle \) is a potentially denoting context-free grammar. If for each \( p \) in \( P \), \( \phi(p) \) has exactly one member, then \( G \) is said to be simple.

The simplicity and abstractness of the definition may be misleading. In the case of a formal language, e.g., a context-free programming
language, the creators of the language specify the semantics by defining \( \phi \). Matters are more complicated in applying the same idea of capturing the semantics by such a function for fragments of a natural language. Perhaps the most difficult problem is that of giving a straightforward set-theoretical interpretation of intensional contexts, especially to those generated by the expression of propositional attitudes of believing, wanting, seeking and so forth. I shall not attempt to deal with these matters in the present paper.

How the set-theoretical functions in \( \phi(p) \) work was illustrated in the preceding section; some empirical examples follow in the next section. The problems of identifying and verifying \( \phi \) even in the simplest sort of context are discussed there. In one sense the definition should be strengthened to permit only one function in \( \phi(p) \) of a given number of arguments. The intuitive idea behind the restriction is clear. In a given application we try first to assign denotations at the individual word level, and we proceed to two- and three-word phrases only when necessary. The concept of such hierarchical parsing is familiar in computer programming, and a detailed example in the context of a question-answering program is worked out in a joint paper with Hélène Bestougeff. However, as the examples in the next section show, this restriction seems to be too severe for natural languages.

A clear separation of the generality of \( \phi \) and an evaluation function \( v \) is intended. The functions in \( \phi \) should be constant over many different uses of a word, phrase or statement. The valuation \( v \), on the other hand, can change sharply from one occasion of use to the next. To provide for any finite composition of functions, or other ascensions in the natural hierarchy of sets and functions built up from a domain of individuals, the family \( \mathcal{K}'(D) \) of sets with closure properties stronger than needed in any particular application is defined. The abstract objects \( T \) (for truth) and \( F \) (for falsity) are excluded as elements of \( \mathcal{K}'(D) \). In this definition \( \mathcal{P}A \) is the power set of \( A \), i.e., the set of all subjects of \( A \).

Definition 3. Let \( D \) be a nonempty set. Then \( \mathcal{K}'(D) \) is the smallest family of sets such that

(i) \( D \in \mathcal{K}'(D) \),

(11) if \( A, B \in \mathcal{K}'(D) \) then \( A \cup B \in \mathcal{K}'(D) \),
A model structure for $G$ is defined just for terminal words and phrases. The meaning or denotation of nonterminal symbols changes from one derivation or derivation tree to another.

**Definition 4.** Let $D$ be a nonempty set, let $G = \langle V, V_N, P, S \rangle$ be a phrase-structure grammar, and let $v$ be a partial function on $V_T^+$ to $\mathcal{K}(D)$ such that $v$ is defined for $\alpha$ in $V_T^+$ and if $\gamma$ is a subsequence of $\alpha$, then $v$ is not defined for $\gamma$. Then $\mathcal{S} = \langle D, v \rangle$ is a model structure for $G$. If the domain of $v$ is exactly $V_T$, then $\mathcal{S}$ is simple.

We also refer to $v$ as a valuation function for $G$.

I now define semantic trees that assign denotations to nonterminal symbols in a derivation tree. The definition is for simple potential potentially denoting grammars and for simple model structures. In other words, there is a unique semantic function for each production, and the valuation function is defined just on and not on phrases of $V_T^+$.

**Definition 5.** Let $G = \langle V, V_N, P, \phi \rangle$ be a simple, potentially denoting context-free grammar, let $\mathcal{S} = \langle D, v \rangle$ be a simple model structure for $G$, let $\tau' = \langle T, V, R, L, f \rangle$ be a derivation tree of $\langle V, V_N, P, S \rangle$ such that if $x$ is a terminal node, then $f(x) \in V_T$ and let $\psi$ be a function from $\tau$ to $\mathcal{K}(D)$ such that

1. if $\langle x, f(x) \rangle \in f$ and $f(x) \in V_T^+$, then
   
   $\psi(x, f(x)) = v(f(x))$,

2. if $\langle x, f(x) \rangle \in f$, $f(x) \in V_N$, and $y_1, \ldots, y_n$ are all the direct descendants of $x$ with $y_i \not= y_j$ if $i < j$, then
   
   $\psi(x, f(x)) = \phi(\psi(y_1, f(y_1)), \ldots, \psi(y_n, f(y_n))$,

where $\phi = \phi(p)$ and $p$ is the production

$\langle f(x), \langle f(y_1), \ldots, f(y_n) \rangle \rangle$.

Then $\tau = \langle T, V, R, L, f, \psi \rangle$ is a simple semantic tree of $G$ and $\mathcal{S}$.

The extension of Definition 5 to semantic trees that are not simple is relatively straightforward, but is not given explicitly here in
the interest of restricting the formal parts of the paper. The empirical examples considered in the next section implicitly assume this extension, but the simplicity of the corpus makes the several set-theoretical functions \( \psi \) attached to a given production easy to interpret.

The function \( \psi \) assigns a denotation to each node of a semantic tree. The resulting structural analysis can be used to define a concept of meaning or sense for each node. Perhaps the most natural intuitive idea is this. Extend the concept of a model structure by introducing a set of situations. For each situation \( s \langle D, \psi \rangle \) is a model structure. The meaning or sense of an utterance is then the function \( \psi \) of the root of the tree of the utterance. For example, using the analysis of John hit Mary from Section 3, dropping the redundant notation for the identity function and using the ordinary lambda notation for function abstraction, we obtain as the meaning of the sentence

\[
\psi(1,S) = (\lambda \sigma f(\nu_\sigma (\text{John}), \nu_\sigma (\text{hit}))) \nu_\sigma (\text{Mary}),
\]

but this idea will not be developed further here. Its affinity to Kripke-type semantics is clear.

4. Noun-Phrase Semantics of Adam I

In part A, I proposed and tested a probabilistic noun-phrase grammar for Adam I. The context-free grammar for the noun phrases of Adam I is as in A.3. To the right of the production rules, are also shown the main set-theoretical functions that make the grammar potentially denoting. These semantic functions, as it is convenient to call them in the present context, are subsequently discussed extensively. I especially call attention to the semantic function for Rule 5, which is formally defined.

<table>
<thead>
<tr>
<th>Production Rule</th>
<th>Semantic Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. NP → N</td>
<td>Identity</td>
</tr>
<tr>
<td>2. NP → AdjP</td>
<td>Identity</td>
</tr>
<tr>
<td>3. NP → AdjP + N</td>
<td>Intersection</td>
</tr>
<tr>
<td>4. NP → Pro</td>
<td>Identity</td>
</tr>
<tr>
<td>5. NP → NP + NP</td>
<td>Choice function</td>
</tr>
<tr>
<td>6. AdjP → AdjP + Adj</td>
<td>Intersection</td>
</tr>
</tbody>
</table>
7. **AdjP + Adj**

Table II lists in the fourth column the observed frequency with which the "standard" semantic function shown as above seems to provide the correct interpretation for the five most frequent types (compare Table I). Of course, in the case of the identity function, there is not much to dispute, and so I concentrate entirely on the other two cases. First of all, if the derivation uses more than one rule, then by **standard interpretation**.

<table>
<thead>
<tr>
<th>Noun phrase</th>
<th>Observed frequency</th>
<th>Theoretical frequency</th>
<th>Stand. semantic function</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1445</td>
<td>1555.6</td>
<td>1445</td>
</tr>
<tr>
<td>P</td>
<td>388</td>
<td>350.1</td>
<td>388</td>
</tr>
<tr>
<td>NN</td>
<td>231</td>
<td>113.7</td>
<td>154</td>
</tr>
<tr>
<td>AN</td>
<td>135</td>
<td>114.0</td>
<td>91</td>
</tr>
<tr>
<td>A</td>
<td>114</td>
<td>121.3</td>
<td>114</td>
</tr>
</tbody>
</table>

I mean the derivation that only uses Rule 5 if it is necessary and that interprets each production rule used in terms of its standard semantic function. Since none of the derivations is very complex, I shall not spend much time on this point.

