

## INFORMATION TO USERS

This dissertation was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.
2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.
3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.
4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

### **University Microfilms**

300 North Zeeb Road  
Ann Arbor, Michigan 48106

A Xerox Education Company

73-14,378

KALISH-LANDON, Nancy, 1947-  
PROFESSIONAL'S TITLE AS A DEMAND CHARACTERISTIC:  
A POTENTIAL METHODOLOGICAL CONFOUND IN LANGUAGE  
RESEARCH.

The City University of New York, Ph.D., 1973  
Language and Literature, linguistics

University Microfilms, A XEROX Company, Ann Arbor, Michigan

PROFESSIONAL'S TITLE AS A DEMAND CHARACTERISTIC:  
A POTENTIAL METHODOLOGICAL CONFOUND  
IN LANGUAGE RESEARCH

by

NANCY KALISH-LANDON

A dissertation submitted to the Graduate  
Faculty in Linguistics in partial ful-  
fillment of the requirements for the  
degree of Doctor of Philosophy, The City  
University of New York.

1973

This manuscript has been read and accepted for the Graduate Faculty in Linguistics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

January 30, 1973 Peter W. Carey  
 date Chairman of Examining Committee

January 30, 1973 D. Terence Langendoen  
 date Executive Officer

Peter Suedfeld

Peter Carey

D. Terence Langendoen

William Oakes

Supervisory Committee

PLEASE NOTE:

Some pages may have  
indistinct print.

Filmed as received.

University Microfilms, A Xerox Education Company

## ABSTRACT

This study examined the role of intuition in linguistic methodology. Intuition as a scientific method was criticized because the intuitions are usually not tested empirically. When the intuitions have been tested, the outcomes have often not supported the intuitions. Moreover, many sentences in English have questionable grammatical status; as a result, intuitions concerning their grammaticality have often been unreliable in terms of interlinguist and intra-linguist reliability. Evidence was cited to support the contention that bias plays a role in the intuitive process.

Early experimental attempts in linguistics sought to avoid the above difficulties with intuition. These studies were generally confounded with sample selection errors, however, such as unrepresentative subject populations, errors in domain sampling of sentence types, and a failure to take account of situational variables.

The present experiment was conducted to isolate one of these variables, namely a suspected demand characteristic of the professional's title. It was hypothesized that introducing an experimenter as a linguist would affect the subject's grammar, while introducing an experimenter as an anthropologist would affect the subject's story content, compared to simply introducing an experimenter as a researcher (control condition).

Forty-five college students were recruited and randomly assigned to each of the three treatment conditions for a total of 15 per condition.

Each subject was told to tell the story of Cinderella over an intercom to a child and an anthropologist, a linguist, or a researcher. All stories were recorded and later transcribed for analysis. A panel of anthropologists judged all 45 stories for story content and ranked them from best to worst. A panel of linguists judged and ranked all 45 stories in terms of prescriptive grammar.

The results of the experiment indicated that the professional's title (anthropologist, linguist, or researcher) did not alter the grammar of the subject's stories. However, when the experimenter was introduced as an anthropologist, the stories were shorter and had less anthropological content than when the experimenter was introduced as either a researcher or a linguist. An unexpected finding was that some subjects reported not hearing the professional's title. Those subjects who were aware of the title produced uniformly shorter stories with less anthropological content than those subjects who were not aware of the title.

The results of this experiment were explained in terms of evaluation apprehension theory.

## ACKNOWLEDGMENTS

First and foremost, I would like to express my thanks to Dr. D. Terence Langendoen for his constant support. Since he dissuaded me, en route, from registering for a terminal Masters degree, in a real sense I owe my Ph.D. to him.

Special thanks are also due to Dr. Peter Carey for serving as my chairman and for introducing me to psycholinguistics.

The two other members of the dissertation committee, Drs. Peter Suedfeld and William Oakes, made helpful suggestions from the time of the proposal, and the final version owes much to their guidance. I am also grateful to Drs. Joseph Barone and R. M. R. Hall for serving on the final examination committee.

Several individuals contributed to the progress of the experiment. I am grateful to Bell Laboratories for the use of their tape recorder, to Edward and Iris Rosenblum for recording the voice of their son, to Drs. Arnold Lazarus and Peter Suedfeld for the use of their offices and for tolerating the holes I made in the wall, and to all the Rutgers University professors who encouraged their students to participate in the experiment.

I would like to express my gratitude to Ralph Smallberg, Camille Lynch, and Pat Sullivan for serving as the Panel of Anthropologists, and to Helena DePaola,

Judy Farber, and Peter DePaola for serving as the Panel of Linguists.

The aid of Bruce Landon, Richard Pargament, and Sam Kingsley in the analysis of the data is gratefully acknowledged.

Finally, I would like to thank Athena Michailidis for transcribing the tapes and for bringing form to the manuscript. May she never see another version of "Cinderella" again.

## CONTENTS

	Page	
I.	Introduction	10
	A. Intuition in Linguistic Theory	10
	B. Early Experimental Attempts	14
	C. Demand Characteristics	20
II.	Methodology	27
	A. Subjects	27
	B. Apparatus	27
	C. Environment	29
	D. Procedure	32
	E. Transcribing Procedure	34
	F. Judge Recruitment	35
	G. Evaluation Procedure for Anthropologists	36
	H. Evaluation Procedure for Linguists	37
III.	Results	40
	A. Anthropologist Data	40
	B. Linguist Data	56
IV.	Discussion	85
V.	Appendices	91
	A. Script for Experimenter Explaining Task to Subject	91
	B. Questionnaire Completed by All Subjects After Task	93
	C. Coding Sheet for Panel of Anthropologists	95
	D. Coding Sheet for Panel of Linguists	97
VI.	References	99

## LIST OF TABLES AND FIGURES

	Page
Figure 1. Experimental environment	31
Table 1. Anthropologist inter-judge reliability of overall story rankings	41
Table 2. Anthropologist inter-judge reliability for ranked number of words per story	42
Table 3. Anthropologist inter-judge reliability for ranked number of anthropological features per story	43
Table 4. Analysis of variance source table of averaged rankings from anthropologist panel	46
Table 5. Averaged rankings for the treatment conditions from the anthropologist panel	48
Figure 2. Averaged rankings for the treatment conditions from the anthropologist panel	50
Table 6. Analysis of variance source table of the number of words per story	51
Table 7. Averaged number of words per story by treatment condition	52
Figure 3. Averaged number of words per story by treatment conditions	55
Table 8. Interrelationships among the dependent measures from the panel of anthropologists	57
Table 9. Average number of anthropological features per story by treatment condition	58
Table 10. Linguist inter-judge reliability of overall story rankings	59
Table 11. Analysis of variance source table of averaged rankings from the panel of linguists	60
Table 12. Summary statistics for the linguistic features	62
Table 13. Analysis of covariance source table of the number of long sentences	64
Table 14. Analysis of covariance source table of the number of three syllable words	65
Table 15. Analysis of covariance source table of the number of short sentences	66
Table 16. Analysis of covariance source table of the number of contractions	67

	Page
Table 17. Analysis of covariance source table of the number of slang and colloquial words	68
Table 18. Analysis of covariance source table of the number of misused words and expressions	69
Table 19. Analysis of covariance source table of the number of false starts	70
Table 20. Analysis of covariance source table of the number of hesitations	71
Table 21. Analysis of covariance source table of the number of sentence fragments	72
Table 22. Analysis of covariance source table of the number of sprawling and . . . and constructions	73
Table 23. Analysis of covariance source table of the number of incorrect plurals	74
Table 24. Analysis of covariance source table of the number of pronoun agreement errors	75
Table 25. Analysis of covariance source table of the number of verb tense errors	76
Table 26. Analysis of covariance source table of the number of subject-verb agreement errors	77
Table 27. Analysis of covariance source table of the number of imperative sentences	78
Table 28. Analysis of covariance source table of the number of subjunctive sentences	79
Table 29. Analysis of covariance source table of the number of passive sentences	80
Table 30. Analysis of covariance source table of the number of coordinating conjunctions	81
Table 31. Analysis of covariance source table of the number of subordinating conjunctions	82
Table 32. Analysis of covariance source table of the number of conjunctive adverbs	83
Table 33. Raw data table of linguistic features by treatment condition	84

## INTRODUCTION

Linguistic intuitions do not necessarily directly reflect the structure of language, yet such intuitions are the basic data the linguist uses to verify his grammar. This fact could raise serious doubts as to whether linguistic science is about anything at all, since the nature of the source of its data is so obscure. (Bever, 1970)

Chomskyan methodology, with its emphasis on the role of intuition to test a grammar, has been controversial almost since its inception and is gradually being replaced by a more experimental approach to data collection. This study will review some of the major criticisms of the introspective approach in linguistics, along with some of the early post-Chomskyan linguistic studies. Further, it will seek to demonstrate that some of these studies included research artifacts resulting from the sampling procedures employed. Finally, a new source of artifact, the professional's title as a demand characteristic, will be identified and evaluated in an empirical study.

### Intuition in Linguistic Theory

Bertrand Russell (1960) asserted that "if you have a good scientific imagination you can think of all sorts of things that might be true, and that's the essence of science. You first think of something that might be true - then you look to see if it is, and generally it isn't." Many generative-transformationalists

unfortunately have stopped after the imagination, the intuition, which has not been tested, and hence cannot be shown to be right or wrong.

Although intuitions about sentence grammaticality have often not been tested, in many instances psychologists have put linguists' global impressions about language to a test and the outcomes have failed to support these impressions. Many linguists of the Chomskyan tradition considered most speech to be highly ungrammatical (Chomsky, 1965) whereas a sociolinguist has shown that this "fact" is not borne out (Labov, 1970). Further, first language acquisition is not the miraculously rapid process, completed by age four, that these linguists believed (Chomsky, 1965; McNeill, 1966). Rather, it is a slower process (C. Chomsky, 1969). Nor does learning proceed in terms of set milestones invariant in order of acquisition (Bloom, 1970). How many other impressions, or intuitions, in linguistics will be shown to be false once the data are in?

Chomsky repeatedly states that the proper subject matter of linguistics is the study of intuitions, that there is "the necessity for present-day linguistics to give priority to introspective evidence and to the linguistic intuition of native speakers" (Chomsky, 1965). Chomsky hoped that disagreements would turn out to be insignificant, and in 1964 he noted that "consistency among speakers of similar backgrounds and consistency

for a particular speaker on different occasions is relevant." In practice, however, almost every paper in the generative framework includes judgments that are questionable to some or most other generative linguists (Labov, 1971). For example, Labov (1971) points out that the sentence

The fact that he lost disturbed each candidate.  
is cited by Postal (1970) to be ungrammatical because it violates a constraint against Postal's intuited backward pronominalization rule, although most readers accept the sentence. On the other hand, Jackendoff (1968) cites as acceptable

The three of the men that you met yesterday  
have not left.

although Labov finds that most speakers questioned do not accept this judgment. The linguist can then defend his intuition by asserting that he is dealing with his own dialect (meaning that someone disagrees with him) or that the judgment holds for most dialects (in which case he has asked several friends or graduate students who agree with him) (Labov, 1971). This situation raises serious doubts about the inter-linguist reliability.

