THE RELATION BETWEEN SENTENCE ORDER AND THE COMPREHENSION OF WRITTEN ENGLISH

Thesis for the Degree of M. A.

MICHIGAN STATE UNIVERSITY

Donald Keith Darnell

1960

THE RELATION BETWEEN SENTENCE ORDER

AND

THE COMPREHENSION OF WRITTEN ENGLISH

By

Donald Keith Darnell

AN ABSTRACT

Submitted to the College of Communication Arts
Michigan State University in partial
fulfillment of the requirements
for the degree of

MASTER OF ARTS

Department of General Communication Arts

1960

Approved David K. Berlo

Donald Keith Darnell

The purpose of this study was to investigate the relationship between the order of sentences in a message and the comprehension of that message by the intended receiver.

The independent variable in this study was sentence order. The two dependent variables were (1) comprehension of the message as measured by close procedure and (2) relative redundancy of the message as estimated from the responses of subjects to blanks in a mutilated form of the message.

The general hypothesis of the study was: Successive steps of removal from the "because" order of sentences will reduce the accuracy of respondents' predictions about the missing parts of the message.

A fifteen sentence message was written and arranged in a "because" order. Sentences were then transposed in the message creating six alternate forms. The six alternate treatments range from seven to forty-three transpositions from the original order. The seven treatments were prepared for close procedure and presented, along with a control message, to 140 subjects (20 per treatment).

The mean close proportion and the average relative entropy were computed for each treatment. An analysis of variance was used to test for differences among mean close and entropy scores for the seven treatments. Rank order correlation coefficients were also computed to determine the relationship between the number of transpositions from the "because" order, the mean close scores,

and the average relative redundancy for the seven treatments. In addition, judgments of difficulty, interest, and utility were obtained from each subject, and a Pearson r computed to determine the relation between an individual's close score and each of his evaluations of the message.

The conclusions of the study were:

1. There was a significant relation between close scores and the transpositions from a "because" order.

2. There were significant differences among relative redundancy scores for the seven treatments, but the rank order correlation between relative redundancy and transpositions from a "because" order was not significant.

3. No evidence was obtained of a relation between close scores and judgments of difficulty, interest, or utility.

An alternate method was suggested for obtaining approximations to a given order of a message, and a correction was made in the procedure for estimating the proportion of relative entropy from the responses of a group of subjects.

THE RELATION BETWEEN SENTENCE ORDER

AND

THE COMPREHENSION OF WRITTEN ENGLISH

By

Donald Keith Darnell

A THESIS

Submitted to the College of Communication Arts
Michigan State University in partial
fulfillment of the requirements
for the degree of

MASTER OF ARTS

Department of General Communication Arts

1960

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I wish to express my appreciation to my adviser, Dr. David K. Berlo, who assisted me throughout this program of research; to Dr. Paul Deutschmann, Dr. David Potter, Dr. Bernard Engle, and Dr. John Street, who helped in the initial planning stages; and to Dr. Malcolm MacLean Jr. and Dr. Erwin Bettinghaus who criticized the final report.

I must also thank the several professors who contributed class time and subjects for this research.

The highest honor is reserved for the family, friends, and colleagues who listened patiently.

PREFACE

The communication process often involves two people, and always involves two roles, the source and the destination. A different classification can also be made of the communication actors, the initiator and the respondent. The initiator and the respondent may be discriminated by the fact that a willingness to expend the necessary effort to accomplish a specific purpose may be assumed on the part of the initiator but not for the intended respondent. That is, when one of the actors initiates a communication it may be assumed that the ratio of expected reward to the expected effort is equal to, or exceeds, unity.

The initiator may assume that the probability of the desired response is approximately equal to the perceived reward/effort ratio for the respondent; (for example, if he is certain that the intended respondent knows what response is desired). He may, then, wish to increase the probability of the desired response by decreasing the effort that will be required of the respondent in giving that response, if he can do so without reducing critically his own reward/effort ratio.

Communication research is directed toward those variables which affect the reward/effort ratio, and one of its goals is to minimize the total effort required of the actors in a given communication chain. Even in the simplest form of the communication

chain, however, where the initiator performs the source, encoder, and transmitter functions, and the respondent performs the receiver, decoder, and destination functions, it is not easy to select the approach that will require the minimum total effort.