The fundamental ideas of denoting grammars as defined in the preceding section come naturally into play when a detailed analysis is undertaken of the data summarized in Table I. The most important step is to identify the additional semantic functions if any in \$p(p)\$ for each of the seven production rules. A simple way to look at this is to examine the various types of utterances listed in Table I, summarize the production rules and semantic functions used for each type, and then collect all of this evidence in a new summary table for the production rules.

Therefore I now discuss the types of noun phrases listed in Table I and consider in detail the data for the five most frequently listed.

Types N and P, the first two, need little comment. The identity function, and no other function, serves for them. It should be
clearly understood, of course, that the nouns and pronouns listed in these first two lines—a total of 1833 without immediate repetition—do not occur as parts of a larger noun-phrase. The derivation of N uses only Pl (Production Rule 1), and the derivation of P uses only P4.

The data on type NN are much richer and more complex. The derivation is unique; it uses P5 then Pl twice, as shown in the tree. As before, the semantic function for Pl is just the identity function, so all the analysis of type NN centers around the interpretation of P5.

To begin with, I must explain what I mean by the choice function shown above as

```
NP
/|
NP N
```

the standard semantic function of P5. This is a set-theoretical function of A and B that for each A is a function selecting an element of B when B is the argument of f. Thus

$$\phi(A, B) = f_A(B) \epsilon B.$$  

I used 'A' rather than an individual variable to make the notation general, but in all standard cases, A is a unit set. (I emphasize again, I do not distinguish unit sets from their members.) A standard set-theoretical choice function, i.e., a function f such that if B is in the domain of f and B is nonempty, then f(B) \epsilon B is a natural device for expressing possession. Intuitively, each of the possessors named by Adam has such a function and the function selects his (or hers or its) object from the class of like objects. Thus Daddy chair denotes that chair in the class of chairs within Adam's purview that belongs to or is used especially by Daddy. If we restrict our possessors to individuals, then in terms of the model structure $\mathcal{S} = \langle D, v \rangle$, $\phi(A, B)$ is just a partial function from $D \times \mathcal{P}(D)$ to $D$, where $\mathcal{P}(D)$ is the power set of $D$.

The choice function is justly labeled the standard semantic function for P5, but at least four other semantic functions belong in $\phi(P5)$. One of these is the converse of $\phi(A, B)$ as defined above, i.e.,
\( \psi(A, B) = f_B(A) \), which means the possessor is named after the thing possessed. Here are examples from Adam I for which this interpretation seems correct: part trailer (meaning part of trailer), part towtruck, book boy, name man, ladder firetruck, taperecorder Ursula.

The third semantic function is a choice function on the Cartesian product of two sets, often the sets' being unit sets as in the case of Mommy Daddy. Formally, we have

\[ \psi(A, B) = f(A \times B), \]

and \( f(A \times B) \in A \times B \). Other examples are Daddy Adam and pencil paper. The frequency of use of this function is low.

The fourth semantic function proposed for \( \psi(P5) \) is the intersection function,

\[ \psi(A, B) = A \cap B. \]

Examples are lady elephant and lady Ursula. Here the first noun is functioning like an adjective.

The fifth semantic function, following in frequency the choice function and its converse, is the identity function. It seems clear from the transcription that some pairs of nouns are used as a proper name or a simple description, even though each noun is used in other combinations. (By a simple description I mean a phrase such that no subsequence of it denotes (see Definition 4). Some examples are pin game and Daddy Cromer.

I do not consider in the same detail the next two most frequent types shown in Table I, namely, AN and A. The latter, as in the case of N and P, is served without complications by the identity function. As would be expected, the picture is more complicated for the type AN. Column 4 of Table I indicates that 91 of the 135 instances of AN can be interpreted as using intersection as the semantic function. Typical examples are these: big drum, big horn, my shadow, my paper, my tea, my comb, oldtime train, that knee, green rug, that man, poor doggie, pretty flower. The main exceptions to the intersection rule are found in the use of numerical or comparative adjectives like two or more. Among the 116 AN phrases standing alone, i.e., not occurring as part of a lower utterance, 19 have two as the adjective; for example, two checkers, two lights, two socks, two men, two boots, two rugs. No numerical adjective other than two is used in the 116 phrases.
I terminate at this point the detailed analysis of the Adam I corpus, but it should be evident that this is only a beginning. For extensive analyses along the same lines, I refer to Smith (1972) and Suppes, Smith and Léveillé (1972).
BIBLIOGRAPHY

Ellis, C., Probabilistic Languages and Automata, Doctoral Dissertation, University of Illinois at Urbana-Champaign, 1969.


Suppes, P., Probabilistic Grammars for Natural Languages, Synthese 22 (1970) 95-116. (a)


REFERENCES

*This research has been supported by the National Science Foundation under grant NSF-GJ-443X. I am indebted to Pentti Kanerva for help in the computer analysis and organization of the data presented in Part B, Section 4, and I am indebted to Dr. Elizabeth Gammon for several useful ideas in connection with the analysis in Section 4. D.M. Gabbay and George Huff have made a number of penetrating comments on Part B, Section 3, and Richard Montague trenchantly criticized an unsatisfactory preliminary version.

+ I am indebted to D. Reidel Publishing Company for the permission to draw part of the material from the two publications.

1 Roger Brown has generously made the transcribed records available and given us permission to publish any of our analyses.

2 W.C. Watt has called my attention to an article by Harwood(1959), which reports some frequency data for the speech of Australian Children, but no probabilistic grammar or other sort of model is proposed or tested. As far as I know, the explicit statistical test of probabilistic grammars, including estimation of parameters, has not been reported prior to the present paper, but given the scattered character of the possibly relevant literature I could just be ignorant of important predecessors to my own work.

3 I have let the words of $V$ serve as names of themselves to simplify the notation.

4 As Richard Montague pointed out to me, to make context-free grammars a special case of phrase-structure grammars, as defined here, the first members of $P$ should be not elements of $V_N$, but one-place sequences whose terms are elements of $V_N$. This same problem arises later in referring to elements of $V^*$, but treating elements of $V$ as belonging to $V^*$. Consequently, to avoid notational complexities, I treat elements, their unit sets and one-place sequences whose terms are the elements, as identical.

5 Other possibilities exist for the set-theoretical characterization of possession. In fact, there is an undesirable asymmetry between the choice function for Adam hat and the intersection function for my hat, but it is also clear that $v(\text{my})$ can in a straightforward sense be the set of Adam's possessions but $v(\text{Adam})$ is Adam, not the set of Adam's possessions.
DISCUSSION:

Potts:
Thank you very much.
I think that we might do well to spend just a minute or two before beginning substantive discussion in identifying the topics which the members of the panel would like to discuss so that we can do so in a fairly orderly way. From hearing the papers it struck me that there are at least three overlapping topics which the panelists may wish to discuss. First, as between Dr. Keenan and Professor Suppes, the question whether or not predicate logic should be used in semantics or, more generally, in grammatical analysis. Second, can the type of componential analysis proposed by Professor Joshi be integrated with Dr. Keenan's logical method? And can it also be integrated with the set-theoretical method outlined by Professor Suppes? The third topic is that of levels in Frege's sense. In Professor Suppes' account we have a hierarchical system and a method of generating new sets in a hierarchy, whereas the functors used both by Professor Joshi and Dr. Keenan are all first-level ones. It might also help members of the audience if the panelists would care to put to each other some simple questions on the more technical notions which have been employed in the three papers and which may not have been fully assimilated. Some examples from Professor Suppes' paper are his notion of congruence, of 'the same proposition', and his procedure for the creation of new sets. May I now ask each member of the panel if there are further topics, which he would like to add to this list?

Keenan:
We could discuss the general adequacy of using a context-free grammar to represent natural language sentences and the possibility of defining important semantic relations on the structures generated by such a grammar.

Potts:
Professor Joshi?

Joshi:
Well, I would like to have some discussion about the types of semantic representations that are selected and the extent to which the person, or the persons, proposing such semantic representation feel as their obligation to show how these are eventually mapped...
onto actual sentences or the extent to which they think that this is not their immediate concern.

Potts: Professor Suppes?

