1973

The Case For Probabilistic Grammars

H. L. Berghel
University of Nebraska-Lincoln

TRANSACTIONS OF THE NEBRASKA ACADEMY OF SCIENCES

THE CASE FOR PROBABILISTIC GRAMMARS

H. L. Berghel
University of Nebraska

The purpose of this paper is to briefly examine two proposed extensions of statistical/probabilistic methodology, long familiar to the sciences, to linguistics. On the one hand it will be argued that the invocation of probabilistic measures is indispensable to any sensible criteria of grammatical adequacy, and on the other hand it will be suggested that probabilistic automata can be relevant to studies of language behavior.

I. The fully adequate (categorial/generative) grammar is one with which there corresponds an algorithm by means of which we can (recognize/generate) all and only those syntactically correct sequences in the corresponding language. At this writing, there does not exist any such 'ideal' grammar for any natural language; and as long as this situation remains, it will be necessary for the linguist to 'rank' competing grammars for both reasons of suitability for corpora, and assessment in terms of potential adequacy. Because of the prima facie potential of the transformational grammars introduced since the mid-1950's, linguists have not made any rigorous attempt at providing a measure of descriptive adequacy of grammars. Lately, such intuitive criteria as simplicity, intuitivity, economy, etc. have been levied against competing grammars, in adjudication of adequacy. But these are certainly not the kinds of objective criteria necessary to any independently valuable method of resolving disputes over relative adequacy. This is not to say that these quasi-criteria are without import to the linguist. Surely, in a ceteris paribus situation it is reasonable to prefer the simpler model to the more complex. But up to now there is no method of 'ranking' available by which we can determine when a ceteris paribus situation obtains. In linguistics, just as in the sciences, only when the adequacies of competing models are established are issues of simplicity, economy and the like, germane.

Certainly, the application of statistical/probabilistic procedures to the field of linguistics is not new. Precedents have been established in taxonomic studies, analyses of distributions of word types in corpora (viz. Zipf's Law), etc. But the notion of using an interjacent probabilistic grammar in determining descriptive adequacy is quite innovative. Of the recent developments in this area, perhaps the most notable is that of Suppes (1970). Suppes' motivation for this paper was the disregard of conventional grammatical models to such fundamental and universal characteristics of natural languages as relatively short utterance length, predominance of grammatically simple utterances, etc. It seems irrational to Suppes to be tolerant of grammars which pay an inordinate amount of attention to those syntactic structures

which are 'deviant,' or at least atypical of general usage, and whose relative frequency of occurrence in the corpus is low. To put the matter differently, if any putatively adequate grammar is to be of value, it must be able to account for a sizeable portion of the corpus, thereby identifying those grammatical types which demand further scrutiny. In order to establish the relative values for alternative grammars, Suppes suggests we consult a probabilistic grammar.

The construction of a probabilistic grammar for any given corpus is a relatively easy task. In terms of a generative grammar, we simply assign to each production or rewrite rule in the grammar a certain probability of use in generating the sequences of terminals in the corpus. The parameters of the probabilistic grammar are associated with probabilities of occurrence of all of the productions for any given non-terminal. Thus, the probability of any given structure is expressed as a function of the parameters involved in the requisite productions. Once this probabilistic grammar is formed, a sample is drawn from the corpus at hand, the frequency of occurrence of the varying syntactic structures is calculated, an estimate is placed on our parameters, and a goodness-of-fit is calculated for the grammar at test. The better the goodness-of-fit, the more adequate the grammar for the corpus considered. To illustrate, consider the following productions common to many base components of current transformational grammars:

1. $S \rightarrow NP + VP$
   $NP \rightarrow NP + S$
   $NP \rightarrow (ART) + N + (S)$
   $VP \rightarrow VB + NP + (NP)$
   $VP \rightarrow VB + NP + (S)$

(Of course the optionality of some constituents, indicated by parentheses in the schema above, would have to be treated separately. These productions were selected because they are so common in the literature, not because they lend themselves easily to the methodology.) Inasmuch as the first production is obligatory, it is quite naturally assigned the probability of 1. Since there are two productions each associated with the other non-terminals, we can express their probabilities as monomial functions of parameters α and β, respectively. That is,