A number of different approaches might be taken. For example, the initiator could select a very redundant code--one that uses a large number of symbols in relation to the amount of information, knowledge, to be transmitted--intending to minimize the load on the encoder and decoder functions. This approach, however, tends to increase the burden on the transmitter and receiver functions and may result in no significant change in the total effort expended. The opposite approach could also be adopted, to minimize the ratio of symbols to content, but again the result is a shifting of effort from one function to another rather than a reduction in the total effort of the system. A third approach (the one most often adopted) is to minimize the effort required of the initiator or the respondent with the consequence that an excessive burden is placed on the other. As has been suggested, the solution to the problem is not an easy one.

The experimental approach adopted in this study is (1) to hold the number of symbols constant, (2) to vary the order of the symbols, producing seven treatments of the message, and (3) to observe the behaviors of respondents to see if there are differential effects among the treatments. If significant differences are observed, and if it is assumed that the same amount of effort is expended by the initiator and respondent on each of the seven treatments, then, that treatment which shows the greatest effect

(e.g., yields the highest mean comprehension score) can be said to require the least total effort in the transmission of a constant amount of information.

LIST OF TABLES

Table | Page

LIST OF ILLUSTRATIONS

INTRODUCTION

The purpose of this study is to investigate the relationship between the order of sentences in a message and the comprehension of that message by the intended receiver. The assertion has been made repeatedly by authors of textbooks and teachers of English composition that it is important to attend to order in creating a written or spoken message. That is, if an individual transmits a message in whatever order the units occur to him, the desired effect may not be obtained, while another order of the same units might achieve the desired effect. The present research is an attempt to obtain some empirical support for a specific assertion which is deducible from the more general one stated above. This specific assertion is that sentence order is a significant factor in the comprehension of English messages.

The significance of the study lies not in proving the experts right or wrong but in the possibility of increasing the efficiency of communication. There are several reasons for assuming that order makes a difference in the meaning elicited by a set of symbols (compare dab with bad, back out with out back, or 12 with 21), but there is no reason to assume that one arrangement is necessarily more difficult to produce than any other.

The method of this study is to compare mean comprehension scores for groups of subjects on treatments of a message which

differ only in the order that the sentences occur. The sentence is chosen as the unit to be manipulated because: (1) the effect of order on comprehension can be clearly demonstrated at the lower levels, (2) it has often been asserted that there is a "best" order for the larger units, and (3) at least one investigation at a higher level (paragraphs) failed to yield a significant result.

It is assumed that an effect of sentence order can only be predicted if dependency can be demonstrated between the sentences, and if the destination is cognizant of the dependencies. With that assumption, the following definitions are critical to this study:

1) Given a set of units, (a), (b), and (c); given that unit (a) occurs, if one can predict the occurrence of (b) or (c) with greater than chance success some degree of dependency obtains between the units (a), (b), and (c).

2) Given a message, to the extent that the order of the units is determined by a dependent relationship between units there is <u>structure</u> in the message.

3) Given a structured message, to the extent that the destination has accurate expectations about the ordering of the units within that message, the destination is <u>organized</u>.

The definitions of other concepts pertinent to this study must be withheld momentarily until they can be presented in theoretical perspective; however, the purpose of the present research can now be restated. It is to investigate the relation between structural variation and organization.

This study is to be presented in four parts. Chapter I includes a theoretical background for the research, a description of the measuring instrument, the general hypothesis, and a rationale.

Chapter II includes the design of the experiment and a description of the technique of administration. Chapter III reports results of a statistical analysis of the data. Chapter IV includes a short summary of the experiment, conclusions, a discussion of the results, some observations about Information Theory, and finally, suggestions for further research.

CHAPTER I

This chapter includes a theoretical background for the research, a description of the measuring instrument, the general hypothesis, and a rationale for the study.

INFORMATION, STRUCTURE, AND ORGANIZATION

A major part of the reasoning behind the present study is based on Information Theory. It is therefore important to introduce at this early point some of the concepts of that theory and to define them linguistically and mathematically. One of the central concepts of information theory is information or entropy. These terms are commonly used with divergent meanings, information referring to "knowledge," and entropy, in the theory of thermodynamics, being associated with the probability of a given distribution of momentum among molecules. In information theory, however, the two terms are equivalent, referring to the uncertainty or lack of knowledge in a communication system. Warren Weaver expresses the general meaning symbolized by the terms in this way.