Suppes: Yes, I would like to have some discussion of Professor Joshi's interesting analysis of perceptual verbs. When you try to systematize this analysis, do you find in various languages a kind of saturation property? I mean do you find something like the introduction of categories in traditional philosophy so that the verbs saturate the possibilities in different ways, of course, in different languages.

Potts: Good. Well, I suggest that we begin with Professor Joshi's question about types of semantic representation. This would lead us fairly naturally into the question about the context-free grammars and that in turn to the controversy about the use of predicate logic. So, Professor Joshi?

Joshi: Psychologists appear to be interested in such factorizations (e.g., move - cause to change location etc.) They will be interested in such an enterprise only if the factors turn out to be psychologically relevant. They are not necessarily obligated to show how one might go from these semantic representations to the sentences. Within linguistics, whatever representation one may adopt, it is generally understood that it is the linguist's business to show how these representations can be mapped onto sentences. Unfortunately, if you look at some recent linguistics papers, one gets the feeling that even some linguists do not consider this as their responsibility either. The question then is: whose job is it?

Potts: Do you see what either Dr. Keenan or Professor Suppes has said as producing a difficulty here?

Joshi: I would say that if you are starting with the phrase structure tree or other closely related representations, you are closer to realizing that goal because one can see how the surface syntax could also be represented by similar structures. If the formal objects you choose for the semantic representations are also useful in characterizing the surface structures, then you are much better
Suppes: I think it relates to the topic about predicate logic because part of my criticism of the representation of the semantics of ordinary language in predicate logic, whether it is the more powerful version Dr. Keenan presented or traditional first-order predicate logic, is without explicit rules as to how to go from the logic to the ordinary language, to the natural language sentence, then one has a major gap in the analysis. And I think it is a gap in the same way as you are talking about for the componential type analysis. It is characteristic for example, it seems to me, of Fillmore's analyses to have the same weakness. He does not give in many ways very good analyses of verbs, he does not give enough structure to show how in a systematic way you get to the actual order of words in the surface of natural language.

Keenan: I take the point that we would like to have fully explicit rules, that we would like completely adequate grammars for all languages; I feel that that is an enormous task that is the subject-matter for an entire field, and it is just unreasonable to impose that requirement before we begin. In fact with respect to the logic I have proposed the reasons these structures have the form they do is because they look a lot like surface structures. I have one major problem: I have lots of pronouns kicking around and I have tried to justify that; in fact, they are extremely useful, and we consider a wide range of languages. I have in fact proposed a format for writing down a relative clause transformation which is independent of the syntax of any language. Of course, it is wrong, but I mean I have actually written down such a procedure and tried to indicate where different categories of languages, so to speak, follow different subroutines in their realization of the relative clause. Some of them delete the pronoun, some of them do not, some of them have the head noun on the right, some on the left, some have a relative pronoun, some do not, and so on. I would submit that this system here is basically very amenable to this treatment, but I do not have a completely worked-out analysis. In the second part of this paper I do argue that the major noun-phrase transformations that linguists have worked with, conjunction, reduction, equi-NP deletion, pronominalization and so on, work very nicely on this system if the identity condition used is "being bound by the same quantified noun-
phrase". To this extent, I think there is a lot of syntactic motivation for doing it this way, (but the) you know completely detailed analysis we may never have them, that is true.

Suppes:
I think we are not really in a disagreement about that, and I want to come back to it, but I would contrast your attitude with the example of Quine, for instance, especially as presented in his book "Word and Object". Many philosophers have not analysed actual usage in ordinary language, but want to eliminate it. There are, of course, proper uses for such a reductive analysis. One can be interested in what is the minimal apparatus to be used to have a given expressive power, but I do feel that much analysis that has been given by philosophers is opposed to what you are really after, and does not recognize the importance of trying to bridge that gap. Now my second remark is this. It does seem to me there are two kinds of strategies and I am not suggesting for a moment that both should not be explored, but there is a strategy different from yours. The strategy I am advocating is to deal with a smaller fragment of the natural language, but then to do the semantics directly on that structure and it is obviously too early in the game to see which strategy can be pushed further. But that is an alternative to what you are proposing.

Keenan:
It is an alternative, but I submit it is inadequate on the grounds that the only adequacy criterion that I know which will allow us to choose among various analyses I find hard to apply directly to surface structures. If we claim to represent the meaning of a sentence, let us say, we surely have to show that it has the semantic properties that native speakers judge it to have. If it implies some other sentence, we have to show that; if it presupposes some other sentence, we have to show that; if it is independent of another, if it answers a question, all of this has to be shown - which means that these semantic relations must be defined on the meaning representations; you simply cannot define consequence on sentences that are semantically ambiguous. The definition of consequence and all the other logical relations is stated in terms of relative truth conditions. Now if a sentence could be assigned more than one truth value in a fixed interpretation, you could argue both that it does imply this and does not imply that depending on how you select the truth value that is convenient for your purpose.
It seems to me then that in defining the semantic relations above, on which we get quite reliable intuitions from native speakers, we have to define them on objects which at least factor out the ambiguity of natural language sentences. And to this extent we are moving away from surface structures. Note that any context-free grammar of a natural language is essentially ambiguous.

Suppes:
I agree about the ambiguity problem, but I disagree about the significance of it because the exact point is that in defining consequence one will use the semantic trees of a sentence. You must deal with the ambiguities, but to look at them in my terminology is to deal with the set of semantic trees that represent the meanings, the possible meanings of a sentence. Now it is quite true that the move to the semantic trees is a move away from the surface structure, but, and this is an important but, as far as I am concerned, those trees reflect directly the structure, which is not the case when you do use a separate logic. There is a difference in the intimacy of the connection.

Keenan:
There is no inherent reason why the trees determined by the strings in the logical language could not be exactly the trees we want to make up. Thus, the logic I have made up, determines a set of trees which are in most important respects the trees we need to represent the few properties of natural language that we can say something about at all; I mean there are lots of things I simply do not know. On the other hand, I found in your discussion the representation of meaning similarity was one which ignored the labels on trees. And I find this surprising. Linguists at least are concerned with what the labels are, and it is not hard to find examples of sentences which have exactly the same tree except for different labels and which are judged semantically ambiguous for those reasons.

Suppes:
What is the example?

Keenan:
I will give some, yes. The problem comes up precisely with the consideration of the sort of semantic property that it is not natural to represent in a context-free grammar, namely the binding properties of quantifier phrases. The example I have in mind will be something like this: "John thought that he was sick" let us say. Somewhat grudgingly we have come to regard the sentence as ambiguous
because most linguists found it is; on the one hand "the" can be considered as bound by "John" and on the other hand, it can be considered as referring differently. Now in English that sentence is ambiguous, in my logic it is not. We have got "John x is such that x thought that x was drunk" on one reading and we get "John x x thought that y was sick" on the other reading. And it comes out that these differ only with respect to label now. That is the only difference; otherwise their structure is similar. Many languages in fact preserve this label distinction. In Yoruba we get two different pronouns in surface depending on whether "he" is bound by "John" or not. In a slightly different context we will get different pronouns in Malay; we get different possessive adjectives of the same sort in Swedish. In fact in simplex sentences the distinction is present between reflexives and non-reflexives. O.K. But this one here is ambiguous and as far as I can see the only thing that represents it is a difference in label.

Suppes:
Oh no, that is a complete misunderstanding of my set-up. The two readings have distinct semantic trees, which reflect the semantic ambiguity.

Keenan:
... different reference?

Suppes:
Yes, they have a different reference and therefore they have different semantic trees, and thus you have exactly the two trees you want, the two semantic trees.

Joshi:
I just want to make sure that some of my earlier comments are not misunderstood. To the extent Dr. Keenan's representations are motivated by syntactical considerations, he would have an easier time going from these representations to the surface syntax. The contrast I want to draw is between those representations which have a syntactic-semantic basis and those which have a purely semantic basis (or conceptual basis). Both Dr. Keenan's and Dr. Suppes' representations essentially belong to the former category and therefore, from my point of view the contrast between them is not really very sharp. They both are interested in eventually mapping their structures onto the surface structures.