How reliable are the intuitive data produced by each linguist? As the sentence judgments above suggest, the theorist's desire to have his intuitive data fit his theory can, consciously or unconsciously, influence his judgments. This artifact has been demonstrated often in psychological research and has been termed Experimenter

Bias (Rosenthal, 1966). Experimenter effects can, in the case of linguistics, lead one to be content with evidence for a theory, as opposed to continuing to search one's intuitions for counter-examples. What is missing in the introspective framework is an operationalized way to test those intuitions and a way to prove oneself wrong; the ability to apply Karl Popper's principle that to be right means that "you have finally abjectly, hopelessly failed to prove yourself wrong" (cited in Dingwall, 1971).

Another problem with the introspective approach is the circularity inherent in intuitive grammaticality testing, that is, that the true test for grammaticality is whether the grammar of the language generates it; but this is no help because it is precisely that grammar that linguists are trying to discover (Langendoen, Kalish-Landon, and Dore, 1972).

Further, a basic psychophysical principle from adaptation-level theory (Helson, Blake, Mouton, and Olmstead, 1956) gives reason to doubt the reliability of intuitive judgments, namely the law of contrast which stresses the effect of context on the absolute judgment of the stimulus. For example, if a person dips both hands into lukewarm water after having had one hand in hot water and the other hand in cold water, the hand that was initially in the hot water feels cold and the hand that was initially in the cold water feels

hot. Helson's AL theory has been extended to social behavior and this, Bever asserts, could operate easily on intuitions about grammaticality; the company a sentence keeps can influence the reader's judgments if that sentence stands in contrast, either much more acceptable or much less, to the rest of the sentences in the group (Bever, 1970). Even when the other sentences do not appear to contrast with the stimulus sentence, judgments of grammaticality can still be affected by set bias. Bever hypothesizes that if a group of examples were shuffled, subjects would be less consistent in their grammatical assignments than if the sentences remained in the order in which they were originally conceived by the linguist (Bever, 1970).

#### Early Experimental Attempts

Since the reliability of the introspective method has been found to be questionable and since psychological investigations of some of these intuitions showed them inadequate, many linguists have recently been turning to experimentation to uncover grammaticality and acceptability, using intuitions now only as hypotheses to their studies. Postal (1970) has isolated informally four distinct dialects with regard to pronominalization constraints. The Langendoen, Kalish-Landon, and Dore (1972) study demonstrated that questions such as

Who did you send the man?

are accepted by many native English speakers, contrary to the intuitions of some linguists (Jackendoff and Culicover, 1971). Schapiro (1972) further clarified this issue by testing the acceptability of dative questions of this sort using the indefinite, rather than the definite, article, and finding that the question's interpretation was related to the article in the question.

Archibald Hill (1961) took eight of Chomsky's (1957) sentences from Syntactic Structures and put to test Chomsky's intuitions about them.

1. Colorless green ideas sleep furiously.
2. Furiously sleep ideas green colorless.
3. have you a book on modern music?
4. the book seems interesting.
5. read you a book on modern music?
6. the child seems sleeping.
7. I saw a fragile whale.
8. I saw a fragile of.

Of these sentences, Chomsky writes

Sentences (1) and (2) are equally nonsensical, but any speaker of English will recognize that only the former is grammatical (p. 15).

Similarly, there is no semantic reason to prefer (3) to (5) or (4) to (6), but only (3) and (4) are grammatical sentences of English (ibid.)

... there are deep structural reasons for distinguishing (3) and (4) from (5) and (6). . . . (ibid.).

Yet (1), though nonsensical, is grammatical while (2) is not. Presented with these sentences, a speaker of English will read (1) with a normal sentence intonation, but he will read (2) with a falling intonation on each word. . . . He treats each word in (2) as a separate phrase. Similarly, he will be able to recall (1) much more easily than (2), to

learn it more quickly . . . sentences such as (1) "might" be uttered in some sufficiently far-fetched context, while (2) would never be . . . (p. 16).

Hill randomized the sentences and tested them on professors, graduate students, one undergraduate, and one secretary. Although Hill recognizes his unrepresentative sample, he asserts that Chomsky wrote that the statements applied to "any speaker."

Hill found that only 7 of his 10 informants rejected sentence (2), only 4 out of 10 subjects rejected sentences (5) and (6), and sentence (1), perhaps Chomsky's most famous example of grammatical nonsense, was rejected by one, accepted by nine. Hill also found that his subjects did not find sentence (2) hard to remember; rather, they were intrigued by it and retained it (Hill, 1961). Finally, his subjects did ascribe to each of the sentences a normal intonation pattern rather than simply a falling inflection on each word, as Chomsky predicted. While Hill's method has been criticized, more recent and careful experimentation has shown that his conclusions are valid (Spencer, 1972).

Other studies were faltering attempts at empirical linguistics, generally because they included sample selection errors. The first category to be discussed deals with population sampling. Many of the early studies based conclusions on a subject population of two or three individuals at best (Brown, 1968; McNeill, 1970; Bloom, 1970; Braine, 1970). Subjects were not

randomly selected. In some cases they were drawn from private schools (Anglin, 1970; Chomsky, 1969; Ervin-Tripp, 1970), where they would be expected to be more intelligent than the population at large to which the results were supposed to generalize; in one case only male subjects were used (Ulatowska, 1972), and very often the researcher's child served as subject (Leopold, 1939, 1947; Greenfield, 1967; Bloom, 1971; Braine, 1970). This latter point constitutes a population sampling error because a language researcher might rear a child who was systematically different in his acquisition of language from children with parents with no linguistic training (further research will have to answer this question) and because many of the linguists, like Leopold, were rearing bilingual children and then attempting to generalize their findings to language development in monolinguals. Not only were these populations not chosen randomly, the method of selection was in fact cited in a methods manual as the most effective way to do research, even in light of the fact that such selection procedures greatly limited the subject pool to only a few who could qualify:

. . . these children should (1) have a high rate of speech, (2) be largely intelligible, and (3) have a high rate of verbal interchange with mother or mother substitute. Brown and his coworkers found it necessary to visit a large number of homes in order to select three children fitting these criteria. (Slobin, 1967).

Another type of sample selection error which occurs in the language research has to do with domain sampling. Some of these studies elicited a limited type of sentence, such as commands only (Smith, 1970) or only basic sentences such as the simple, active, affirmative, declarative (SAAD) sentence, and then attempted to generalize their conclusions to English sentences in general (Brown and Hanlon, 1970).

While these studies sampled a minute portion of the language domain, it is questionable whether other studies were sampling the language domain at all; these studies, in their rigorous attempts to structure the language task such that answers obtained from one subject could be easily and reliably compared to answers from other subjects utilized a task unlike normal speech - the very phenomenon they were purportedly studying. The nature of the task was shown to be relevant to what data were collected in a study which administered two tasks. One involved answering questions, a situation approximating normal conversation, to determine how subjects interpreted these questions. The other task used the same questions but required inserting a word to clarify the question in any one of several spaces left between the typed words of the questions. The way the subject interpreted the questions was significantly related to the task he performed (Langendoen, Kalish-Landon, and Dore, 1972). This finding suggests that

there may be some inappropriateness to using sentence completion tasks (Greenbaum, 1970; MacKay, 1966) as a method of data collection as well. A method even further removed from conversational speech is that of Anglin (1970) who requested subjects to sort slips of paper with words on them in order to study word meaning. Most recently, language experiments are being designed with similarity to normal sentence processing in mind (Carey, de Boysson-Bardies, Ioup, and Ahammer, 1971).

Once the methodological problems of population and domain sampling have been dealt with, the situational context still remains as a potential bias; the specific experimental context should be considered a sample of all the contexts in which normal sentence processing occurs. Many recent descriptions of syntactic development have failed to take account of situational variables and freely admit this as a defect (Brown and Fraser, 1963; Klima and Bellugi, 1966). It has been found that the absence of passive sentences from a sample of a child's language could be attributed either to a lack of capacity to produce this structure or to the absence of occasions in the experiment or observation sessions for its production (Svartvik, 1966; Dakin, unpublished research). It would be a sampling bias to sample an unrepresentative situation and then to generalize one's results to other, noncomparable contexts.

The fields of sociolinguistics and anthropological linguistics have illustrated that style shifting is common for adult speakers and that the particular style of speech they choose is related to the experimental or observational context (Labov, 1966; Gumperz, 1965, 1967; Hymes, 1964). Indeed, there is evidence that even children under three years of age differentiate styles according to the age of the addressee (Slobin, 1967). If this is so, it would behoove one doing language research to know exactly what characteristics of language are shifting and what contextual situations are causing the shifts; once these variables are isolated, they can be controlled such that the researcher is on firm ground for concluding that his language experiments or observations are tapping a representative sample of the subject's or informant's speech. If formal language is the phenomenon being studied, then this style can be elicited; if the subject's or informant's relaxed style is the phenomenon being studied, then this style can be elicited. The researcher would know at all times the extent of his external validity.

#### Demand Characteristics

The present experiment was conducted to isolate one of these variables, namely one dealing with the demand characteristics of the experimental context in language research. The term demand characteristics derives from the work of a psychologist, Martin T. Orne

(1969). Orne asserted, like the sociolinguists, that subjects are concerned with their performance in an experiment and that situational variables may affect their performance by distorting the information being elicited. The cues which may alter the subject's performance by communicating to him what he thinks the investigator expects of him and what the investigator hopes to find are the demand characteristics of that situation. Orne (1969) stressed that the shift in performance is not necessarily a deception by the subject; demand characteristics may be so subtle that the subject is not aware of their effect.

Labov (1970) has isolated some of these demand characteristics for language research. It has become apparent through his research that if the investigator's race is different from the subject's race (also Vaughan, 1963; Womack, 1967), if the investigator's interview with the subject or informant is very formal, if the investigator did not know the subject previously, and even if the investigator is much taller than the subject, the speech corpus will be short. Presumably these factors increase the subject's nervousness and he responds to the demand characteristics by talking as little as possible (Labov, 1970).

Beryl Loftman Bailey also recognized the possibility of unconscious demand characteristics and attempted to control for them in her field work by using

nonreactive measures. She observed that Jamaicans who speak Creole unconsciously shift to a form closer to standard English when in the presence of an outsider. She therefore gathered much of her corpus unobtrusively by taking notes on speech behavior on public transportation, street corners, and in local stores (Bailey, 1966). Further, she noted that a tape recorder was a demand characteristic, as informants shifted to non-Creole forms in its presence; she then took mental and pencilled notes instead (see also Labov, 1970; Slobin, 1967).

Anthropologist Margaret Mead (1931) found that the investigator's family status could also be a demand characteristic in fieldwork. She recommended that "a woman investigator should attempt to discover the native attitudes toward barrenness and marital status . . . and then disclose herself as . . . whatever fable is most likely to give the natives confidence in her" (Mead, 1931). This method of deceiving subjects has a long history in psychology and is still considered an ethical procedure to avoid demand characteristics (Stricker, 1967; Stricker, Messick, and Jackson, 1967).

Demand characteristics have also been found to influence the area of semantic research. Cataldo, Silverman, and Brown (1967) replicated a study by Livant (1963) on the semantic differential ratings of nouns and verbs and found Livant's results to be a function

of the demand characteristics associated with the ordering of the nouns and verbs.