> Entropy associated with a situation is a measure of the degree of randomness, or "shuffledness" if you will, in the situation.... Thus for a communication source one can say. . . "this situation is highly organized, it is not characterized by a large degree of randomness or of choice -- that is to say, the information (or entropy) is low."[1]

---

[1] Claude E. Shannon and Warren Weaver, The Mathematical Theory of Communication (Urbana: The University of Illinois Press, 1949), p. 103.

From this explanation, it would seem that a negative relation obtains between entropy and organization. Since organization is a key concept in this study, a more exact definition of entropy is required to establish this relationship.

Given an information source which is producing a message by successively selecting discrete symbols from a set of symbols, if the symbols are independent and equally probable that source has maximum entropy. The value of maximum entropy for a set of n symbols may be expressed

$$H_{max} = n\left(-\frac{1}{n} \log_2 \frac{1}{n}\right).$$

Thus, if n equals four, H max = 4 $(-\frac{1}{4} \log_2 \frac{1}{4})$ = 4 (.5) = 2. That is to say, a situation in which the source must decide between four equally probable alternatives contains two "bits" of information. The maximum number of "bits of information" or "units of entropy" in a system is equal to the number of times the number of alternatives must be divided by two to obtain a quotient of one.

Given an information source which is producing a message by successively selecting discrete symbols from a set of symbols, if the probability of choice of the various symbols, at any stage in the process, is dependent on the previous choices the source has less than maximum entropy. In the situation where the symbols are dependent or unequally probable, the expression for the absolute (observed) entropy is

$$H = -[p_1 \log_2 p_1 + p_2 \log_2 p_2 + \cdots + p_n \log_2 p_n].$$

Thus, if n equals four as in the previous example, and p_1, p_2, p_3, and p_4 equal .50, .25, .15, and .10 respectively, the absolute entropy

of this system equals 1.7427 "bits." The relative entropy of this system is the ratio of the actual entropy to the maximum entropy or .8713 (87%), and one minus the relative entropy is the relative redundancy of the system. A formula for expressing relative redundancy follows;[2]

$$1 - \left[\frac{-\Sigma p_i \log_2 p_i}{n-\left(\frac{1}{n} \log_2 \frac{1}{n}\right)}\right].$$

If, in fact, entropy is negatively related to organization, then a positive relation should obtain between organization and redundancy. At various points throughout the remainder of this thesis it will be convenient to refer to either entropy or redundancy, but in each case, the obverse is implied.

The concept of relative entropy assumes a determiner or determiners. Without predetermination the source would always have complete freedom of choice -- maximum entropy would obtain. For the human source the determiners are the rules which the source adopts for the construction of a particular message. These rules may be of three types: (1) those which are commonly accepted in a language community (e.g., frequency of previous usage), (2) those which are based on specific agreement between the source and destination, and (3) those which the source selects arbitrarily.

---

[2] Ibid., pp. 103-106. Tables are available for $(-p \log_2 p)$ in E. B. Newman, "Computational Methods Useful in Analyzing a Series of Binary Data," American Journal of Psychology, LXIV (April, 1951), 252-262.

The significance of the distinction is that for each of the three types there is a different probability that the source and the destination will agree on the set. The existence of the third possibility argues that entropy is a characteristic of the source, and that a consideration of relative entropy as a characteristic of a message or a language involves the assumption of one of the first two types of rules.

Suppose a given source composes a message and transmits part of it to a receiver who is expected to complete the message (i.e., play the source role). The relative entropy of the two sources in composing the withheld part of the message, and consequently the two versions of it, would be expected to differ to the extent that the rules they have adopted differ. Therefore, to the extent that the two versions of the withheld portion of the message are similar it can be assumed that the rules adopted by the two sources are similar. And, if the two versions of the message are identical, or very similar, it would seem to follow that the situation is highly redundant for both sources, or that the withheld part of the message contains little "information."

Now it is possible to contrast and compare the three basic concepts of this study. Given a set of symbols among which dependent relationships obtain, and given that a message is constructed from the set; if the order of the symbols in the message is determined by the dependent relationships there is structure in the message; if the order is determined by the rules which the source has adopted there is redundancy in the source; and if a receiver is able to

predict the symbols and the order of the symbols with greater than chance success there is organization in the destination.

If, for a given message, the rules which the source has adopted are based on a knowledge of dependent relationships in the symbol set, the proportion of relative redundancy in the source is equal to the degree of structure in the message. If, for a given message, the source and the receiver adopt the same rules, relative redundancy and the degree of organization are equal. If, for a given message, both of these conditions are met, relative redundancy, structure, and organization have the same value.