Suppes:
I do not know how many people here are familiar with Winograd's
work. This is a recent dissertation written at MIT, on the processing of natural language in a computer. The dissertation will be published in the "Journal of Cognitive Psychology". It is a much discussed example of the presumed successful attempt to process natural language in a computer. Winograd is concerned to deal with a simple perceptual situation and to process commands or questions about this situation.

Now I agree very much with your remarks about that, but there is a theoretical point that I would like to amplify, and we may want to get on to this afternoon with our chairman who is unfortunately somewhat muzzled this morning. The point is the following: Winograd makes a very big point that the semantics is in terms of procedures and it is part of his approach not to make a sharp separation of procedures and data. It is natural to ask of such procedures how they can fit in with the purely set-theoretic semantics that I have described here. There are some direct remarks to be made on this topic, for example, in the case of arithmetic. In many cases the denotation of a terminal is not the set but a procedure, where we have procedures we will want to make the set-theoretical operations combine with procedures in a natural way. So, for example, we might replace intersection by conjunction. In terms of how we want to represent semantically the facts of life, the point is not a philosophical commitment to set-theory but the fact that the techniques of set-theory are simply the most familiar and the simplest techniques of computation of this kind. If we have a very firm grip on the situation, the kind of thing Winograd has attempted, we can replace a set-theoretical version by a much more algorithmic or procedure-orientated version and conceptually for my standpoint that is not a major change.

Keenan: I would just like to make one remark about the lexical decomposition of the sort that Professor Joshi was discussing in this respect. I think probably linguists have been simply very unclear about exactly what properties of surface sentences are supposed to be determined by their underlying structures and which of these properties then are supposed to be either preserved or modified in a fixed way by the transformation. In almost all the cases of proposed lexical decomposition in the literature it is not difficult to find cases where plausibly the underlying structures have distinct semantic and syntactic properties from the things you
derived from them. Even the people who advocate this sort of thing have provided the arguments. Postal's article on anaphoric islands will sometimes give sentences like Mary is an orphan as suggested coming from something like Mary is a young person whose parents are dead. But obviously that thing provides many noun-phrases that can be referenced by pronouns and so on which cannot be referenced once we have replaced that sub-tree by a single lexical item like orphan. Similarly the other cases, the "kill" being paraphrased by cause to become dead and so on. Everybody has counterexamples to the full paraphrase of those two surface structures, and I think before we can therefore propose this sort of decomposition, we simply have to stipulate exactly which properties we are considering, which properties of the surface sentences we are considering determined by these underlying structures and only then can we really make a convincing case.

Potts: Well, should we move on now to another topic? Dr. Keenan, perhaps you would like to pursue your point about context-free grammars?

Keenan: It is not natural in a context-free grammar to represent the sort of structures you get with a variable binding operators of any sort; what you normally have to do in a context-free grammar is generate some sort of quantified phrase irrespective of whether anything in the sentence quantified into matches it so that in a context-free grammar you cannot generate only sentences like every man is mortal, but you also necessarily generate things like every man John loves Mary where the subordinate sentence quantified into in no way talks about the phrase announced in the quantifier phrase. From the point of view of natural language this is surely one of the most unnatural things I have ever heard of. You never find in natural language relative clauses like The man that John loves Mary or The man that all girls are mortal. If you announce this in a head noun-phrase, the man that, the next thing that comes along has to talk about it. Now of course you can always fiddle; you can generate the wrong thing to begin with and then put a filter on the end and get rid of the garbage. I submit that is a completely unintuitive way to represent the meanings of natural languages. What we should do is generate what we want to begin with because that is the most explicit and direct way to describe what the natural language structures are.
Suppes:
I think we agree on the point about the variable binding; we could disagree about its use. Still, I can take your point and not quarrel with this as a feature of context-free languages that we would like not to have in order to match as closely as possible the sentences that occur in natural languages. Fortunately I am not really disturbed by your remark, because we can use indexed grammars which are slightly richer than context-free grammars. They introduce indices to keep track of exactly the variable binding problems you raise, and yet it is known that such indexed grammars do not go far into the hierarchy of context-sensitive languages. For example, we can still draw trees. I feel that trees constitute a certain kind of paradise; there is a great naturalness to the use of trees in the analysis of the structure of sentences and I am very reluctant to give up trees. I am not unwilling to give up the more severe restriction to context-free grammars.
I will make another point. I do accept transformations that map trees into trees, using a context-free base. But I think independent of the transformation issue I agree with your remarks about variable binding and in fact we have found in our own empirical work that it is desirable to carry along a rather substantial set of indices along the lines we have been talking.

Keenan:
If you were to modify the grammar in that way, that would be a significant departure from a context-free grammar. The structures determined by these sentences in a standard logic or any of the usual extensions can also be represented as trees. One very slight inexplicitness, namely the sameness of label relation among the different nodes, is only indicated by writing the same label down. There is nothing in the actual tree structure that tells you that and you have to have a clear idea what the vocabulary of labels is so that you know that these two things are not, so to speak, graphemic variants, but they are different occurrences of the same symbol. Basically the tree seems to represent most explicitly the constituent structure properties of the sentence and it leaves slightly inexplicit, I think, the identity of label functions. And if we look at transformations that only changed labels but did not change constituent structure, in a sense the tree structure does not change that much, but other than that I think the proposals we had been making then would be very similar. I mean surely the
grammar of the standard logic cannot be written as an indexed context-free grammar because it is no longer context-free anymore; it generates a bigger class than that.

**Suppes:**
Yes, certainly. If we exclude vacuous quantification, logical notation, then so you cannot say for every x there is a y such that P(y). In this case it is exactly the indexed grammar that we are talking about. Concerning the problem about the labels that you mentioned, it is a question of what you want to buy. If we require as we do in a phrase-structure grammar, and we do our logic that way, that we have a fixed finite set of non-terminals, then of course we must necessarily have this problem of, in my terms, the denotations in a given tree. For example, the same label occurring in various places in a tree will be at nodes that have different denotations. That is an obvious necessary consequence of the fixed finite nature of the set and of course in traditional logic, by traditional I mean say, Hilbert-Ackermann, syntax is not discussed in terms of a phrase-structure grammar, but the difference is not important.

**Keenan:**
If you limit yourself to finite indices, surely there will be semantically essentially ambiguous sentences, cases where the possible antecedents of the pronouns are more in number than the number of indices you have got to discriminate among them - something which is admittedly natural; that is the way natural languages are. The discriminating powers of pronouns can only distinguish masculine and feminine, singular, plural, a few things like that, and if you have more noun-phrases than that, essentially the sentence is ambiguous. This does mean that the only well-understood semantic relation, logical consequence, can, as I see it, not be really defined on these things. I cannot quite get away from that. I think you answered that once but I perhaps did not completely understand the answer. The point is if the structures are essentially ambiguous, we do not have a consequence relation.

**Suppes:**
What I have said is that I agree with your remark about the surface structure being ambiguous, so it means that logical consequence must be defined for the semantic trees, and, exactly as in the pronoun case, you will have very different tree structures and those tree structures will tell you what the different possible
denotations of the pronouns are. So the response is that the dis-
ambiguation occurs at the level of the set of semantic trees of the
surface structure.

**Keenan:**
I see where the issue is. Usually in a standard logic bound
variables are not considered as having denotation at all, I mean
consider a well-formed sentence like *every barber who shaved just
those barbers who do not shave themselves cut himself.* Now what is
the denotation of the last *himself*? Well, there is not any, because
if it is co-referential with its antecedent which it has to be,
there cannot be one. So there is not any object here that you are
referring to. The sole function of that pronoun is to be anaphoric,
but it does not denote, it cannot be treated as a constituent in-
dependent of the noun phrase that binds it. Maybe that gets into a
bit more of, I think, of the kind of important difference between
the intuition of a context-free grammar and the added power you get
with variable binding operators.