A study by Bruce (1964) found that subjects reported the content of a tape recorded "story", played in the presence of noise so intense that the voice was just audible but not intelligible, according to the topic about which they expected to hear. Then the topic was switched while the tape recording was held constant and the subjects' accounts of the content of the tape altered in the direction of the alleged topic. Bruce concluded that semantics can be influenced by the demand characteristic of what one expects to hear.

In a study on language acquisition, Carol Chomsky (1969) neglected to control for demand characteristics in one of her episodes. She placed a blindfolded doll on the table and then asked if the doll was "easy to see"; her results on this task indicated to her that children acquire the meaning of "easy to see" late in their linguistic development. However, the task is somewhat tricky, as the children she tested may have been susceptible to suggestion and perhaps put aside their intuitions about what they thought about the blindfolded doll in favor of the suggested answer. This task tells the child that there is something relevant about the blindfold since they rarely, if ever, see blindfolded dolls, and this might have created a bias to conform to the demand characteristics.

Other demand characteristics which have been isolated in non-language studies but which may be relevant to language research as well include age of investigator (Riesman and Ehrlich, 1961), sex of investigator (Harris, 1971), and social class differences between investigator and informant (Riesman, 1956). The present study was designed to isolate the effects on language of another potential demand characteristic, the professional's title.

It was hypothesized that the professional's, or investigator's, title (field of study) would systematically influence an informant's speech. Although linguistics is not a prescriptive discipline, its associations with prescriptive grammarians in the minds of many individuals not acquainted with linguistic research might cause these individuals to attempt to control their grammar to conform to formal, prescriptive standards if they were confronted with the task of talking to a linguist. Whether they indeed alter their usual sentence structure is an important question for language fieldwork and experimentation. Does the mere fact that one introduces himself to the naive informant as a linguist cause his informant to produce better-formed sentences, judged prescriptively, than he would otherwise produce? This suspected demand characteristic became the basis for this study.

To render untenable the competing alternative

hypothesis that experimenter bias may be responsible for the results, a panel of judges was recruited to analyze the stories collected from the subjects; they were blind to the experiment's purpose. Labov (1966) has also employed judges to rate language data. A child was recruited to be tape recorded; the child's voice played over an intercom served as an excuse to collect a simple story ("Cinderella" was arbitrarily chosen) from the subject to avoid making him feel silly or suspicious in telling a story to an adult investigator and to facilitate collecting the same basic story from each subject. This technique of focusing the subject's attention away from the experimenter to relax him follows from work done by Labov (1970) and Slobin (1967).

In this study, the independent variable was the demand characteristic of professional's title, operationally defined at three levels - linguist, anthropologist, and researcher. The researcher level was included as a control group, as the professional's title is unspecified. The anthropologist level was included to avoid the confound that specificity alone might have accounted for the results (linguist being specific in title vs. researcher being unspecified in title). By including an anthropologist condition it was possible to compare two specific titles, in fact two social science titles, for differences, and any

differences could be attributed to the particular title involved. It was hypothesized that just as subjects might strive to produce "better" grammar for a linguist, they might also strive to produce a "better" story for an anthropologist. The dependent variables were the subject's story content and grammar in his telling of Cinderella. The story content was rated for analysis by a panel of anthropologists. The grammar was rated for analysis by a panel of linguists.

If the investigator was identified as a linguist, it was hypothesized that the subject's grammar would be prescriptively better than if the investigator was identified as an anthropologist or as a researcher. If the investigator was identified as an anthropologist, it was hypothesized that the subject's story content should be better (longer and more involved) than if the investigator was identified as a linguist or as a researcher.

## METHODOLOGY

### Subjects

The sample was drawn from a population of undergraduate students taking psychology courses in the summer session of Rutgers - the State University. They were paid one dollar and also given course extra credit. Their age range was from teenage to middleage, and both sexes, different ethnic backgrounds, and other subject variables were represented in the sample. All subjects were native speakers of English.

The sample of 45 subjects was randomly assigned to the two experimental groups and to the one control group, with 15 subjects per group.

### Apparatus

A male two year old child was recruited to be recorded on a 12 minute continuous tape loop. The tape loop ensured that, regardless of the time of the subject's appearance or the length of time it took him to complete the task, the tape of the child would always be running without interruption. Twelve minutes was considered adequate to mask the repeating nature of the tape and, indeed, no subject reported noticing it. The child was recorded in his darkened bedroom in the evening at which time he felt alone and relaxed enough to talk freely to himself, experimenting with meaningless sound combinations and sing-song babble. The context-free nature of the corpus was an asset because, whatever

the subject said, the child appeared to respond. The tape was carefully edited to erase background noises or context-specific material, and periods of silence were interspersed.

This 12 minute tape loop of the child remained constant throughout the study to control against story differences which might have been elicited as a function of differences in a child's verbal behavior, had a real child been present. Of course, a tape had the additional advantage of avoiding scheduling difficulties of having the same child present for all 45 subjects. According to information gathered from the post-experimental questionnaires and debriefing sessions, this method did not impair credibility.

The tape was played on an Ampex 2000 stereo tape recorder at a constant volume as determined by a few practice sessions. This tape recorder played the tape loop into an adapted intercom in the experimental office, as well as into a speaker in the same office as the tape recorder (see Environment section).

Another tape recorder, a Tandberg tape deck, was used to record the subject's voice from a microphone in the intercom.

The intercom was a one channel communication system. The outside of the intercom box was labelled "Intercom" and a "call" button could be pressed to change it from receiving the sounds of the child to

recording onto the Tandberg. The inside of the intercom was not adequate for the study, so it was adapted. The original microphone was replaced by a microphone of higher quality, an Ampex 1500 microphone, in order to allow for a clearer recording.

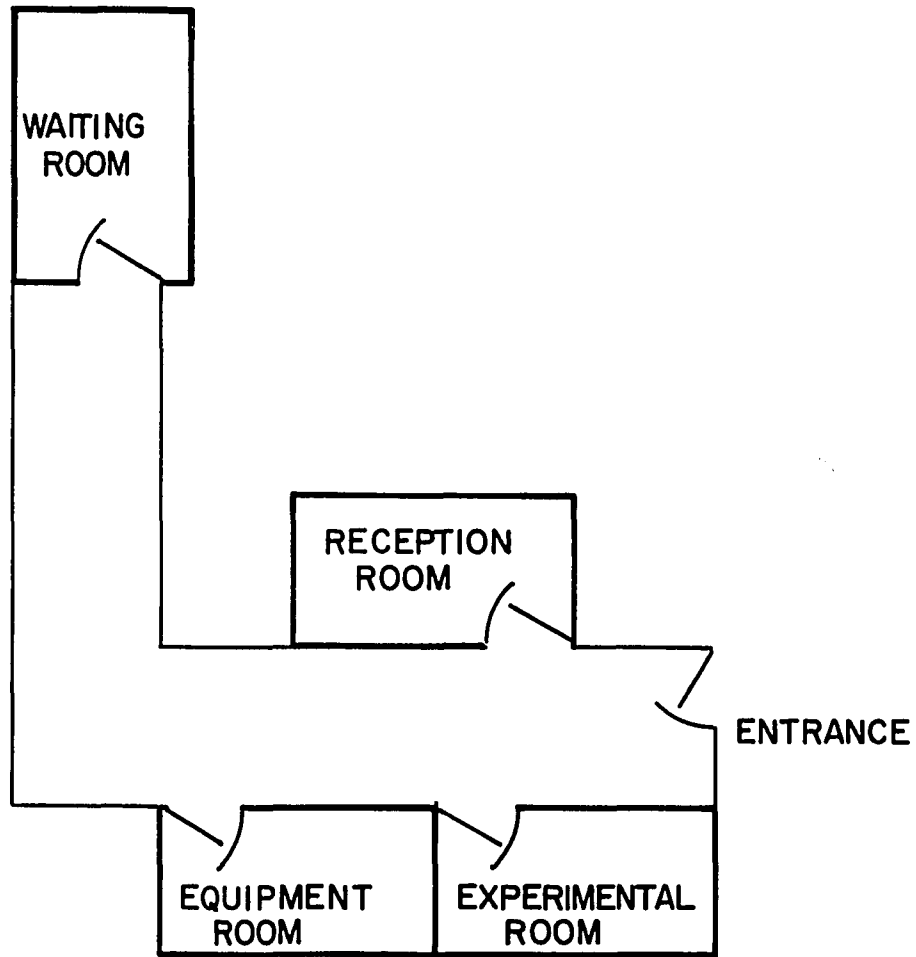
The recording level of the Tandberg was individually adjusted for each subject for maximum clarity. The tapes were recorded at  $7\frac{1}{2}$  i.p.s., and a separate tape was used for each subject.

### Environment

The research environment consisted of four adjacent offices on the university's campus. One office, the Reception Room, was used to greet subjects and, after the task, to administer the post-experimental questionnaire, to pay the subject one dollar, and to debrief him. Another office, the Waiting Room, was used to retain subjects who arrived early. This office was far down the hall from the experimental offices so that it was impossible for the waiting subject to hear the tape of the child; this eliminated any possibility that a subject waiting a long time could detect the repeating nature of the tape and thereby become suspicious. Also, this waiting subject could be kept completely apart from the subject who just completed the task (see Figure 1).

The two remaining offices were next-door to each other; these were the experimental offices. The Equipment Room contained the tape recorder playing the tape

Figure 1. Experimental environment



loop of the child, and also the tape recorder which was recording the subject. There was a sign on this door which read "Experiment in Progress; Do Not Disturb," and the door was locked at all times so that a subject could not open it accidentally and thereby discover the deception. None of the doors had any windows. The volume of the tape recorder playing the tape loop was set such that the child could faintly be heard from outside this office, enhancing the effect that the child was inside.

The Experimental Room was the office in which the interaction took place. It contained a desk and two chairs - one for the experimenter explaining the task, and one for the subject. On the wall was the intercom which played the tape loop from the Equipment Room, and also, when the "call" button was depressed, stopped playing the child's voice through the intercom (although it did not stop the child's voice from being heard outside of the Equipment Room door) and began recording the subject. In this way, only the subject was recorded for analysis, without noise interference from the child tape.

### Procedure

The experimenter always turned on the tape loop before the subject arrived to ensure that credibility was not questioned if the subject came early. The tape that would be used to record the subject was also activated at this time. Then the Equipment Room was

locked. The experimenter waited in the Reception Room for the subject. When the subject arrived, he was taken into the Experimental Room.

The experimenter then introduced herself as a graduate student helping to run the study. This introduction was used to reduce the status of the experimenter, thereby focusing attention away from herself and onto the hypothetical investigator in the other room (Equipment Room). The hypothetical investigator, Dr. Koslin, was fictitious to avoid any possible subject bias in response to the age, sex, race, or demeanor of an investigator. The only fact that the subject could respond to was the investigator's title.

Subjects were assigned randomly in the order of their appearance to one of the three conditions in such a way as to complete each condition at the same rate. This was done to control for any changes that occurred over time.