CLOZE PROCEDURE

Cloze procedure is the measuring instrument chosen for use in this study. It is defined by its originator as "a psychological tool for gauging the degree of total correspondence between (1) the encoding habits of transmitters and (2) the decoding habits of receivers."[3] The procedure is quite simple. One takes a written message and deletes a portion of the words. The deletion is performed in such a way as to be independent of the words themselves; i.e., from a random start every n^{th} word is deleted. This is done because "it seems necessary to let the occurrences of all sorts of words be represented according to the proportion of their occurrence."[4]

---

[3]Wilson Lewis Taylor, "Application of 'Cloze' and Entropy Measures to the Study of Contextual Constraint in Samples of Continuous Prose " (unpublished Ph.D. dissertation, The University of Illinois, 1954), p. 3.

[4]Ibid., p. 4.

Blanks of uniform size replace the deleted words, and the mutilated message is presented to a group of subjects who attempt to replace the missing words. The number of deleted words which a subject correctly replaces is his "close score," which may be represented as a proportion for comparative purposes.

> The "close score" for a given subject and a given passage is taken to be a measure of the degree of correspondence between the language habits used by the source while "encoding" the message (fitting sequences of language symbols to the meaning) and the language habits used by the receiver while "decoding" it (fitting meaning to the mutilated message) and, on the basis of meaning perceived, attempting to encode those elements that will make the message's form whole again.[5]

In a later publication Taylor anticipates some of the questions that might arise and gives these answers:

> Findings up to now indicate that the easiest ways of applying close procedure may be best for most uses. There seems to be little advantage in preclassifying words and limiting deletion to them, and no advantage to putting oneself to the trouble of judging and scoring synonyms. Also, it appears that an every-fifth-word deletion system spaces blanks as far apart as they need to be. Further, a series of about 50 blanks is roughly sufficient to allow the chances of mechanically selecting easy or hard words to cancel out and yield a stable score of the difficulty of a passage, or the performance of an individual, despite what specific words the counting-out process may delete.[6]

Some information about what close measures as well as information about its reliability is provided by the following correlations. In a before-after learning experiment, close scores

---

[5] Ibid., pp. 2-3.

[6] Wilson L. Taylor, "Recent Developments in the Use of Close Procedure," Journalism Quarterly, XXXIII (Winter, 1956), 48.

were obtained along with comprehension scores from independently validated multiple choice tests and AFQT intelligence scores. Close correlated with the "before" comprehension test .70 and with the "after" comprehension test .80. Close correlated with itself before and after .88. Comprehension scores correlated with AFQT .65 and .70, while close and AFQT correlated .73 and .74. All these r's are significant at the .01 level and all are large and positive.[7]

Close procedure was originally intended as a measure of "readability." Previous research in that area had shown that messages could generally be ranked according to difficulty by considering a very few variables. The Dale-Chall formula, for example, is based on average sentence length and percentage of unfamiliar words.[8] The Flesch formula considers the number of words per sentence and the number of syllables per hundred words.[9] It was evident, however, that these formulas did not take into account all the variables that affect readability. "One may think of close procedure," says Taylor, "as throwing all potential readability influences in a pot, letting them interact, then sampling the result."[10]

---

[7] Ibid., 45.

[8] Edgar Dale and Jeanne S. Chall, "A Formula for Predicting Readability: Instructions," Educational Research Bulletin, XXVII (February, 1948), 37.

[9] Rudolph Flesch, "A New Readability Yardstick," Journal of Applied Psychology, XXXII (June, 1948), 223.

[10] Wilson L. Taylor, "'Close Procedure': A New Tool for Measuring Readability," Journalism Quarterly, XXX (Fall, 1953), 417.

Cloze tests correlate highly with findings of the Dale-Chall and Flesch formulas on standard materials, but when the mechanical formulas are applied to non-standard materials cloze scores seem to be better measures of real difficulty.[11] Cloze scores, then, may be an index of learning, comprehension, intelligence, or message difficulty, and it is a matter of control that determines what they index in a particular situation.

The practical significance of cloze procedure for this study is that cloze procedure is a method of obtaining predictions from an individual. It is assumed that an individual's predictions are an index of his expectations. The proportion of his predictions that are "correct" is, then, an estimate of his organization, and the mean cloze score for a group of subjects is an estimate of the average organization those subjects have in relation to a given message. The assertion that cloze scores are an index of organization in addition to being an index of learning, comprehension, intelligence, or message difficulty, should not be surprising, since all of the latter concepts are certainly related to the process of forming correct expectations (organization).