**Suppes:**
Well, I think the following is the case. It is true that we can give,
let us say, in first order logic a characterization of satisfaction
that does not require the bound variables to denote. An even better
example perhaps is this. We do not require that syntential
connective denote, for example, we do not in the ordinary definition
of satisfaction or truth for a first order theory have a denotation
for the connective "and". On the other hand in the case of the
variables we must discuss their denotations very explicitly because
we talk about the values of the variables which must lie in the
domain of the model. I completely agree that in a standard classical
formulation we have things that do not denote, but that in terms of
what I am saying do denote. But I do not see that as a major
problem. I also think if I say *John went to the store and he bought
a new snirt*, in ordinary intuitive ideas *he* denotes John, and it is
quite nice to have pronouns denote.

**Keenan:**
Even if . . .

**Suppes:**
You do not think that *he* denotes John?

**Keenan:**
I did not say that pronouns never denote. I am saying that there are
many uses of pronouns whose purpose is simply to indicate cross-
reference in a sense which is captured at least minimally by the use of variable binding operators in logic, where they do not denote. Even if you formulate a semantics of a standard logic in such a way that variables always denote, the denotations are ignored in the truth conditions of a sentence like \( \text{For all except } x \ldots \). It does not matter what \( x \) denotes when we look at the truth of that because we require always that we can look at, we have to look at, all the things. It does not matter what \( x \) refers to; if you make up your semantics your way, it refers to something. I think it is essential for your point that these things do denote because otherwise I do not understand how you will discriminate the trees for sentences which have deictic pronouns in them which are supposed to denote, and ones that have these others in them. I mean originally you said you did it because they denote different things, but in the cases I gave it is quite clear that the pronoun does not denote; in the case you gave it does, o.k., but in the one I gave, \text{The barber who shaved himself} etc. etc. the last \text{himself} does not denote because there is not any such individual. And it is not really understood as a referring expression in that sense.

Suppes: In that particular example I do not agree with you. In the phrase \text{The barber who shaved himself}, the reflexive pronoun \text{himself} has as a denotation the barber.

Keenan: Sorry about the whole example, it was one of those tricky ones. Would you consider \text{every student shaved himself}, what is the denotation of \text{him} then? It does not denote, it is not understood to refer to an entity in the same way as \text{every student shaved him} is. There \text{him} has to denote; we do not know what its referent is, but it denotes, but you are telling me the only way the semantic trees are different is that the \text{him} and \text{himself} in this case denote different things, but I think that is intuitively incorrect because the occurrence of \text{himself} is not understood as a denoting expression.

Suppes: Yes, I think I agree that I must be careful in now I would formulate \text{every man loves himself}, as an expression of the same type that is where there is a quantification binding on the \text{himself}. I will reflect on what I would say about the denotation, I quite agree with you that it is a good case distinguished from the others.
The discussion is now open to the floor. Various people have already asked to speak and I shall ask them to speak in order, so as to keep the discussion on one topic for a period and then move to another topic.

I begin by asking Professor Schnelle to raise a point concerning presupposition, which is, of course, directed to Dr. Keenan.

Schnelle:
This is not really a question, the question is behind my critical remark. The notion of presupposition and all the definitions which were introduced by Keenan are an example of what is usually called a transposed mode of speech which should not be allowed in theoretical linguistics. The result is that some reading of what is presented gives rather odd examples. For instance, in the hand-out of Mr. Keenan's paper we have the example The fact that Fred left early really surprised John, and he said that if the other sentence, Fred left early, is not true then the first one is understood to be vacuous. I find this quite strange. I can very well imagine a situation where Fred left early is in fact false, but still the situation where people communicate the fact that Fred left early really surprised me is not at all understood as being vacuous. Such a situation may be that the communicating people do not know that Fred left early. The others are the examples The man who won won and every student who left early left early and so on, then it is said their natural denials can never be true. Whatever this may mean, if somebody utters the sentence The man who won did not win, he may present a proposition which may be true, or false, or whatever. Now Keenan says: this is contradictory on literal reading. I think presupposition, and that is my main point, should be introduced as being primarily a pragmatic concept. I could give an example of such a definition as I would imagine it: Speaker x presupposes presupposition q in uttering sentence a if and only if whenever speaker x utters sentence a with proposition p as its meaning, proposition p contains q, then during the act of uttering x does not question the truth of q. The linguist very often does not wish to bother about all these pragmatic problems and he could make the following statement: a description for the pragmatic concept of presupposition is usable if it is a basic one in situations which are standard relative to the grammatical analysis I propose. He assumes, of course, that there are such standard situations. Now my definition

Potts:

The discussion is now open to the floor. Various people have already asked to speak and I shall ask them to speak in order, so as to keep the discussion on one topic for a period and then move to another topic.

I begin by asking Professor Schnelle to raise a point concerning presupposition, which is, of course, directed to Dr. Keenan.

Schnelle:
This is not really a question, the question is behind my critical remark. The notion of presupposition and all the definitions which were introduced by Keenan are an example of what is usually called a transposed mode of speech which should not be allowed in theoretical linguistics. The result is that some reading of what is presented gives rather odd examples. For instance, in the hand-out of Mr. Keenan's paper we have the example The fact that Fred left early really surprised John, and he said that if the other sentence, Fred left early, is not true then the first one is understood to be vacuous. I find this quite strange. I can very well imagine a situation where Fred left early is in fact false, but still the situation where people communicate the fact that Fred left early really surprised me is not at all understood as being vacuous. Such a situation may be that the communicating people do not know that Fred left early. The others are the examples The man who won won and every student who left early left early and so on, then it is said their natural denials can never be true. Whatever this may mean, if somebody utters the sentence The man who won did not win, he may present a proposition which may be true, or false, or whatever. Now Keenan says: this is contradictory on literal reading. I think presupposition, and that is my main point, should be introduced as being primarily a pragmatic concept. I could give an example of such a definition as I would imagine it: Speaker x presupposes presupposition q in uttering sentence a if and only if whenever speaker x utters sentence a with proposition p as its meaning, proposition p contains q, then during the act of uttering x does not question the truth of q. The linguist very often does not wish to bother about all these pragmatic problems and he could make the following statement: a description for the pragmatic concept of presupposition is usable if it is a basic one in situations which are standard relative to the grammatical analysis I propose. He assumes, of course, that there are such standard situations. Now my definition
reads: If $S$ is a standard situation for a sentence $q$ relative to my
standard grammatical analysis, then $x$ presupposes $q$ ..., as above,
and if and only if the structural presuppositional notion, which
Keenan for instance has given, is true.
But this reduces everything to a literal meaning.

Keenan:
With regard to the sentences I claimed under my definitions pre-
supposed themselves, The man who won won and so on, I would
certainly agree that a sentence like The man who won did not win
could be used in a meaningful way on many occasions. One of the
things I could produce to help us analyze what its meaning was would
be that if we had understood it in my literal sense, it would be ob-
viously untrue; so the speaker must be getting at something else.
This sort of thing happens all the time. Suppose someone says John
is a good father and John is not a good father. On the face of it
is an obvious contradiction, but we give a normal speaker enough
credit for not being so simple-minded as to say an"obvious" contra-
diction. So we think a ,"he must mean something slightly different
from what it literally means", so we reinterpret good perhaps to be
slightly different in meaning in the second case; so we have not
exactly denied what we just asserted. That I think is the kind of
thing we are doing here: If we adopt the relatively literal-minded
analysis I am proposing, we have a basis for saying what this
extended usage is, I have in no wise described the extended usage,
I have no pretence of doing so, but it does give us a basis for
saying why it is not meant in exactly the literal way. Your alter-
native is a more pragmatically based notion of presupposition, my
initial feeling is that it replaces a clear but very narrow-minded
notion, mine, a very useful notion, by one that covers a lot more
ground but is much less clear and therefore, I think, much less use-
ful.
I do not easily see how I can use your notion to discuss particular
speech of speakers. I imagine I would always be having recourse to
a d h o c criteria. You have to refer to things like The speaker
means $q$. Well, what does that mean? If we talk about speech acts,
we have to consider all the ways of uttering it. If I say John
thinks that all men are mortal, well, I have just uttered the
sentence All men are mortal, but do I mean it or not? And in many
funny cases like this, all the accidental uses, you have to have
recourse to further sort of a d h o c non-sociological analysis
to get out what you mean by speaker means q and so on. Admittedly the notion I have got is an abstract notion, it has nothing whatever to do with speaker's beliefs or intents. I give it as a property of a sentence in the same way that logical consequence or logical implication is defined in terms of truth relations of sentences quite independently of whether anyone utters them or believes them. The question is whether the world is the way the sentence says it is. It is an abstraction, we in a sense abstract away from a speech-situation. My argument is that it is a useful one; it allows me to distinguish without recourse to vague things about, you know, what it means to say that p contains q and x means p and so on, it allows me to distinguish similarities and differences in meanings between sentences of the sort that I mentioned. So I give you a pragmatic answer: I can use it to say interesting things about language.