The subject was informed that a child and Dr. Koslin, a linguist, would listen to him from the next room (experimental condition 1), or that a child and Dr. Koslin, an anthropologist, would listen to him (experimental condition 2), or that a child and Dr. Koslin, a researcher, would listen to him (control condition). Then the subject was told that all he had to do was to tell the story of "Cinderella" to the child and Dr. Koslin, and then come out whenever he finished

his story. The subject was to push the "call" button whenever he was ready to begin his story and to keep the button depressed while talking. If he asked if he should expect Dr. Koslin to interact with him over the intercom, he was told, "No, just tell your story straight through and come out whenever you're ready" (for script, see Appendix A).

Using a story which is part of our culture, "Cinderella", ensured that the subjects' stories could be directly compared for length and content.

Upon leaving, the experimenter suggested that the subject take a few minutes to consider what he wanted to say. Then the experimenter closed the door and left the subject alone. Leaving the subject alone to tell his story to the fictitious child and investigator further minimized the impact of the actual experimenter.

After the subject came out of the room, he was directed back to the Reception Room where he completed the post-experimental questionnaire (see Appendix B), was debriefed, and was paid. He was then free to leave. When it was certain that the subject was no longer in the building, the Tandberg tape recorder in the Equipment Room was stopped, the recorded tape was labelled, and a fresh tape was quickly mounted and begun for the next subject.

#### Transcribing Procedure

At the completion of the study, the 45 subject

tapes were each assigned a code number and were transcribed for analysis into typed protocols. The paid transcriber was not told the hypothesis of the experiment or which subjects were in each condition. She was instructed to be very careful in her transcription, making sure long pauses were transcribed by elipsis (the experimenter and transcriber together decided at the beginning that approximately a 3 second pause or longer be transcribed as an elipsis) and that word fragments (where a subject began a word and then changed his mind and stopped in the midst of it) were also carefully transcribed. An Ampex 2000 tape recorder was used to play the tapes for transcription. Each tape was first transcribed by hand and then the protocol was rechecked against the tape an average of five times to ensure accuracy. Then the handwritten version was typed and this final transcription was also checked against the handwritten version and against the original tape. These typewritten protocols were sequentially numbered and xeroxed copies of the complete set of 45 stories were produced for each judge (see Judge Recruitment section).

#### Judge Recruitment

Three graduate students in anthropology were recruited to serve as an expert panel of judges to evaluate the stories on anthropological criteria. The selected judges were volunteers in response to an advertisement placed on the anthropology bulletin board at

The Graduate School, City University of New York. The notice read: "Anthropologists - rank 45 stories of Cinderella from best to worst by October 30 and earn \$50. An extra \$25 will be paid if you finish by October 16." Three graduate students in linguistics were also recruited to serve as an expert panel of judges to evaluate the linguistic characteristics of the stories. The same conditions of payment were offered.

Three anthropology students responded to the ad, so all respondents became panel members. The anthropologist panel included two females and one male, with an average age of 26 years. They were all second year graduate students with at least one course in anthropological linguistics. Each of the three judges worked independently.

Three linguistics students responded to the ad, so all respondents became panel members. The linguist panel included two females and one male, with an average age of 27 years. All were third year graduate students. Each of the three judges worked independently.

#### Evaluation Procedure for Anthropologists

Each anthropologist judge was given a full set of the 45 stories and 45 coding sheets. On the coding sheets the judges were requested to complete the following type of information for each story: a breakdown of the way the subject described the characters (physical appearance, clothing, possessions, kinship, and

emotions), physical locations (number mentioned, objects described, colors of room or object), and actions in the story, and a total word count. This task was similar to folk myth analysis (McLendon, 1971) (see Appendix C). All breakdowns were encoded as frequencies so that they could be directly compared from story to story. It was hypothesized that the subjects in the anthropologist condition would use more of these features in their stories than the subjects in the linguist or researcher conditions.

After all the stories were encoded in this fashion, each anthropologist was instructed to rank the stories, using these guidelines, from best (1) to worst (45).

When unanticipated difficulties arose with regard to coding particular data, the anthropologist contacted the experimenter and was given further instructions; then this added information was transmitted to the remaining judges.

#### Evaluation Procedure for Linguists

Each linguist was given a full set of the 45 stories, 45 coding sheets, and several xeroxed pages of a grammar book for reference (Birk and Birk, 1969). On the coding sheets the judges were requested to complete the following information for each story:

1. Number of long sentences 10 words or longer (+)
2. Number of words three syllables or longer (+)

3. Number of short sentences 6 words or shorter (-)
4. Number of contractions (-)
5. Number of slang and colloquial words (-)
6. Number of words and expressions misused (-)
7. Number of false starts (-)
8. Number of hesitations (uh, er, etc.) (-)
9. Number of sentence fragments (-)
10. Number of sentences within a sprawling "and-and" structure (-)
11. Number of incorrect plurals of nouns (-)
12. Number of agreement mistakes between pronouns and antecedents (-)
13. Number of verb tense errors (shifts from one to another inappropriately) (-)
14. Number of agreement mistakes between verbs and subjects (-)
15. Number of sentences in the imperative mood (+)
16. Number of sentences in the subjunctive mood (+)
17. Number of sentences in the passive voice (+)
18. Number of coordinating conjunctions (non-run-ons) (+)
19. Number of subordinating conjunctions (+)
20. Number of conjunctive (transitional) adverbs (+)

The (+) and (-), indicated the feature's desirability in prescriptive grammar according to a leading college English Composition course text (Birk and Birk, 1969), were included to aid the judges; the judges were instructed to use these frequencies by dividing the number of positive ones by the number of negative ones in

each story to arrive at a ratio for each story. Then these individual ratios were used to arrive at a ranking of stories from best (1) to worst (45) (for actual coding sheet, see Appendix D). It was hypothesized that the subjects in the linguist condition would use more positive features and fewer negative features, and hence have a higher ratio, than subjects in the anthropologist or researcher conditions.

When unanticipated difficulties arose with regard to coding particular data, the linguist contacted the experimenter and was given further instructions; then this added information was transmitted to the remaining judges.

## RESULTS

Anthropologist Data

The judgements of the three anthropologists on the panel were summarized into three measures for each story: the story's overall rank order, the number of words in the story, and the total number of anthropologically relevant features per story (see Appendix C).

The inter-judge reliability for the overall story rankings is presented in Table 1. Only the rankings between judges B and C were reliably similar. However, the relationship was small ( $r_{BC} = .44, p < .01$ ).

The inter-judge reliability for the rankings of the number of words per story is presented in Table 2. These reliabilities are extremely high, approaching 1.0.

The inter-judge reliability for the rankings of the total number of features per story is presented in Table 3. The pattern of reliabilities was not uniform, with the direction of the coefficient changing between judges A and B and between B and C. Since the pattern of significant relationships in the three tables was not consistent, none of the judges were dropped. In Table 1, to achieve a semblance of reliability one could have dropped judge A; however in Table 3 a positive reliability could have been achieved by dropping judge C. Because none of the coefficients in Tables 1 and 3 were large and all of the coefficients in Table 2 indicated

Table 1  
Anthropologist inter-judge reliability  
of overall story rankings

Judges	A	B	C
A	-		
B	.18	-	
C	.21	.44**	-

\*\*  $p < .01$

Table 2  
Anthropologist inter-judge reliability for  
ranked number of words per story

Judges	A	B	C
A	-		
B	.9958**	-	
C	.9998**	.9984**	-

\*\*  $p < .01$

Table 3

Anthropologist inter-judge reliability for ranked  
number of anthropological features per story

Judges	A	B	C
A	-		
B	.36*	-	
C	-.10	-.37*	-

\*  $p < .05$

uniform inter-judge reliability, there was no good justification for dropping any judge from the major analyses.

In each of the three title conditions there was a group who reported in the post-experimental questionnaire that they remembered hearing the professional's title and a group in each of the three title conditions who reported not hearing the professional's title. Of the 15 subjects in each condition, 8 heard the anthropologist title, 9 heard the researcher title, and 7 heard the linguist title.

The statistical procedure selected to evaluate the effects of treatment conditions was analysis of variance. The use of analysis of variance requires that three assumptions of the data be made (Li, 1964):

- a) randomness of samples
- b) normality of populations
- c) homogeneity of variances

In the present experiment the procedure for assigning subjects to conditions satisfied the randomness of samples assumption which, incidentally, is the most important assumption.

According to the central limit theorem (Li, 1964), as the size of the sample increases, the distribution of means of all possible samples drawn from the same population becomes more and more like a normal distribution, providing the population has a finite variance.

Thus by using the mean judgments of the panel of judges as the dependent variable scores, it can be assumed that these variables will be normally distributed in the case of mean rankings.

The homogeneity of sample variances can be assumed on the same central limit theorem premise as the normality of the samples.

Furthermore, there are additional reasons why analysis of variance was used. First, analysis of variance is an extremely robust test, which means that small departures from the assumptions make little difference in the outcomes of the test. Second, and perhaps more importantly, analysis of variance was used in the present study for the purpose of making weak inferences about the data. The desired inferences were simply whether or not the conditions differed at all, as opposed to the strong inference purpose of estimating the real magnitude of the differences between the conditions.

The overall panel of anthropologists rankings were averaged across judges. These averaged ranks were then used as scores in a 2 X 3 factorial analysis of variance with unequal cell sizes using an unweighted means procedure to accommodate the "missing" data. The summary of this analysis is presented in Table 4.

Both the awareness and the title effects were significant. Further analysis of the title means found

Table 4  
 Analysis of variance source table of averaged  
 rankings from anthropologist panel

Source	SS	df	MS	F	p
Title	868.54	2	434.27	3.44	p<.05
Awareness	1383.21	1	1383.21	10.96	p<.01
Interaction	54.43	2	27.22	< 1	NS
Residual	4922.08	39	126.21		

that the anthropologist condition had significantly lower ranking (researcher vs. anthropologist conditions  $\bar{t} = 2.37$ ,  $df = 39$ ,  $p < .05$ ; linguist vs. anthropologist conditions  $\bar{t} = 2.04$ ,  $df = 39$ ,  $p < .05$ ), while the mean rankings for the researcher and the linguist conditions were not different ( $\bar{t} = 0.04$ ,  $df = 39$ , NS).

The awareness of the professional's title resulted in a uniform reduction in the rankings of the stories, indicating that the stories where the title was heard had less anthropological content. The mean rankings of the stories by the panel of anthropologists for each treatment condition are presented in Table 5 (rank 1 to 45 was best to worst in terms of anthropological judgments). The results are also presented graphically in Figure 2.