By observing the predictions of groups of subjects, some information can be obtained about the choices that a source has under specific conditions and their relative probabilities. From these an estimate of the relative redundancy or entropy of an "average" source can be calculated. To obtain this estimate one treats each

---

[11] Ibid.

blank of the mutilated message as a discrete system and the response of each subject as a possible outcome of that system. By computing the probability of each word's occurrence, it is possible to approximate empirically the degree to which the blank is determined by its context.[12] Maximum entropy would then be indicated if all subjects give different responses to a blank, and zero entropy if all subjects give the same response. Given the responses of a group of subjects it is possible to compute maximum entropy, relative entropy, and the relative redundancy of a particular blank (see pages 5 and 6). The average entropy or redundancy for the message is simply the arithmetic mean of those scores for all blanks.

Taylor offers still another measure that may be obtained in the same computations, that of "misdirection." The cloze proportion (proportion of right answers) is removed from the set of alternatives and an entropy value computed from the remaining responses. This absolute value is then divided by the maximum entropy value for the whole system to obtain remaining relative entropy. Remaining relative entropy plus the cloze proportion subtracted from one gives the misdirection figure (i.e., CP+RRH+M= 1).[13] Table 1 shows a sample calculation of entropy, redundancy, cloze proportion, and misdirection. This table assumes twenty respondents for a single blank, of whom ten give the right answer, five give a single wrong answer, and the remaining five give different wrong answers.

---

[12] Taylor, "Application of 'Cloze' and Entropy Measures. . .," p. 15.

[13] Ibid., 59-61.

TABLE 1.--Sample Calculations of Entropy, Redundancy, and Misdirection

Responses	Frequency	Proportion	$-p \log_2 p$
a	1	.05	0.2161
b	1	.05	0.2161
c	1	.05	0.2161
d	1	.05	0.2161
e	1	.05	0.2161
f	5	.25	0.5000
g (c P)	10	.50	0.5000
	20	1.00	2.0805

Maximum entropy	=	20 (0.2161)	= 4.3220
Relative entropy	=	2.0805 / 4.3220	= .46
Rel. redundancy	=	1.00 - .46	= .54
Cloze proportion			= .50
Rem. rel. ent.	=	1.5805 / 4.3220	= .36
Misdirection	=	1.00 - (.50 + .36)	= .14

A TREATMENT OF STRUCTURE

Because cloze procedure requires a message approximately 250 words in length, and because the Flesch formula indicates that a sixteen word sentence is fairly difficult, it was decided to create a fifteen sentence message. Since it would be impossible to vary structure through all possible orders of fifteen sentences (there are 15! possible orders) it was necessary to select a specific kind of message structure.

The "because" order seemed to provide the necessary conditions for a structured message since, as Mills says, "There is one essential relationship among ideas: each subpoint must assist in proving the point to which it is immediately subordinated. . ."[14] In the "because" order the test of subordination is the insertion of "for" or "because" between a topic point and its subpoints.[15] Coordinate subpoints under the same heading can be connected by "and" and are logically interchangable.

Admittedly, a "because" order outline is not the way to structure a message, nor is it applicable to all kinds of materials, but it is one way on which many authorities agree, and is therefore selected for testing the hypothesis.

---

[14] Glen E. Mills, Composing the Speech (New York: Prentice Hall, Inc., 1952), p. 207.

[15] Ibid.; and Donald C. Bryant and Karl R. Wallace, Fundamentals of Public Speaking (New York: Appleton-Century Company, Inc., 1947), p. 229.

To simplify the process of creating a message a fifteen unit outline in skeleton form was set up and sentences which met all the requirements for logical subordination and coordination were written to fit. Two independent judges verified the writer's judgment. An outline in which each main point has the same number of subpoints was used to equalize contextual differences and to simplify the experimental manipulations of structure. The resulting message, in outline form, can be found in Appendix A. The dependency between sentences is clear; given any one of the major heads, only two of the fourteen other sentences may logically follow; given one of those, there are two coordinate points which may follow.

HYPOTHESIS

The general hypothesis of this study is:

Successive steps of removal from a "because" order will reduce the accuracy of respondents' predictions about the missing parts of the message.

RATIONALE

The general hypothesis stated above is a conclusion based on a series of premises. These are:

1) Structure in a system increases the internal predictability of the elements in that system.