Potts: Do you want to pursue on that, Professor Schnelle?

Schnelle: Just one sentence: You should have said that it is an abstract.

Keenan: In part II I give a complete definition. It might have been useful to point that out. I take your point.

Potts: Dr. Posner also has a question on presupposition.

Posner: If you take the sentence I thought that he was sick, there is a difference between the information in I thought that, whatever this information may be and the information in he was sick. You describe the difference by saying the one is a presupposition and the other is an assertion. Perhaps the whole sentence is asserted the presupposition being a part of it. But take sentences like That he was sick is possible where he was sick is not presupposed in the sense it follows from the sentence That he was sick. So you must have a concept to describe the difference between the information he was sick and it is possible that, which is very similar to the difference between the two informations contained in he was sick and I thought that.

Keenan: A small point concerning your examples which I am a little unclear about: In none of those cases I thought that he was sick, he was
sick is possible did I say that the embedded sentence was presupposed. I discussed, admittedly, with one example a class of predicates of the surprise sort which take factive embeddings; they can accept the paraphrase the fact that Fred left surprised John. Your examples are simply not the cases I am discussing; I admit, however, your point that there is a distinction in these sentences between the more dominant and the less dominant information, although I think it is actually a much less clear distinction than in the case of the predicates I come to say something about. In very many languages the locutions for I thought that he was sick are almost indifferent as to where the negation is placed: I thought that he was not sick, I did not think that he was sick. The distinction between the levels on which the information is presented there is almost non-existent. You can get other examples where it is not, and I admit that the distinction, the logical distinction I am making between assertion and presupposition does not capture completely the distinction between, so to speak, most dominant and less dominant. It covers several very interesting cases. This amounts to saying that there is more than one way that information can be subordinate in a sentence — which any linguist would agree with, surely. I think I leave my answer at that, if that satisfies you.

Posner:
Just one remark: You said that this distinction between the different dominances, high, less, low, is unclear. I think it can easily be made clear if you take a concept of protest and if you make a test by protesting partial information of the sentence.

Keenan:
If we take the examples you gave, say I think that John is sick. If I respond, "Do you?" as a weak protest, it means "Do you think that John is sick?" Well, there is a difference between "Do you?" and "Is he?" If you could represent in some sort of formally clear way what you meant by natural protest, you know, I could certainly consider using that notion. The advantage of the one I have got, is that it abstracts from this sort of intuition what a natural denial is. We extend the notion to many sentences that do not have any natural denials; conjunctions, for example, make presuppositions here, but there is no natural negation. Disjunctions of sentences in this system make presuppositions, but there is no natural negation of such sentences and so on. That is what I define in
terms of falsehood conditions. The basic cases I made this system up to handle do have natural negations and that works out right. Well, I leave it at that.

Posner:
I think one can define a situation of protesting which is clearly enough and which is restricted enough to get a test for differentiating relevance.

Keenan:
The advantages of the approach I have is that it is all done formally. I know what sorts of things in models look like and what truth values look like and I know how to reason clearly with them. But the more pragmatic notion of a natural protest is something that needs formalizing before we can work with it. But I accept the intuition and we will close the discussion on that.

Potts:
We have quite a number of other questions addressed to Dr. Keenan, but I think we must give him a breather at this point. So I shall ask Herr Wett to put a question to Professor Joshi on componential analysis.

Wett:
I have some remarks on your system of componential analysis. When you have analyzed push, you found at least two elements: move and cause to. I do not see the reason why you stopped there because "cause" can be decomposed again and you find lots of components and you can continue further on. To be honest, I do not think that the two elements are really in the meaning of "push" because you can imagine sentences like I pushed the table, and I pushed it again, but I could not move it.

My question is: why do we do componential analysis?

Joshi:
First of all, the analysis I presented was only an example (as a matter of fact the example is Fillmore's) and some other analysis could also be justified. The purpose of my talk was not to justify one or another factorization but to present the kinds of arguments that are given for justifying factorization. Of course, this does not yet answer your question.

Your question really is: Where do you stop? From a linguist's point of view the main purpose is to establish closer syntactic-semantic correspondences. The idea is that such decompositions would help establish these correspondences better than the
undecomposed items. For example, the do in the analysis of action verbs helps one to relate the question what did he do? to the sentence he studied. One also looks for independent justification. For example, in this case one can claim (as Ross appears to do) that this do is really the phonetic representation of the intentionality predicate that you may want to set up for all action verbs. Whatever other justifications one might give, one always wants to make sure that such decompositions aid in setting up better syntactic-semantic correspondences. A psychologist would like to see these components related to the conceptual categories he may have set up.

Of course, there are the usual considerations which also appear elsewhere in linguistics, viz., we would like to see to what extent the components reappear in other semantic domains, to what extent they group together, to what extent you can compose them to describe lexical items other than those you started out with, etc. Within these guidelines there is room for variation and clearly to that extent the factorizations arrived at are arbitrary.

Potts:
Would you like to pursue that?

Wett:
You can continue the decomposition until you get an infinite description of a sentence like "I push the table".

Joshi:
Oh no, that is not quite the case.

Wett:
Maybe it is an extreme case.

Joshi:
Aside from the considerations that I outlined in my previous answer to you, we have yet another important consideration. Although our components are abstract in the sense that they are not the lexical items themselves, we are clearly interested in having these components or their compositions lexically realized. This constrains us significantly. In fact, if one looks at the components that have been proposed for various verbs, one notices that most of the components are fairly easily lexically realizable. In my talk, I outlined the various positions one can take in this respect.

Potts:
Well, we will come back now to Dr. Keenan. There are questions on two aspects of his formal language, his atomic sentences and his
noun phrases. Although it seems the wrong way round, I am going to
take a question on the noun phrases first because that on the
atomic sentences will lead us into questions about the reference of
pronouns which will invoke Professor Suppes as well. So I am going
to ask Mrs. Barth to speak next.

Barth:
I would like to propose the following questions to Mr. Keenan:
How can in your logic a negation sign which is a first prefix of a
noun phrase be moved beyond the quantified noun phrase, or are they
not moved but absorbed by these noun phrases? Can you in your logic
avoid a conclusion that One cat has two tails starting from the
truth premise that No cat has two tails?

Keenan:
First as regards how negation works in these sentences, it behaves
in a completely standard way. Let me just illustrate in one phrase
what the actual rule in the syntax of the logic is. It says roughly
that if S is a sentence and this thing NP is a noun phrase, a
quantified noun phrase, and x is a variable pronoun, then (NPx,S)
is a sentence, and we require that x occur free in S and so on.
If S is a negative sentence, well, we get every student was not
drunk. Quantified noun phrase addition is the same rule if S is an
affirmative sentence or S is a negative sentence. In other words,
the noun phrases make sentences out of sentences, we can negate
them or we can operate on negative sentences. So there are two
completely natural representations for All mathematicians are not
smart. Let us try your sentence One cat has two tails, I did not
explicitly put in number quantifiers but we can treat them that
way, i.e. as quantifiers, let us say: (a cat x)(2 tails y)(x has y)
I can represent here in a perfectly standard way, a sound way, the
usual quantifier ambiguities in, every boy loves a girl, some girl
loves every boy and so on. I take it that that basic way would
handle your No cat has two tails or whatever it is. I do not know
how to handle the essential possession of has, and it seems to me
pretty silly to think of it like a two-plan predicate for the two
tails that the cat has are something that does not exist
independent of the cat.
So I am not bothered by that particular syntax, but the basic
relative scope ambiguities of the quantifiers come out exactly
right in this system. Regarding the specific problems on
existential instantiation I have a complete definition of truth in
a model, but I have not written down a deductive system. So off the top of my head I do not know what a rule of existential instantiation would look like; so I will not try and make it up on the spur of the moment.