The number of words in the stories (length) was also analyzed with an unequal cell size 2 X 3 factorial analysis of variance. The word counts from the panel of anthropologists were averaged to form the scores for this analysis. The summary of this analysis is presented in Table 6. Both of the main effects of title and awareness were significant and again there was no interaction. The mean number of words in the stories is shown in Table 7. Those subjects who were aware of the professional's title produced shorter stories. In addition, regardless of the awareness condition, the stories in the anthropologist condition were

Table 5  
 Averaged rankings for the treatment conditions  
 from the anthropologist panel

Treatment Condition	Anthropologist	Researcher	Linguist	Com- bined
Reported hearing the title	32.29	24.04	27.05	27.67
Reported not hearing the title	24.38	14.33	14.29	17.67
Combined	28.60	20.16	20.24	23.00

Figure 2. Averaged rankings for the treatment conditions from the anthropologist panel

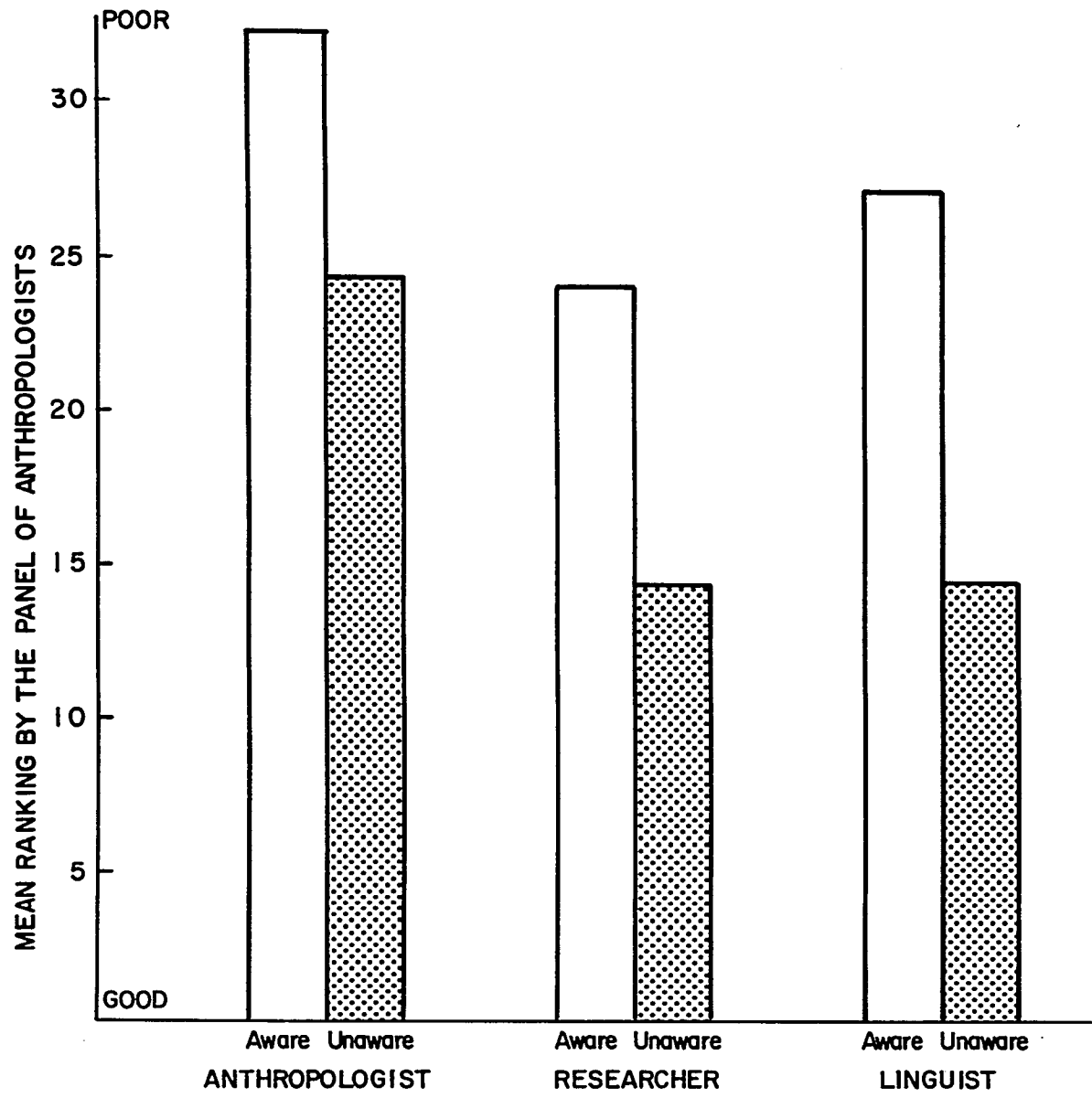


Table 6  
Analysis of variance source table of  
the number of words per story

Source	SS	df	MS	F	p
Title	581,679	2	290,839	3.70	p<.05
Awareness	1,238,755	1	1,238,755	15.77	p<.001
Interaction	174,449	2	87,224	1.11	NS.
Residual	3,063,787	39	78,559		

Table 7  
 Averaged number of words per story  
 by treatment condition

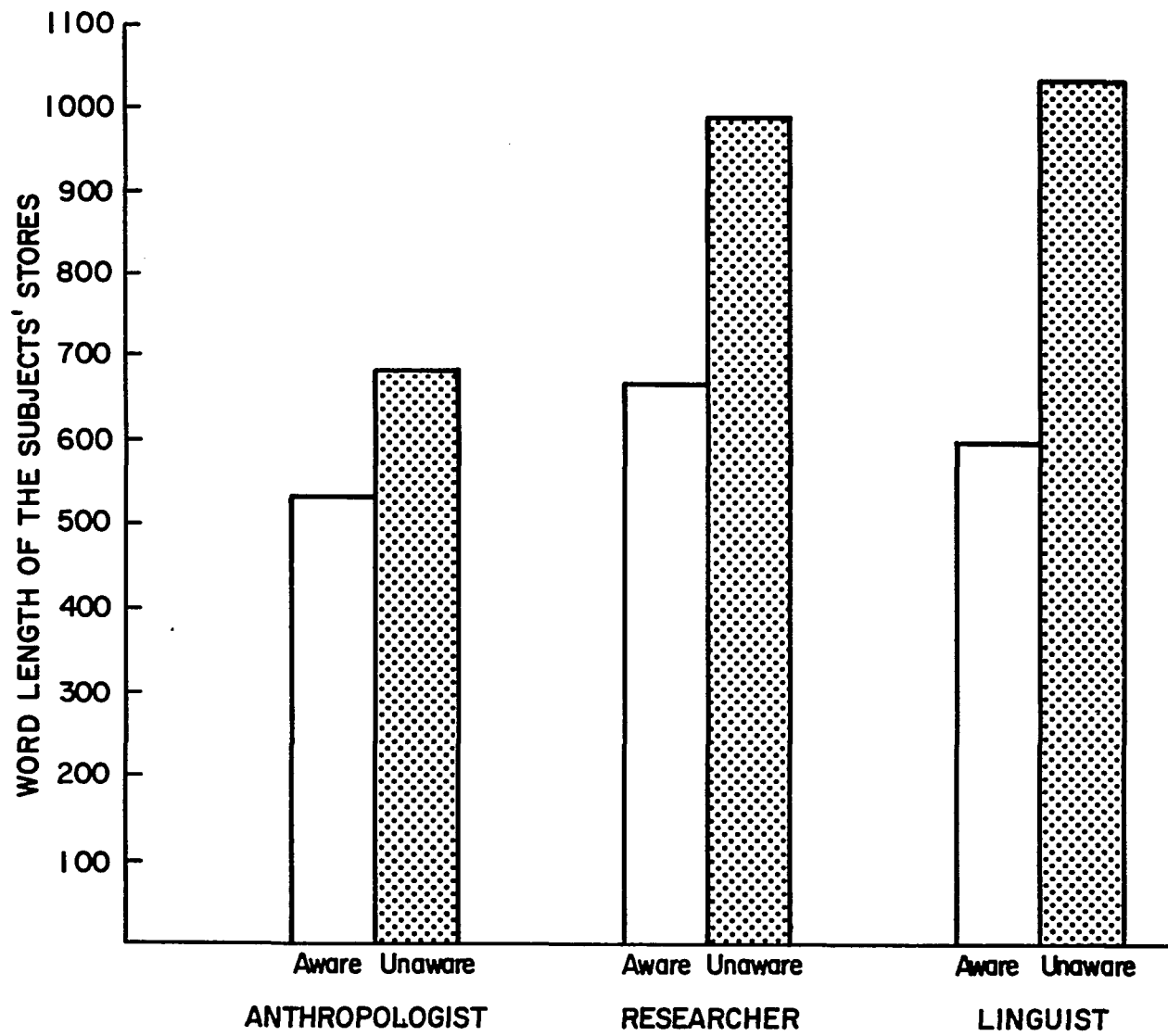
Treatment Condition	Anthropologist	Researcher	Linguist	Com- bined
Reported hearing the title	524.5	674.9	595.7	601.7
Reported not hearing the title	682.4	990.2	1031.3	903.2
Combined	598.2	801.0	828.0	742.4

significantly shorter (researcher vs. anthropologist condition  $\underline{t} = 2.45$ ,  $df = 39$ ,  $p < .05$ ; linguist vs. anthropologist condition  $\underline{t} = 2.25$ ,  $df = 39$ ,  $p < .05$ ). As with the rankings, the average number of words did not differ between the researcher and the linguist conditions ( $\underline{t} = 0.21$ ,  $df = 39$ , NS). A graph of these results is presented in Figure 3.

The fact that the length of the stories differed between conditions is especially important since the other measures of interest (i.e., the number of anthropological features and linguistic features) are correlated with story length, and therefore are confounded with story length. Because this form of confounding could not be experimentally controlled, an attempt was made to statistically unconfound the results on the measures potentially affected by story length. The method used was a 2 X 3 factorial analysis of covariance for unequal cell size with the length of the story as the covariate. This technique evaluates the effects of the treatments in the hypothetical situation where the stories are all the same length (which they were not). The use of analysis of covariance does not imply that the length of the stories per se "caused" any change in the other measures or vice versa, but just that they co-occurred.

The interrelationships among the anthropologists' ranking, the number of words, and the number of

Figure 3. Averaged number of words per story by treatment conditions



anthropological features are presented in Table 8. It is of importance to note that the number of words and the number of features in the story are positively correlated ( $r = .39, p < .01$ ).

Consequently, the average number of anthropological features (averaged for the panel of anthropologists) was analyzed in a 2 X 3 factorial analysis of covariance for unequal cell sizes.

The summary of this analysis is presented in Table 9. None of the effects of the conditions were significantly different, which means that the number of anthropological features in the stories was not influenced by the professional's title or the awareness of that title when the story length is controlled statistically.

#### Linguist Data

The rankings from the panel of linguists were analyzed with a Spearman Rank Order correlation coefficient. The correlations between the judges' rankings in Table 10 were all significant, indicating that the panel's judgments were reliable.

These rankings were averaged across judges. The resulting averaged ranks were then used as scores in a 2 X 3 factorial analysis of variance with unequal cell sizes. The summary of this analysis is presented in Table 11. No differences were found between any of the conditions which indicates that at a global level the

Table 8  
 Interrelationships among the dependent measures  
 from the panel of anthropologists

Dependent measures	Overall rank	Number of words	Number of anthropological features
Overall rank	-		
Number of words	.09	-	
Number of anthropological features	.35*	.39**	-

\*  $p < .05$

\*\*  $p < .01$

Table 9  
Average number of anthropological features  
per story by treatment condition

Source	SS	df	MS	F	p
Title	60.38	2	30.19	<1	NS
Awareness	6.74	1	6.74	<1	NS
Interaction	299.12	2	149.56	1.73	NS
Residual	3276.61	38	86.23		

Table 10  
Linguist inter-judge reliability of  
overall story rankings

Judges	D	E	F
D	-		
E	.72**	-	
F	.45**	.56**	-

\*\*  $p < .01$

Table 11  
 Analysis of variance source table of averaged  
 rankings from the panel of linguists

	SS	df	MS	F	p
Title	276.16	2	138.08	1.07	NS
Awareness	1.74	1	1.74	<1	NS
Interaction	178.29	2	89.14	<1	NS
Residual	5022.33	39	128.78		

grammar of the subjects was not affected by the experimental conditions.