2) The "because" order is a kind of structure.

3) The "because" order is a familiar pattern to most users of English.

4) Language users are capable of responding to variations in structural predictability.

The first of these premises is assumed as a matter of definition. Given that the "because" order imposes dependencies among units within the message, it is structured. The "because" order is taught in many schools and is recommended by textbooks as a good way to "organize" messages, so it seems reasonable to assume that most university students have had experience with the "because" order. The fourth is supported by empirical data.

Studies which have dealt with message order, structure, or organization will be discussed in three groups: (1) those which treat but do not support the present hypothesis, (2) those which follow similar procedures but are irrelevant to this argument, and (3) those which support the general hypothesis of this study.

The only study which has come to this writer's attention that falls in the first category is that of K. C. Beighley. Beighley had several speech students outline the material from two speeches and arrived at the "best" order by democratic means. The paragraphs were then randomized with the following restrictions: paragraph number one was not allowed to be in the first five, no two paragraphs were permitted to be in the original order, and no three paragraphs of the same main point were allowed to follow each other. Comparison of the mean comprehension scores for the "organized" and "disorganized" versions failed to show a significant difference.[16]

Though Beighley's result does not support the present hypothesis,

---

[16] K. C. Beighley, "An Experimental Study of the Effects of Four Speech Variables on Listener Comprehension," Speech Monographs, XIX (November, 1952), 249-258.

it can not be considered as negative evidence because: (1) failure to reject the null hypothesis does not necessarily disprove the theoretical hypothesis, (2) it is possible that a passage of paragraph length provides sufficient context to reach an optimum level of organization beyond which subjects do not attend to structural dependencies, and (3) Beighley states,

> Where it was desirable to alter the wording of a transition at the beginning of a paragraph in order to make the material read more smoothly, such a change was made. Slight rewording was necessary at the beginnings of only two paragraphs in each of the two speeches.[17]

This could be interpreted as saying that the random versions were nearly as well patterned as the structured versions, and the small differences which did occur were promptly corrected. The negative result might have been predicted, depending upon what Beighley meant by "read more smoothly."

The following studies are included only because they have dealt with the effect of message order on receiver comprehension. Gulley and Berlo, defining a cell as "a message unit, consisting of an assertion and the evidence supporting that assertion," varied the order of cells as well as the order of assertions and evidence within the cell, producing six experimental messages. They found no significant difference among the six treatments on the retention of the propositions, assertions or evidence.[18] Sponberg found total

---

[17] Ibid., p. 251.

[18] Halbert E. Gulley and David K. Berlo, "Effect of Intercellular and Intracellular Speech Structure on Attitude Change and Learning," Speech Monographs, XXIII (November, 1956), 287-297.

retention of three assertions were significantly superior when the strongest assertion was presented first; that is, in anti-climax order.[19] However, replications of that study by Gilkinson et al have not produced the same result.[20] Though studies of this kind have investigated effects of order of message units on receiver comprehension, they have generally compared two or more "structured" orders.

A study which dealt with some of the same concepts as this study, but which used a different approach, is that of Peterson. Peterson took two passages from a social studies book and modified them for "better organization." Peterson considered fourteen items in her modification of the passages, and though significant results were obtained, it is impossible to discover which or how many of these variables were responsible for the increase in comprehension test scores.[21]

The remaining studies to be cited fall in the third class; i.e., tend to support the present hypothesis. Hamilton conducted a series of experiments in reading and states:

---

[19]Harold Sponberg, "The Relative Effectiveness of Climax and Anti-Climax Orders in an Argumentative Speech," Speech Monographs, XII (No. 1, 1946), 35-44.

[20]Howard Gilkinson, S. F. Paulson, and D. E. Sikkink, "Effects of Order and Authority in an Argumentative Speech," Quarterly Journal of Speech, XL (April, 1954), 25-26.

[21]Eleanor M. Peterson, Aspects of Readability in the Social Studies (New York: Bureau of Publications, Teachers College, Columbia University, 1954), pp. 25-26.