Potts:
There are now two questions on the atomic formulas which we can best take together, the first is from Dr. Posner.

Posner:
Could you please make some more remarks on the status of your noun phrase representations?

Potts:
The second question is from Dr. Stachowitz on the reference of pronouns.

Stachowitz:
I do have some questions about representations of quantifiers. Could you try and show me how you represent "All my friends, who are in Italy now, are wise," and the same sentence beginning with "Almost ...". And then the same sentences without commas, and also these sentences beginning with "All my friends but one ...".

Keenan:
The justification for treating complex noun phrases in the way that I am doing is that if I do it this way, I can show that certain sentences are logically independent, have certain topical consequences, and that certain sentences make certain presuppositions. There is no guarantee whatever that it could not be done in some other way; this just happens to be a way I have proposed because it gets me the comest results for the class of sentences I was looking at. Let me give just one or two examples. My rule says I get a quantified noun phrase as follows: Step one: Put any determiner "some", "every" or "the" in front of a common noun followed by a pronominal index. O.K. Determiner, noun, x, that is a simple quantified noun phrase, this gets put into quantified sentences by a rule that says: Take any sentence S with x is free, then put a quantified noun phrase in x in front of the sentence. That is a quantified sentence. We have a second definition of a quantified noun phrase: Take any quantified noun phrase, and here is a lot of freedom, convert it, sorry, take any quantified sentence, convert it into a noun phrase, and here all we really need to do is give that sentence a pronominal index because we are going to use that for quantifying into further sentences; so I take simply this
thing, call it $x$ for the moment, and write an index after it and call the whole thing a quantified noun phrase. You can do much more, you can put a relativization marker after it, here, if you wanted to, but you do not actually have to make the distinction between a head noun and a subordinate sentence because it is already marked here, by the parentheses. Now the motivation for distinguishing that structure in the sentence is that in order to practically state the truth conditions and falsehood conditions which determine the consequences and presuppositions of the sentence, it is natural to be able to reference this noun phrase; for example, suppose I want to show that each student in Philadelphia likes horses. I want to show that presupposes there exist students in Philadelphia; that is if it is false, we still understand it there are students in Philadelphia but there is at least one who does not like horses. Well, since I have this whole thing as a noun phrase, it looks very roughly like, something like, you know, Every student $x$ is such that $x$ in Philadelphia... the whole thing is a noun phrase, $x$ likes cheese. Now, this is the thing that I have to stipulate the truth-in-the-model conditions for, and I stipulate it as true, roughly, just in case the set which this noun phrase is interpreted as is not empty and this sentence $x$ likes cheese holds of every member of the set. This sentence is false in an arbitrary interpretation; just in case this set is not empty and this sentence fails for one of its members; otherwise it gets the third value, in particular if the set is empty or if this sentence quantified into had more complexively and made presuppositions that did not have anything to do with this noun phrase.

The reasons then for discriminating this structure is that I find it natural to refer to the sets specified by these noun phrases in the truth conditions. The second reason I indicated briefly in the presentation, namely if I represent them in this way, I get natural representations for transparent and opaque readings. An illustration of this would probably introduce more formalism than is useful at this moment, but it is not very complicated; the entire semantics I have given in about two or three pages in part II of this handout.

Concerning the second question on the representations of the various sorts of relative clauses you suggested the answer is that I do not have any answer. I have no explicit representation in this system of non-restrictive relative clauses. The only relative clauses of
complex noun phrases I even attempt to say anything about are restrictive ones, and I have no way to represent quantifier modifiers like "almost". O.K. I do not know how to write the semantics for that. I can represent these things like "all but one" as long as it is specific I can in effect - in fact in an article in "Foundations of Language" I worked that out slightly informally, so I will not go into it further here.

Stachowicz:
Is it identical to "only one is not in Italy"?

Keenan:
"Only one of my friends is not in Italy". No! I defined it in terms of "only". What I defined was an "only"-quantifier and then the "all but x" was stated as a function of that. I do not actually remember what the function was. But that one could be handled with a certain minimal degree of adequacy here and then it would be the case that the whole thing had the category noun phrase or that "one of my friends" would be a constituent in an underlying structure in that representation.

Stachowicz:
Now since that "only" assumes that "all my friends but one are in Italy" is equivalent to "only one of my friends is not in Italy", I would like to know what your semantic representation of these sentences is. Is it identical or can we derive it from the other?

Keenan:
What I have proposed in that article in fact was - it was in a sort of logical framework - to derive the "all but one" from "only... not". I introduced it by definition - a transformational. You could say that we derived that by a transformation, and so they will have the same underlying representation, yes.

Potts:
While we have this last example, I shall take a question from Professor Staal on Dr. Keenan's notion of 'making more explicit'.

Staal:
I think that is an interesting notion, but I do not know exactly what you need it for. The way in which you define it is all right, in the way you talk about it, it seems to me, you have something in mind that may not be all right. In your examples you indicated that it is not the case that the King of France is bald if it is not the case that there exist, a King of France. In fact it seems to be that one wants to make these presuppositions explicit and that
is what I am interested in in your definition, that those presuppositions also have to be constructed, they cannot just be enumerated. Keenan;

Within this system we can ask the following and interesting question for a given sentence S: Is it possible to find another sentence whose assertions in this logical sense are the consequences of the other one? In other words, a sentence which makes fully explicit all the things that the other sentence implies, its assertions as well as its presuppositions. And the answer is: No, it cannot be done.

The claim here is that I have defined a notion of explicitness and I am not saying that we cannot find a sentence S whose consequences equal the presuppositions of some other sentence. We can do that, and that is a theorem in the system. What I am saying is that it is not the case that for every sentence S there is another sentence T such that the assertions of T, the things it implies without presupposing them - are equal to the non-trivial consequences of the original; in other words, the non-trivial ones are ones that can be false in some state of affairs; they exclude the logically true ones. That fails in this system for several reasons which you might argue are shortcomings of the system. I do not want to argue that one, I will just give the examples. Thus consider proper names in this system: a sentence with a proper name like John left, i.e. John is such that he left, presupposes John exists. We can say that John x is such that x exists. This is one of the sentences that presuppose themselves. It cannot be false; it can be third valued if the name John is not interpreted as an individual of the universal discourse; that is if a non-refering name, and so the sentence can be untrue, but it cannot be false in this system; that is a fact of life of this system. But there is no way in this system to assert the presuppositions of the proper names because any sentence that appears to assert it is one of those that presuppose themselves it does not then logically assert it. You can also not make fully explicit the presuppositions of the factive predicates because the only way I can say that "Fred's leaving surprised John" is to put it in a position where it is presupposed in this system. Maybe I should augment the syntax; I am not claiming otherwise; I can of course take this sentence and look at that as a presupposition in the metalanguage. I can take it up and say this is a presupposition, in fact all its consequences are
presuppositions of the original. In that sense I can make it explicit; what I simply cannot do is fully paraphrase every sentence in this system by another in this system which does not make any presuppositions other than the trivial class of logically true sentences - which under these definitions are presuppositions of every sentence, - something which is sort of natural actually. Note that it is impossible in this system to assert a logically true sentence, and furthermore they never assert anything - which I look upon as a small but positive property of this system. The conjuncted sentences I have proposed were paraphrases, admittedly in a weak sense; but they are not complete paraphrases because they do not make the same presuppositions and assertions. However, it is a common practice in Linguistics and Philosophy to attempt to explain the meaning of a sentence by pulling out each of its individual ideas and saying them separately. To some extent that process can be justified in this system by showing that if we have two sentences, let us say roughly one of them complex without conjunctions, the other a conjunction of somewhat simpler sentences, there is a theorem here that says roughly that the simple one will assert some of what the complex one presupposes, given that they have the same truth conditions to begin with. And this justifies that conjuncting makes some of the original information more explicit. Admittedly, the result is not a complete paraphrase because you have lost the presupposition structure.