Further analyses were conducted to discover if the experimental conditions affected any of the twenty linguistic features (averaged across judges) that had been summarized in the linguists' rankings. Since the number of instances of a linguistic feature could be affected by the length of story, correlations were computed between each of the measures and length of the stories as judged by the panel of anthropologists. These coefficients along with mean, standard deviation, lowest score, and highest score are presented in Table 12 for the twenty linguistic features. In all but two instances the relationship between the length of the story and the number of occurrences of the linguistic feature is positive and in most of those cases the relationship is significant beyond the .01 level. The two negative correlations occurred with the number of incorrect plurals and the number of subjunctive sentences, both of which were quite rare, occurring less than five times altogether in the forty-five stories in the experiment. One characteristic in common to all twenty variables was high variability, as indicated by both the standard deviation and the range from lowest to highest score. Only four features - long sentences, short sentences, coordinating conjunctions, and subordinating conjunctions - were found in all forty-five

Table 12 Summary statistics for the linguistic features

Feature	Mean	Standard deviation	Lowest score	Highest score	Correlation with story length
Long sentences	20.02	11.43	2	55	.80**
Three syllable words	13.22	5.65	3	27	.69**
Short sentences	3.56	6.01	0	30	.50**
Contractions	7.73	5.58	0	21	.70**
Slang and colloquial words	4.67	5.00	0	20	.32*
Misused expressions	4.18	2.72	0	10	.42**
False starts	9.16	7.07	0	31	.34*
Hesitations	9.16	7.88	0	35	.13
Sentence fragments	1.00	1.07	0	4	.49**
and . . . and . . .	8.56	5.92	0	26	.47**
Incorrect plurals	0.09	0.42	0	2	-.17
Pronoun agreement errors	0.42	0.62	0	3	.19
Verb tense errors	1.29	1.90	0	9	.20
Verb-subject agreement errors	0.22	0.47	0	2	.28
Imperative sentences	0.96	1.31	0	6	.53**
Subjunctive sentences	0.04	0.30	0	2	-.11
Passive sentences	0.93	1.16	0	4	.25
Coordinating conjunctions	25.56	14.01	5	71	.77**
Subordinating conjunctions	6.20	4.42	1	18	.58**
Conjunctive adverbs	2.40	1.42	0	7	.69**

\*p < .05

\*\*p < .01

stories. This would seem to indicate that there are sizeable style differences between subjects.

The numbers of each of the linguistic features were analyzed with a 2 X 3 factorial analysis of covariance for unequal cell sizes. The length of the story was used in all analyses as the covariate. It will be recalled that there were differences between the treatment conditions on the length of the stories. The analysis of covariance effectively analyzed the number of linguistic features that have been produced had the stories all been of equal length.

The summaries of these analyses are presented in Tables 13 through 32, and the raw data in Table 33. In none of these analyses were any of the treatment effects significant which suggests, as did the analysis of linguistic rankings, that the experimental conditions did not influence the grammar of the subjects in a consistent manner.

The linguistic data results of the experimental manipulations of professional's title and the reported awareness of that title can be summarized quite simply. There were no detectable differences between the experimental conditions on any of the measures of grammar recorded.

Table 13  
Analysis of covariance source table  
of the number of long sentences

Source	SS	df	MS	F	p
Title	108.05	2	90.03	1.87	NS
Awareness	2.70	1	2.70	< 1	NS
Interaction	69.02	2	34.51	< 1	NS
Residual	1832.92	38	48.23		

Table 14  
Analysis of covariance source table of the  
number of three syllable words

Source	SS	df	MS	F	p
Title	4.04	2	2.02	< 1	NS
Awareness	1.73	1	1.73	< 1	NS
Interaction	38.88	2	19.44	1.07	NS
Residual	688.18	38	18.11		

Table 15  
Analysis of covariance source table of  
the number of short sentences

Source	SS	df	MS	F	p
Title	74.25	2	37.13	1.29	NS
Awareness	11.73	1	11.73	< 1	NS
Interaction	12.06	2	6.03	< 1	NS
Residual	1093.29	38	28.77		

Table 16  
Analysis of covariance source table  
of the number of contractions

Source	SS	df	MS	F	p
Title	12.75	2	6.37	<1	NS
Awareness	15.64	1	15.64	<1	NS
Interaction	23.16	2	11.58	<1	NS
Residual	645.21	38	16.98		

Table 17  
 Analysis of covariance source table of the number  
 of slang and colloquial words

Source	SS	df	MS	F	p
Title	6.64	2	3.32	<1	NS
Awareness	0.75	1	0.75	<1	NS
Interaction	13.76	2	6.88	<1	NS
Residual	970.24	38	25.53		

Table 18  
 Analysis of covariance source table of the number  
 of misused words and expressions

Source	SS	df	MS	F	p
Title	0.51	2	0.26	<1	NS
Awareness	0.13	1	0.13	<1	NS
Interaction	1.61	2	0.80	<1	NS
Residual	266.32	38	7.01		

Table 19  
Analysis of covariance source table  
of the number of false starts

Source	SS	df	MS	F	p
Title	53.33	2	26.66	< 1	NS
Awareness	59.77	1	59.77	1.27	NS
Interaction	36.47	2	18.24	< 1	NS
Residual	1790.17	38	47.11		

Table 20  
Analysis of covariance source table  
of the number of hesitations

Source	SS	df	MS	F	p
Title	12.09	2	6.05	< 1	NS
Awareness	33.92	1	33.92	< 1	NS
Interaction	21.55	2	10.77	< 1	NS
Residual	2629.65	38	69.20		

Table 21  
 Analysis of covariance source table of  
 the number of sentence fragments

Source	SS	df	MS	F	p
Title	3.24	2	1.62	1.81	NS
Awareness	0.03	1	0.03	< 1	NS
Interaction	0.60	2	0.30	< 1	NS
Residual	34.07	38	0.90		

x

Table 22

Analysis of covariance source table of the number  
of sprawling and . . . and constructions

Source	SS	df	MS	F	p
Title	95.97	2	47.98	1.72	NS
Awareness	13.18	1	13.18	< 1	NS
Interaction	47.90	2	23.95	< 1	NS
Residual	1057.64	38	27.83		

Table 23  
Analysis of covariance source table of  
the number of incorrect plurals

Source	SS	df	MS	F	p
Title	0.10	2	0.05	<1	NS
Awareness	0.09	1	0.09	<1	NS
Interaction	0.20	2	0.10	<1	NS
Residual	7.00	38	0.18		

Table 24  
 Analysis of covariance source table of the  
 number of pronoun agreement errors

Source	SS	df	MS	F	p
Title	0.43	2	0.21	< 1	NS
Awareness	0.89	1	0.89	2.34	NS
Interaction	0.67	2	0.33	< 1	NS
Residual	14.49	38	0.38		

Table 25  
Analysis of covariance source table of the  
number of verb tense errors

Source	SS	df	MS	F	p
Title	3.97	2	1.99	<1	NS
Awareness	0.56	1	0.56	<1	NS
Interaction	3.74	2	1.87	<1	NS
Residual	144.28	38	3.80		

Table 26  
 Analysis of covariance source table of the  
 number of subject-verb agreement errors

Source	SS	df	MS	F	p
Title	0.60	2	0.30	1.50	NS
Awareness	0.01	1	0.01	<1	NS
Interaction	0.81	2	0.40	2.04	NS
Residual	7.67	38	0.20		

Table 27  
Analysis of covariance source table of the  
number of imperative sentences

Source	SS	df	MS	F	p
Title	2.32	2	1.16	<1	NS
Awareness	0.02	1	0.02	<1	NS
Interaction	2.01	2	1.00	<1	NS
Residual	50.25	38	1.32		

Table 28  
 Analysis of covariance source table of the  
 number of subjunctive sentences

Source	SS	df	MS	F	p
Title	0.15	2	0.07	< 1	NS
Awareness	0.17	1	0.17	1.91	NS
Interaction	0.17	2	0.09	1.01	NS
Residual	3.39	38	0.09		

Table 29  
Analysis of covariance source table of  
the number of passive sentences

Source	SS	df	MS	F	p
Title	3.22	2	1.61	1.20	NS
Awareness	0.29	1	0.29	<1	NS
Interaction	0.33	2	0.16	<1	NS
Residual	51.08	38	1.34		

Table 30  
 Analysis of covariance source table of the  
 number of coordinating conjunctions

Source	SS	df	MS	F	p
Title	298.62	2	149.31	2.04	NS
Awareness	44.85	1	44.85	<1	NS
Interaction	387.92	2	193.96	2.64	NS
Residual	2787.10	38	73.34		

Table 31  
 Analysis of covariance source table of the  
 number of subordinating conjunctions

Source	SS	df	MS	F	p
Title	49.12	2	24.56	1.79	NS
Awareness	0.89	1	0.89	<1	NS
Interaction	0.09	2	0.05	<1	NS
Residual	520.57	38	13.70		

Table 32  
 Analysis of covariance source table of  
 the number of conjunctive adverbs

Source	SS	df	MS	F	p
Title	1.71	2	0.86	<1	NS
Awareness	1.07	1	1.07	<1	NS
Interaction	0.67	2	0.33	<1	NS
Residual	43.35	38	1.14		