> It has . . . been found that for all subjects ranging from the age of nine or ten up to maturity the constituent elements of context, i.e., the phrase, the sentence, the paragraph, etc., have each a definite measureable value as an aid in the perception of words. The influence of these factors varies with different individuals, with maturity, with practice, and with different selections of reading matter, the latter varying directly with the perfection of context.[22]

Apparently what Hamilton means by "perfection of context" is the degree of dependency between message units, for his method was to randomize message units of different sizes and to compare the "average minimum reading time per word." That is, for a given passage, one experimental message was created by randomizing the paragraphs, another by randomizing sentences, another by randomizing phrases, and the final one by randomizing words, and the average time required to recognize each word increased with each reduction in the size of the structured units.[23]

Support for the validity of Hamilton's time index is found in a study by Howes. Howes shows that duration threshold (tachistoscopic presentation) and recognition time (continuous presentation) are inversely related to word probability. This result was obtained by correlating the two indexes with word frequencies which were obtained by actual word count.[24]

---

[22] Francis Marion Hamilton, The Perceptual Factors in Reading ("Archives of Psychology," Vol. I, No. 9; New York: The Science Press, 1907), p. 52.

[23] Ibid., 14.

[24] Davis H. Howes, "The Definition and Measurement of Word Probability" (unpublished Ph.D. dissertation, Dept. of Psychology, Harvard University, 1950).

It is inferred from these two studies that the size of the contextual unit (i.e., the length of the unbroken series of dependent elements) affects the conditional probabilities of those elements, and that the effect is subsequently reflected in the time required to recognize each element in sequence. Some support is provided for this inference by another study. Miller and Friedman, using a very short passage with one letter omitted and counting the percent right on the subjects' first guess, conclude that (1) the more content the more accuracy, (2) accuracy is greatest when content is symmetrical around the deleted character, and (3) left content is easier to use than right content.[25]

The Miller and Friedman study also shows the relationship between length of the dependent sequence and subjects' ability to make predictions about missing units. Another study which deals with predictability and contextual constraint is that of Selfridge. Two paragraphs from a children's story were read to a group of subjects; then each subject was asked to guess and write the next word. Subjects were then told the correct word and asked to guess the next, and so on. The mean predictability (percent of right guesses) was .3322. The low was .05 for the first word in an independent clause and the high .48 for the thirteenth word in an independent clause. (So few clauses in the passage contained more than thirteen words that results beyond this point were unreliable.)

---

[25] George A. Miller and Elizabeth A. Friedman, "The Reconstruction of Mutilated English Texts," Information and Control, I (September, 1957), 38-55.

In a different experiment reported in the same thesis, with learning recall as the dependent variable, the first word in each clause was the most difficult to learn, and difficulty decreased with the amount of context preceding the tested word. A positive relation was found between recall and the degree of contextual determination.[26]

One final study relevant to the present research dealt with nonsense material. This study found that subjects recalled more nonsense syllables from a set with a lower average rate of information per syllable. Rate of information was controlled by having the subjects learn patterns in advance, the patterns representing different degrees of contextual determination. Up to a certain point a constant amount of information was learned, but beyond that point syllable learning did not increase in proportion to the reduction in information.[27] This study supports the Selfridge result on the relation between predictability and learning rate. It has the additional advantage of demonstrating the effect of structure in nonsense material where there is little possibility that the result could have been produced by another variable.

---

[26] Jennifer Selfridge, "Investigations into the Structure of Verbal Content" (unpublished honors thesis, Dept. of Psychology, Radcliffe College, 1949).

[27] Murray Aborn and Herbert Rubenstein, "Information Theory and Immediate Recall," Journal of Experimental Psychology, XLIV (September, 1952), 260-266.

It is contended that these several studies, taken as a group, leave no doubt that language users can respond to differences in language structure. If the "because" order is a structure of which most users of English are aware, it seems that alteration of the dependent relationships among sentences to produce structures which are less familiar to the respondent should reduce the accuracy of predictions made about the message. The hypothesis that successive steps of removal from a "because" order will reduce the accuracy of respondents' predictions about the missing parts of the message is therefore subjected to an empirical test.

CHAPTER II

This chapter includes the design of the experiment and a description of the technique of administration.

DESIGN OF THE EXPERIMENT

The present study was designed to compare the effects of varying sentence order on (1) comprehension of the message and (2) the relative redundancy of the message.

The independent variable in this study is sentence order. The two dependent variables are (1) comprehension of the message as measured by close scores and (2) the relative redundancy of the message as estimated from the number and frequency of different responses to blanks in a mutilated form of the message.