2.		probability
	$S \rightarrow NP + VP$	1
	$NP \rightarrow NP + S$	α
	$NP \rightarrow (ART) + N + (S)$	$1 - \alpha$
	$VP \rightarrow VB + NP + (NP)$	β
	$VP \rightarrow VB + NP + (S)$	$1 - \beta$

It is easy to see from the above that the probability of a sequence of the type VB + ART + N + S, say, would be, $(1-\alpha)(1-\beta)$. By appealing to the frequency

distribution of sequences in the corpus, we can place estimates on the parameters, and then test the probabilistic grammar for goodness-of-fit.

An excellent illustration of the application of probabilistic grammars to corpora can be found in Gammon (1970). Gammon is concerned with ranking several primers according to how well the grammars manifest in them correlate with the grammars manifest in the spoken speech of the children for which they were intended. She feels that if the correlation is close, 'only the act of reading and not the structure and sound of the material will be new to the students,' thus facilitating the student's reading progress. The same type of quantatative analysis as outlined above is performed, enabling Gammon to assess the primers in terms of how accurately they represent the grammars employed by the children.

Not surprisingly, work with probabilistic grammars is beset with difficulties. Only those corpora containing the simplest of syntactic structures can be capably dealt with; and in dealing with these, we are at present limited to phrase structure models. But since so many topical issues in theoretical linguistics, like the nature of linguistic universals, are unresolved, it is premature to consider these limitations as vitiation of the study. For one thing, the nature and number of non-terminal constituents is still an open question. It has been suggested (by Bach, McCawley, Fillmore, principally) that radical reconstruction of base components of languages be necessary in order to make any progress toward universality of constituents. This reconstruction may well lessen the number of parameters involved, and simplify the creation of probabilistic grammars for corpora immeasureably. Because the work with probabilistic grammars is the only attempt at establishing objective criteria in order to sensibly evaluate grammars in terms of adequacy, its contributions are important. And if our experiences with the sciences are at all relevant, there is indication that the invocation of probabilistic methodology to linguistics may afford us the only intelligent approach to quantatative analysis.

II. This second section is intended merely to acquaint the reader with some recent research in which probabilistic methodology has been extended to theories of language behavior, learning, etc. Any formulation of this methodology would be beyond the scope of this paper, but the reader is counselled that several rigorous accounts are available for consideration (e.g., Suppes, 1969).

The current contributions in this area rely upon a reintroduction of conditioning theory into notions related to language behavior. They begin ex hypothesi: that stimulus-response theory, in general, is not conceptually inadequate for accounts of language acquisition – only that part of stimulus-response theory which is associated with traditional reflex studies is not full enough to deal effectively with the intricacies of language behavior.

As Suppes points out, rejections of the application of conditioning theory to language behavior frequently confuse 'particular restricted applications of the fundamental theory with the range of the theory itself.' What Suppes claims here is that there is an isomorphic stimulus-response model for all finite automata, and that there is reason to suspect that probabilistic automata can be found which generate languages which are stochastically equivalent to natural languages. One assumption, namely that the internal states of the machine can be likened to the responses of an organism, seems to be of questionable validity (See Block and Fodor, 1972).

Of course it would be infelicitous to take any of the arguments represented in this paper as conclusive. The claim in Section I, however, that tests of adequacy of competing grammars be empirically tied to the relevant corpora lends itself to strong intuitive support. Of course, the case for probabilistic automata rests upon vindication of the stimulus response theory as a rich enough theory to account for any facit of language behavior. Until this fundamental issue is resolved, not a great deal can be said about the future of the study other than that it looks promising.

REFERENCES CITED

Gammon, E. M. 1970. 'A Syntactical Analysis of Some First-Grade Readers'. Ph.D. Thesis, Stanford University. June.

Block, N. J. & Fodor, J. A. 1972. 'What Psychological States Are Not.' Philosophical Review. April.

Suppes, P. 1969. 'Stimulus-Response Theory of Finite Automata.' Journal of Mathematical Psychology.

_____. 1970. Probabilistic Grammars For Natural Languages.' Synthese. December.