Subjective Phrase Structure:
An Empirical Investigation

Z. S. Bond and John Gray

Received July 21, 1972

The present study represents an attempt to investigate the basis of subjective groupings of words in sentences, employing Edwin Martin's subjective phrase structure technique. Results indicate that subjects base their judgments not only on grammatical structure but also on constituent length and on stress and intonation patterns.

INTRODUCTION

Dissatisfaction with the restricted data base of many theoretical syntacticians has led us into a study directed toward eliciting subjective judgments concerning sentences. The scope of such a study is wide and much work remains to be done. However, early results are interesting enough to justify the presentation of a preliminary report, together with some tentative conclusions.

Our study has its origins in a paper by Edwin Martin (Martin, 1970). Martin wanted to “discover empirically whether or not our time-honored, widely accepted views on phrase structure are in fact representative of how language users themselves organize sentences [p. 153].” In order to achieve this aim, he suggested a technique which would produce a “subjective phrase structure tree,” on the basis of the pooled judgments of a group of people.

1 A shorter version of this paper was presented to the Annual Meeting of the Canadian Linguistic Association at McGill University, Montreal, May 1972.
2 Department of Linguistics, University of Alberta, Edmonton, Alberta, Canada.
The technique is quite simple. Subjects were presented with a written sentence and asked to indicate which words of the sentence form natural groups. By counting how many subjects group any two words together, it is possible to arrive at a distance measure showing how closely related those words are judged to be. This distance measure, can, in turn, be represented in a "subjective phrase structure tree."

The technique proposed by E. Martin is attractive, but it can be used profitably only with a clear understanding of the basis on which subjects produce groupings. One of Martin's hypotheses is that "subjective hierarchies are invariant over people [p. 157]." Accordingly, different subjects will give consistent groupings, although they will vary in fineness of analysis.

The present study was undertaken to investigate some of the factors that influence the subjects' groupings of words within sentences. Particularly, we were concerned with the effect of the constituent length and the stability of subjects' judgments of the same constituent in different contexts.

**METHOD**

Sixty-four different sentences were presented, one per page in a booklet, to a group of 35 subjects, who were students in an introductory linguistics course. The sentences were assembled from a list of subject and object constituents, varying in length from one to six words. (In one case, a constituent erroneously consisted of 7 words.) The verb was always a one-word transitive verb in the past tense. The list from which we drew the constituents, and the frequency of use of each constituent, is given in Table I. The sentence varied in length from three words to thirteen, and each constituent appeared in combination with a variety of other constituents. The subjects were tested prior to any discussion of syntactic theory. We used the same order of presentation for all subjects, so that all subjects working on a given sentence would have similar prior experience with the task. No training session was given and, in consequence, early judgments were treated with reservation. The instructions to the subjects were identical with Martin's instructions. These instructions are very broad. Their latitude gives each subject the opportunity to establish his personal style of response. A more constraining set of instructions would have reduced the variation in response style and reduced the usefulness of resulting trees.

The data were processed following the procedure outlined in E. Martin (1970). The subjective grouping of a sentence was coded and submitted to a Fortran computer program which carried out the steps of Martin's method and produced the tree diagram. Each subject/sentence judgment was coded as a vector of numbers. The vector had as many elements as there were words in the sentence being coded. Basically the vector consisted of 1's and 2's. When a