Table 33 Raw data table of linguistic features by treatment condition

Treatment conditions	Linguistic features																									
	Long sentences	Three syllable words	Short sentences	Contractions	Slang/colloquial words	Misused expressions	False starts	Hesitations	Sentence fragments	and . . . and . . .	Incorrect plurals	Pronoun agreement errors	Verb tense errors	Verb-subject errors	Imperative sentences	Subjunctive sentences	Passive sentences	Coord. conjunctions	Subord. conjunctions	Conjunctive adverbs						
Anthropologist aware	06	09	02	00	02	01	01	12	00	04	00	00	00	00	00	00	00	00	00	00						
	07	15	00	07	00	02	00	00	00	10	02	00	00	00	00	00	00	00	00	00						
	15	10	01	07	01	01	00	00	00	10	02	00	00	00	00	00	00	00	00	00						
	06	03	01	02	06	02	06	03	00	03	00	00	00	00	00	00	00	00	00	00						
	16	07	01	05	01	03	05	07	01	02	00	01	00	00	00	00	00	00	00	00						
	16	11	08	02	01	01	08	06	00	01	00	00	00	00	00	00	00	00	00	00						
Anthropologist unaware	13	10	01	19	20	05	14	17	01	19	00	00	03	00	01	00	00	00	00							
	26	16	04	06	01	05	10	06	01	05	00	00	00	00	01	00	02	24	09							
	26	17	04	07	02	04	09	03	00	16	00	01	02	00	02	00	00	31	07							
	11	19	01	02	03	03	03	09	01	06	00	01	02	00	00	00	00	02	08							
	21	12	02	10	02	03	10	06	00	12	00	01	02	00	01	00	00	42	03							
	10	05	01	03	01	03	03	09	00	05	00	00	00	00	00	00	00	13	03							
Researcher aware	11	17	00	03	02	04	29	11	01	06	00	00	01	00	00	00	00	18	06							
	28	16	30	21	18	09	15	15	04	05	00	01	03	01	06	00	02	35	06							
	15	10	03	06	00	04	12	06	00	04	00	01	00	00	00	02	01	20	07							
	13	10	01	12	01	04	11	35	00	05	00	00	01	01	01	00	00	18	03							
	11	14	01	05	00	01	03	13	01	14	00	00	01	00	01	00	00	16	05							
	34	19	02	09	03	05	07	02	01	04	00	00	00	01	01	00	01	33	11							
Researcher unaware	28	25	01	12	10	03	10	14	02	05	00	00	01	00	04	00	03	24	07							
	06	09	01	01	00	04	08	09	00	11	02	00	00	00	00	00	04	12	01							
	14	07	04	04	07	08	09	10	02	02	00	01	00	00	00	00	00	20	09							
	10	04	01	02	03	02	04	02	01	05	00	00	01	00	00	00	00	17	04							
	43	15	02	18	07	08	00	00	02	11	00	01	00	00	00	00	03	45	18							
	21	18	07	12	05	02	13	06	02	10	00	00	01	00	00	00	01	27	05							
Linguist aware	02	05	00	07	19	01	01	07	01	09	00	03	04	00	00	00	00	13	01							
	28	20	03	10	05	04	07	15	01	11	00	00	02	00	01	00	01	49	05							
	40	16	03	11	08	10	06	01	01	00	00	01	01	00	04	00	02	46	17							
	55	18	28	14	06	08	14	12	04	07	00	01	02	00	02	00	02	71	18							
	33	11	03	06	03	01	16	07	01	10	00	00	01	00	03	00	00	41	06							
	28	16	01	05	03	07	12	02	03	04	00	00	00	00	00	00	03	29	12							
Linguist unaware	03	13	00	06	01	00	03	10	00	09	00	00	08	00	01	00	00	12	03							
	18	15	00	11	02	04	18	24	02	16	00	01	01	01	02	00	00	24	04							
	11	10	10	07	06	04	08	02	00	08	00	01	00	00	00	00	01	35	04							
	22	13	00	09	11	08	05	00	01	11	00	00	00	00	01	00	01	35	04							
	16	09	02	01	01	05	02	03	00	01	00	00	01	00	00	00	01	11	06							
	15	09	03	02	01	01	05	07	00	02	00	00	01	00	00	00	00	22	04							

## DISCUSSION

The reliabilities of the overall rankings from the panel of anthropologists were low. This could possibly be attributed to the task; the task of ranking forty-five stories from best to worst may have been too difficult since the capacity to process forty-five items is well beyond immediate memory span. This lack of good ranking reliability also implies that an investigator should always consider the possibility that an individual rater may not be entirely reliable, and consequently it is more appropriate to use several raters for language analysis. Using several raters not only allows assessment of the inter-judge reliability but also affords the possibility of using a pooled rating as a better estimate.

The fact that the inter-linguist reliabilities were all positive and moderately high implies that the components (positive features of grammar and negative features of grammar) were judged with reasonable accuracy, since the rankings derived from the ratio of positive features to negative features. Had any of the components been unreliable, this would have considerably reduced the reliability of the rankings.

The overall rankings, and the individual linguistic features, as judged by the panel of linguists failed to discriminate between any of the conditions, probably because prescriptive grammar is such an

overlearned response for a college student that it simply was not affected by the experimental conditions.

Since the anthropological rankings correlated with the length of the stories, and since the direction of the results of the two analyses was the same, then the most parsimonious interpretation is that the conditions affected the lengths of the stories which in turn affected the ranks.

The fact that the stories were longer when the subjects reported that they were not aware of the title than when they were aware of the title indicates that the subject's post-experimental questionnaire and the debriefing do not necessarily provide a clear picture of what the subject experienced during the session. This conclusion is similar to that made by Orne (1969) in summarizing the validity of the post-experimental inquiry. Orne concluded that many factors mitigate against fully honest communication in the inquiry and further that the subject cannot necessarily verbalize in the inquiry what he may have perceived during the experiment although it affected his performance at the time.

The present study bears more than a passing resemblance to some recent work on evaluation apprehension (Rosenberg, 1969)\*. In both experimental

---

\*This reference was suggested by Dr. Peter Suedfeld

situations the subjects were exposed to a situation which implied that their behavior was being evaluated. The evaluation apprehension theory proposed that "the typical human subject approaches the typical psychological experiment with a preliminary expectation that the psychologist may undertake to evaluate his (the subject's) emotional adequacy, his mental health, or lack of it. Even when the subject is convinced that his adjustment is not being directly studied he is likely to think that the experimenter is nevertheless bound to be sensitive to any behavior that bespeaks poor adjustment or immaturity" (Rosenberg, 1965). In the present study, subjects also expected to be evaluated, and like the expectation that psychologists are concerned with mental health, these subjects expected that the linguist or the anthropologist was concerned with aspects of their verbal behavior, even though considerable care was taken to convince them that a child was the focus of the study.

Rosenberg's studies found that the subjects who felt evaluated presented themselves as being better-adjusted than the subjects who did not feel evaluated. These findings are consistent with this study's original hypothesis that subjects should produce better grammar for a linguist and a better story for an anthropologist.

Contrary to this expectation, the present study

found that the anthropologist condition resulted in shorter, consequently worse, stories. Similar results were found by Rosenberg (1969) in a study involving a high degree of evaluation apprehension. A post-experimental index of the anxiety aroused by the high evaluation apprehension strongly correlated with the degree to which the subjects reacted. Rosenberg interpreted these findings to mean that the subjects utilized a defense mechanism to cope with the anxiety of being evaluated. This evaluation apprehension anxiety, when strongly experienced, leads to reactive anger toward the experimenter. The anger becomes manifest in a desire to disconfirm what the subject perceives as the experimenter's expectations. In the current study, one case illustrates this effect well. One subject, in the post-experimental questionnaire, indicated her anger by responding to the question, "What type of experiment do you think this is?" by writing, "One that dupes the hell out of you". Her anger was probably not an isolated instance in the present study.

The fact that stories were longer when subjects reported being aware of the titles could best be explained by Rosenberg's findings as well. From his post-experimental interviews he proposes that subjects exhibit a "motivated inattention" (Rosenberg, 1969) in an attempt to obscure from themselves the extent to which they regulated their responding. Since it is

unlikely that subjects, who were informed of the title three times during the session, really missed hearing the title altogether and since the awareness condition where they reported not being aware of the titles yielded significant title differences, it is not unrealistic to conclude that at some level they did hear the title, were affected by it, and then after the session failed to recall it. If they had been aware of what anthropologists and linguists look for and knew they were influenced by this, motivated inattention may explain their inability to recall the title.

There was a large amount of variability between subjects on the linguistic grammatical features. It was not uncommon to find a range from zero use of a particular feature to 50 or more instances of it across subjects (see Table 12). This variability may account for the finding that there were no significant differences of title or awareness for any of the grammatical features. Further, this may imply that language studies concerned with finding correlations between situational variables and a particular grammatical form or shift in style should be conducted by observing grammatical shifts within the subject's corpus rather than comparing one subject's corpus to another's.

Moreover, the fact that the stories were shortened when subjects felt their stories were being evaluated is consistent with Labov's (1970) findings that some

subjects are reluctant to speak in the presence of an experimenter authority figure. The implications of these results for anthropological linguistics are that fieldworkers who use their professional title may in fact be causing their informants to withhold the very information they are trying to collect. Consequently, more unobtrusive methods should be considered for the collection of anthropological data.

Future studies might eliminate the problem of inter-judge reliability by simply eliminating judges entirely in favor of analyzing the corpus with a computerized syntactic analysis. This would entail using search procedures from a computer language such as SNOBOL to scan the data for instances of a certain linguistic feature (e.g. to locate any occurrences of but) and to print out a frequency count of each. This would be easy to do with measures like number of coordinating conjunctions, but more difficult with measures like words and expressions misused.

### Summary

There is a demand characteristic of professional's title which causes anthropologists to judge stories produced by anthropological subjects as inferior to those produced by other subjects and which causes stories told to anthropologists to be short. The greater the awareness of the title, the more the demand characteristic operates to shorten these stories.

## Appendix A

## Script for Experimenter Explaining Task to Subject

Hi, M \_\_\_\_\_? Would you come across the hall, please? I'm Nancy Kalish-Landon, a graduate student helping to run the study. Over on the wall pointing you'll see an intercom. The intercom goes to the room next-door where the linguist anthropologist researcher, Dr. Koslin, and child are. The reason we're using an intercom instead of a face-to-face interaction is so that the anthropologist's linguist's researcher's or child's appearance doesn't influence your story, or so your appearance doesn't influence them.

Now what I'd like you to do is tell the story of Cinderella to Dr. Koslin and the child. You'll have a few minutes to think about what you want to say.

This is a one-way intercom. Right now you're listening to them. It's not until you reverse the direction of the sound by pressing this button that they can hear you. Make sure whenever you're talking that you hold the button down.

Now take a few minutes to think about what you want to say; then whenever you're ready just press the button and begin and the linguist anthropologist researcher and child will listen to your story. When you're through, just come out; I have a questionnaire for you to fill out, then I'll pay you and you can leave, OK?

[after completion of task and questionnaire]

It's very important that you don't discuss this experiment with anyone attending Rutgers because we may need their help, too, and we want them to participate on the basis of the same knowledge about the study that you had when you arrived. When the experiment is completed, I'll explain the results to your class. Well, thanks again. Bye.

## Appendix B

## Questionnaire Completed by All Subjects After Task

Name \_\_\_\_\_ Other languages you speak  
fluently \_\_\_\_\_  
 Native Language \_\_\_\_\_

Would you please answer the following questions.  
 In this experiment, there was no way you could  
 have cheated if you "saw through" the experiment,  
 so if you behaved accordingly, please tell me.

- order
- was
- rando-  
mized
1. From your previous knowledge, what is lin-  
guistics and what does a linguist do?
  2. From your previous knowledge, what is  
anthropology and what does an anthropolo-  
gist do?
  3. What type of experiment do you think this  
is? What do you think a "researcher" does?
  4. Do you think you told your story with the  
fact that you were being observed in mind?  
How do you think this knowledge might have  
changed your story?
  5. Whether or not your story was told in a  
manner conscious of the experimenter, what  
do you think other subjects have done in  
this experiment?
  6. When was the last time you told this  
nursery story to a child?

7. What is the experimenter's - Dr. Koslin's -  
field?

Comments or Questions:

## Appendix ⑥

## Coding Sheet for Panel of Anthropologists

	Story Number _____
1. No. of <u>characters</u> mentioned in the story	_____
2. No. of characters <u>physically described</u>	_____
3. No. of characters described by <u>clothing or possessions</u>	_____
4. No. of characters assigned a <u>kinship</u>	_____
5. No. of times <u>emotions</u> are ascribed to	
Character A (mentioned first in story)	_____
Character B (mentioned second)	_____
Character C	_____
Character D	_____
Others	_____
6. No. of <u>locations</u> mentioned in the story (e.g. house, town, Room A of house, Room B of house)	_____
7. No. of locations <u>physically described</u> (e.g. objects in room, size of house, color of trees in forest)	_____
8. No. of <u>actions</u> performed by	
Character A (same character as question 5)	_____
Character B	_____
Character C	_____
Character D	_____
Others	_____
9. Of these total actions, how many were <u>magical</u> ? (e.g. animal talking, person flying)	_____
10. How many times does the story mention a <u>moral</u> , a lesson to be learned, an ethical code, etc?	_____
11. Total no. of words in the story (count <u>everything</u> , including false starts, <u>uh</u> , <u>er</u> , etc)	_____

Now assign this story an overall rating, based on the features above, indicative of the story's overall level of interest, involvement, complexity, and the like. These personal ratings will be used by you to compare stories and rank them 1 (best) through 45 (worst).