Staal:
And what you want to make explicit also by some methods is exactly the presupposition structure.

Keenan:
I can do that, but you will not like it. Note that we can rephrase your question as:
"Is it a case where for every sentence $S$ I can find another sentence $T$, such that the consequences of $T$ are equal to the presuppositions of $S$?" This sentence $T$ then would in a sense tell us what all the presuppositions of $S$ were because they would be $T$ and everything that follows from $T$ and that is all. That answer is yes. Consider: For every sentence $S$ consider the sentence "$S$ or not $S$", and then that will do it. But it somehow does not seem insightful, but still the presuppositions of $S$ are exactly the consequences of "$S$ or not $S$".

but the answer to the other question is "no". If you consider
For each sentence S we can sort of compute, let us say as a function of the structure of S, a finite set of sentences which is such that all the presuppositions of S are assertions of this set of sentences.

We simply cannot make all these presuppositions explicit; that is what we have to show we cannot do in the system - which may be a shortcoming of the system, maybe we want to make it a requirement of logic that it should make everything explicit, but I do not think so because I have tried it out in natural language and it is very difficult to make explicit the presuppositions of proper names without getting completely meta-linguistic and saying the name John denotes. And it is also very difficult to find a sentence which makes explicit, asserts, the presuppositions of a factive sentence. You can make up a conjunction: Fred left early and that fact surprised John, but if we imagine a case where Fred did not leave early, it still seems that the sentence that fact surprised John must have the third value because the fact referred to does not exist. And so it is hard to make that explicit. And that may just be a fact of natural language, it seems it is natural to presuppose this, in a logical sense, and so it cannot be made fully explicit.

I am very interested in Professor Suppes' way of doing semantics; it is familiar, it is intuitive, it is transparent, we know how to work with it. However, the question is whether it is applicable to a natural language. To test it let us take some difficult examples. For instance, when you have a discontinuous morpheme, for instance the familiar "if, ... then" which occurs in different places of a sentence. Of course artificially you can say I represent it as one node, but that is not how it appears in the sentence. You rather would like to attribute to two nodes, to two different words, one semantic concept presumably of implication, and this is not provided by the normal tree, and assigning to each node a semantic quality. Similarly I was not yet quite satisfied with your answer to the enquiry about pronouns and their relation to referentials. You may take another discontinuous morpheme: comparative suppose I say This discussion is of a higher quality than most. Then usually linguists will take that -er of higher and than, and higher-than should be really one morpheme which occurs in a discontinuous way. However, at the same time higher is a modifier of quality and must be akin together and be joined together in form
a higher quality as a noun phrase. So you have here a double-structure; one structure a higher quality which forms a noun phrase, and another, namely -er with than which forms a connexion. Here again it is both discontinuous and over-imposed. If you consider pronouns, an important feature of pronouns is that they are very sensitive to intonation, in particular stress, which again over-imposes another structure. Take an example: If you use an epiphoric pronoun: When he was eighty years old, Bertrand Russel wrote a book. If you stress he, then you say When he was eighty years old, Bertrand Russel wrote a book, then he is no more a referential for Bertrand Russel. And therefore you have to provide some method for stress which is another feature, and you have then to assign to a node not only a grammatical category, but two categories, namely that it is a pronoun and that it is stress. However, then this is not enough because if you consider longer texts, you may have: When he was eighty years old, Bertrand Russel wrote a book, but when I will be eighty years old, I will be senile, where he again is a referential to Bertrand Russel. All this shows that when you are making semantic interpretation of your nodes, you must have quite a lot of rules for how to do it. And those rules are in a way the content of semantic sentence and this is only a representation of one aspect of the result.

Suppes:
I think the examples you bring up are good problems for my kind of formulation. I want to make a couple of distinctions before I come to the pronoun case. Certainly, the discontinuous morphemes are a problem. I have discussed here only what I call the simple case where each node denotes. First, for continuous morphemes, let me give you an example that is not a problem. Take a relation like have diplomatic relations with. For example Does Japan have diplomatic relations with China? Now in this case, I would use a single denotation for the verb phrase, not because this is an ultimate analysis of the phrase, but because it is expedient not to escalate the type so much, so that in actual applications one will let phrases rather than single words denote. And so not every node has a denotation. In the case of discontinuous morphemes, we can begin in this fashion and then by transformation perhaps deal with the matter. But I am not sure this is a satisfactory answer, and I do not want to give a hasty response as to how I would handle discontinuous morphemes.
Now let me turn to pronouns. First on your point that stress or emphasis can carry some additional meaning. I quite agree with you, and one may wish to have more than just a simple denotation carried. So we want to carry at the node some additional information. And that is a case of elaboration, not a fundamental change of my approach to semantics.

I would like to return to give a further example of pronouns, to Keenan's good example of reflexion, and let me indicate how at least in an imprompt fashion I would want to handle that kind of example, just to show how I would hope to move a step further. I think (and maybe this is something that Keenan and I share) that one of the dispairing things about any attempt to be systematic is almost anyone can produce an example you have difficulty with. Let me take three sentences, he loves himself, everyone loves himself and Someone loves himself. I do not want here to go into how I am going to handle quantification; that is a topic that would take a longer discussion. But I would want to say that regarding the himself I would handle it in the quantified cases - I think I perhaps gave a different impression in the earlier discussion -, as reflexive. I would handle it in the same way as here, and let me indicate how I would attempt to do that. So I would roughly have a tree that looks like this

```
S : (x ∈ L ∩ I)
  NP : x  RVP : L ∩ I
    Pro : x  TV : L  RP : I
      he : x  loves : L  himself : I
```

and let me just give on the right-hand side of each node its denotations. So I say a reflexive verb phrase (RVP) goes to the transitive verb (TV) here and the reflexive noun phrase (RP), and then TV go to loves and RP to himself. The key is reflexive. himself denotes the identity relation (I), and loves the binal relation L. The root of the tree, which is labelled S, has T or F as its denotation. Thus (x ∈ L ∩ F) = T if and only if x ∈ L ∩ F.

This then is the kind of analysis I would give of himself and I
think this reflects the character of a reflexive pronoun. The quantification in these two cases takes place on the left side of the tree and is quite separate from the analysis of \textit{himself}. I am under no illusion that what I have said here will necessarily meet the problems of other pronoun cases. I want to be properly and more than duly modest about that point. And also, as I have said, I will need some time to think about discontinuous morphemes. Let me just close by saying that for my standpoint the effort should be to push the framework as far as possible and to recognize where it cannot be made to work.

\textbf{Potts:}
Are you satisfied with that?

\textbf{Hiz:}
Maybe one more example: \textit{He who loves himself does not love others} is a universal sentence in normal understanding. Now you have \textit{he} as a variable and you want that for a variable \textit{he} to have a denotation, however, that would be presumably not quite proper to do here because it is a universal sentence.

\textbf{Suppes:}
I think in that case, we have an idiosyncracy of quantification, that is, where the quantifier is understood. The weakness of my response is that I do not see how to provide an algorithm for recognizing such cases. For example the key here is the modifying relative clause, but I am not at all certain that I have any real clarity on the device that should be used. And by the way, it is even a problem in ordinary mathematical talk whether the variables really are bound or not.

\textbf{Keenan:}
The short remark about your rather ingenious way to handle this on the spur of the moment I think in principle only works where the anaphoric and deictic pronouns are formally distinct in the language. So here it will work for \textit{himself}, but for the example I originally gave, John thought \textit{he was drunk}, we do not have an anaphoric and deictic distinction in the system, we have only got a \textit{he}. So the thing is ambiguous in that respect and to define your function you have to discriminate two different things that, I guess, are in the domain of the function.

\textbf{Suppes:}
You know I have to draw two different trees, one being different from the other to do that.
Keenan:
Yes, right.

Potts:
In drawing the symposium to a close, may I say that there will be an opportunity to continue discussion of these topics this afternoon. There are a few questions from the floor which have not been able to take, but I will pass these on to the chairman of this afternoon's session. It only remains then for me to thank the three symposiasts for the excellent papers they have given us this morning and for provoking a very lively discussion. Thank you very much.