## Appendix B

## Coding Sheet for Panel of Linguists

Story Number \_\_\_\_\_

Note to linguists: This coding sheet is to be used in conjunction with the Xeroxed pages supplied to you from Understanding and Using English by Birk and Birk. You are also advised to consult the same dictionary, Webster's New World Dictionary of the American Language.

Formality Level	Feature Desirability
1. No. of <u>long sentences</u> (10 words or longer, not run-ons)	_____ +
2. No. of words <u>3 syllables</u> or longer	_____ +
3. No. of <u>short sentences</u> (6 words or shorter)	_____ +
4. No. of <u>contractions</u>	_____ -
5. No. of <u>slang and colloquial</u> words (consult <u>New World</u> dictionary)	_____ -
6. No. of words and expressions <u>misused</u>	_____ -
Delivery	
1. No. of <u>false starts</u>	_____ -
2. No. of <u>hesitations</u> ( <u>uh</u> , <u>er</u> , etc.)	_____ -
3. No. of <u>sentence fragments</u>	_____ -
4. No. of <u>sentences within a sprawling and-and structure</u> (refer to Birk and Birk, p. 463)	_____ -
Grammar	
1. No. of <u>incorrect plurals</u> of nouns	_____ -
2. No. of <u>agreement mistakes</u> between <u>pronouns</u> and <u>antecedents</u> (refer to Birk and Birk, pp. 411-413)	_____ -

	Feature Desirability
3. No. of verb <u>tense errors</u> (shifts from one to another inappropriately)	_____ -
4. No. of <u>agreement mistakes</u> between <u>verbs</u> and <u>subjects</u> (refer to Birk and Birk, pp. 425-427)	_____ -
 Style	
1. No. of sentences in <u>imperative</u> mood	_____ +
2. No. of sentences in <u>subjunctive</u> mood	_____ +
3. No. of sentences in <u>passive</u> voice	_____ +
4. No. of <u>coordinating conjunctions</u> (refer to Birk and Birk, p. 438)	_____ +
5. No. of <u>subordinating conjunctions</u> (Birk and Birk, p. 438)	_____ +
6. No. of <u>conjunctive (transitional) adverbs</u> (Birk and Birk, pp. 438-439)	_____ +

Now assign this story an overall rating by weighing its good points (high numbers within positive features, low # within negative features) with its bad points (low # within positive features, high # within negative features). These personal ratings will be used by you to compare stories and rank them 1 (best) through 45 (worst).

## REFERENCES

- Anglin, J. M. The growth of word meaning. Cambridge: The M.I.T. Press, 1970.
- Bailey, B. Jamaican creole syntax: a transformational approach. Cambridge: University Press, 1966.
- Bever, T. G. The cognitive basis for linguistic structures. In J. R. Hayes (ed.) Cognition and the development of language. New York: John Wiley & Sons, Inc., 1970.
- Birk, N. P. and Birk, G. B. Understanding and using English. New York: The Odyssey Press, 1965.
- Bloom, L. Language development: form and function in emerging grammars. Cambridge: The M.I.T. Press, 1970.
- Bloom, L. 1971 (personal communication)
- Braine, M. D. S. The acquisition of language in infant and child. In C. Reed (ed.) The learning of language: essays in honor of David H. Russell. New York: Appleton-Century-Crofts, 1970.
- Brown, R. W. The development of Wh questions in child speech. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 279-290.
- Brown, R. and Fraser, C. The acquisition of syntax. In C. N. Cofer, and J. K. Musgrave, (eds.) Verbal behavior and learning. New York: McGraw-Hill, 1963.
- Brown, R. and Hanlon, C. Derivational complexity and order of acquisition in child speech. In J. R. Hayes (ed.) Cognition and the development of language. New York: John Wiley & Sons, Inc., 1970.
- Bruce, D. J. Effects of context upon intelligibility of heard speech. In R. C. Oldfield and J. C. Marshall (eds.) Language. London: Penguin Books, 1968.
- Carey, P., de Boysson-Bardies, B., Ioup, G., and Ahammer, I. On the scope of negation. Unpublished paper (City University of New York), 1971.

- Cataldo, J. F., Silverman, I., and Brown, J. M. Demand characteristics associated with semantic differential ratings of nouns and verbs. Educational and Psychological Measurement, 1967, 27, 83-87.
- Chomsky, C. The acquisition of syntax in children from 5 to 10. Cambridge: The M.I.T. Press, 1969.
- Chomsky, N. Aspects of the theory of syntax. Cambridge: The M.I.T. Press, 1965.
- Chomsky, N. The logical basis of linguistic theory. In H. Lunt (ed.) Proceedings of the Ninth International Congress of Linguists. The Hague: Mouton, 1964.
- Dakin, J. Unpublished research cited in J. Lyons (ed.) New horizons in linguistics. Baltimore: Penguin Books, 1970.
- Dingwall, W. O. A survey of linguistic science. College Park: Linguistics Dept./University of Maryland, 1971.
- Ervin-Tripp, S. Discourse agreement: how children answer questions. In J. R. Hayes (ed.) Cognition and the development of language. New York: John Wiley & Sons, Inc., 1970.
- Greenbaum, S. Verb-intensifier collocations in English: an experimental approach. The Hague: Mouton, 1970.
- Greenfield, P. M. Who is 'DADA'? Unpublished paper, 1967, cited in D. McNeill The acquisition of language. New York: Harper & Row, 1970.
- Gumperz, J. J. Language. In B. J. Siegel (ed.) Biennial review of anthropology. Stanford: Stanford University Press, 1965.
- Gumperz, J. J. On the linguistic markers of bilingual communication. Journal of Social Issues, 1967, 23, 48-57.
- Harris, S. Influence of subject and experimenter sex in psychological research. Journal of Consulting and Clinical Psychology, 1971, 37, 291-294.
- Helson, H., Blake, R. R., Mouton, J. S., and Olmstead, J. A. The expression of attitudes as adjustments to stimulus, background and residual factors. Journal of Abnormal Social Psychology, 1956, 52, 314-322.

- Hill, A. Grammaticality. Word, 1961, 17, 1-10.
- Hymes, D. Language in culture and society: a reader in linguistics and anthropology. New York: Harper & Row, 1964.
- Jackendoff, R. S. Quantifiers in English. Foundations of Language, 1968, 4, 422-442.
- Jackendoff, R., and Culicover, P. A reconsideration of dative movements. Foundations of Language, 1971, 7, 397-412.
- Klima, E. and Bellugi, U. Syntactic regularities in the speech of children. In J. Lyons and R. J. Wales (eds.) Psycholinguistics papers. Edinburgh: Edinburgh University Press, 1966.
- Labov, W. The social stratification of English in New York City. Washington, D.C.: Center for Applied Linguistics, 1966.
- Labov, W. The logic of nonstandard English. In J. E. Alatis (ed.) Report of the twentieth annual round table meeting on linguistics and language studies. Washington, D.C.: Georgetown University Press, 1970.
- Labov, W. Methodology. In W. O. Dingwall (ed.) A survey of linguistic science. College Park: Linguistics Program/University of Maryland, 1971.
- Langendoen, D. T., Kalish-Landon, N., and Dore, J. Dative questions. Paper presented at Northeastern Linguistic Society (NELS), Amherst, Mass., October 22, 1972.
- Leopold, W. F. Speech development of a bilingual child. Vol. 1: Vocabulary growth in the first two years. Vol. 2: Sound-learning in the first two years. Chicago: Northwestern University Press, 1939, 1947.
- Li, J. C. R. Statistical inference I. Ann Arbor: Edward Brothers, Inc, 1966.
- Livant, W. P. A comparison of noun and verb forms on the semantic differential. Journal of Verbal Learning and Verbal Behavior, 1963, 1, 357-360.
- MacKay, D. G. To end ambiguous sentences. Perception and Psychophysics, 1966, 1, 426-436.

- McLendon, S. 1971, (personal communication).
- McNeill, D. The acquisition of language. New York: Harper & Row, 1970.
- Mead, M. The primitive child. In C. Murchism (ed.) A handbook of child psychology. Worcester: Clark University Press, 1931.
- Orne, M. T. Demand characteristics and the concept of quasi-controls. In R. Rosenthal and R. L. Rosnow (eds.) Artifact in behavioral research. New York: Academic Press, 1969.
- Postal, P. M. Coreferential complement subject deletion. Linguistic Inquiry, 1970, 1, 439-500.
- Reisman, D. Comment on the state of communication research. Public Opinion Quarterly, 1956, 23, 10-13.
- Reisman, D. and Ehrlich, J. Age and authority in the interview. Public Opinion Quarterly, 1961. 25, 39-56.
- Rosenberg, M. J. When dissonance fails: on eliminating evaluation apprehension from attitude measurement. Journal of Personality and Social Psychology, 1965, 1, 18-42.
- Rosenberg, M. J. The conditions and consequences of evaluation apprehension. In R. Rosenthal and R. L. Rosnow (eds.) Artifact in behavioral research. New York: Academic Press, 1969.
- Rosenthal, R. Experimenter effects in behavioral research. New York: Appleton-Century-Crofts, 1966.
- Russell, B. Bertrand Russell speaks his mind. New York: World Publishing Co., 1960.
- Schapiro, B. The underlying structure of indirect objects. Paper presented at Northeastern Linguistic Society (NELS), Amherst, Mass., October 22, 1972.
- Slobin, D. I. A field manual for cross-cultural study of the acquisition of communicative competence (second draft). Berkeley: University of California Press, 1967.

- Smith, C. S. An experimental approach to children's linguistic competence. In J. R. Hayes (ed.) Cognition and the development of language. New York: John Wiley & Sons, Inc., 1970.
- Spencer, N. J. Differences between linguists and non-linguists in intuitions of grammaticality-acceptability. Paper presented at Annual Meeting of the Linguistic Society of America, Atlanta, Ga., December 27, 1972.
- Stricker, L. J. The true deceiver. Psychological Bulletin, 1967, 68, 13-20.
- Stricker, L. J., Messick, S., and Jackson, D. N. Suspicion of deception: implications for conformity research. Journal of Personality and Social Psychology, 1967, 5, 379-389.
- Svartvik, J. On voice in the English verb. The Hague: Mouton, 1966.
- Ulatowska, H. K. Psycholinguistic approach to the study of ambiguity. Papers in Linguistics, 1971, 4, 339-378.
- Vaughan, G. M. The effect of the ethnic grouping of the experimenter upon children's responses to tests of an ethnic nature. British Journal of Social and Clinical Psychology, 1963, 2, 66-70.
- Womack, W. M. Negro interviewers and white patients. Archives of General Psychiatry, 1967, 16, 685-692.