In order to make predictions about differences in expected scores, seven order-treatments of the message were obtained by a systematic method. The first order is a "because" order. Given a "because" structured message, the next step was to control the variation of structure. A random order might have been used, but there was the possibility that a random drawing would produce the structured order or some close approximation of it. This was obviously to be avoided. The problem was somewhat complicated by the fact that there are 128 orders that are logically correct (i.e., any coordinate units and their subunits can be interchanged). Moreover, if this order may be called <u>deductive</u>, there is a like number of <u>inductive</u> orders which are equally well structured and,

possibly, equally familiar to most users of English. Even eliminating these possibilities, there would be no way of estimating the "distance" between the random and structured orders and no basis for making predictions about the relative effectiveness of more than two orders.

It was observed that moving the topic sentence to the middle of the message placed the topic sentence as far as possible from its proper position in either a deductive or inductive pattern. Moving the two subheadings to the center of their cells also placed them as far as possible from their positions in a deductive or inductive sequence. It was further observed that the first of these moves required seven binary transpositions (exchanges of adjacent sentences), and the second required six more. The possibility of an ordinal scale of structure was exposed. Some trial and error shuffling produced an order which was a maximum distance (43 transpositions) from a deductive or inductive order, and any further transposition moved toward one of these orders. Three more orders were produced which are at least 20, 27, and 35 transpositions from a deductive or inductive order. The seven experimental message orders are represented symbolically in Fig. 1.

CONTROL

Differences in difficulty among experimental treatments resulting from variables such as word length and sentence length were controlled by using identical sentences in all treatments.

To control for differences in individual ability to produce close scores, a second message was prepared which matched the ex-

ORDERS						
1	2	3	4	5	6	7
T.S.	I	IA	IA	IA1	IA1	IA1
I	IA	IA1	IA1	IA	IB	IIA
IA	IA1	IA2	IA2	IA2	IA2	IA2
IA1	IA2	I	II	II	II	II
IA2	IB	IB	IB	IB1	IB1	IB1
IB	IB1	IB1	IB1	IIA	IIB	IIB
IB1	IB2	IB2	IB2	IB2	IB2	IB2
IB2	T.S.	T.S.	T.S.	T.S.	T.S.	T.S.
II	II	IIA	IIA	IIA1	IIA1	IIA1
IIA	IIA	IIA1	IIA1	IB	IA	IB
IIA1	IIA1	IIA2	IIA2	IIA2	IIA2	IIA2
IIA2	IIA2	II	I	I	I	I
IIB	IIB	IIB	IIB	IIB1	IIB1	IIB1
IIB1	IIB1	IIB1	IIB1	IIB	IIA	IA
IIB2	IIB2	IIB2	IIB2	IIB2	IIB2	IIB2
0	7	13	20	27	35	43

TRANSPOSITIONS FROM FIRST ORDER

Fig. 1.-- The seven orders of the experimental message (Sentences are represented by conventional outline symbols to show the logical relationships.)

perimental message in difficulty according to the Flesch readability formula. Mutilated in the same way as the experimental message, the control message was submitted to all subjects along with one order of the experimental message.

Differences in classroom groups of subjects were controlled by (1) limiting subjects to undergraduate students in the College of Communication Arts and in the Department of Communication Skills and (2) by giving each classroom group approximately equal proportions of the different experimental treatments.

In addition to the primary variables of this study the experimenter wished to know if subjects could predict the scores that they would make on the tests; that is, if they were consciously aware of any difficulty produced by transposing sentences in a message. A five point scale was prepared on which subjects were asked to rank the test between Easy and Difficult. As a possible measure of motivation, two other scales, Interesting . . . Uninteresting and Useful . . . Worthless, were added. The set of three scales was provided for both the control and the experimental messages.

One page of printed instructions was also prepared. Instructions, messages, and judgment scales were reproduced by mimeograph. A typical experimental packet is presented in Appendix B.

ADMINISTRATION OF THE EXPERIMENT

Test packets were made up of one instruction sheet, the control message, the experimental message, and a sheet of judgment scales, in that order. Each test packet contained only one treatment of the

experimental message. The test packets were arranged in series, so that any seven packets used in series would represent all orders of the experimental message. Subjects were tested during a regular class period. The writer served as experimenter in all cases.

In the early phases of testing, several graduate students were inadvertently tested. They were dropped from the sample for the sake of homogeneity and their test packets were replaced. One hundred forty undergraduates were tested, and these were distributed in seven groups of twenty each; therefore, it was possible to do the analysis with equal n's without discarding qualified sybjects. Subjects were students in the department of Communication Skills and the College of Communication Arts, Summer School, Michigan State University. The sample was composed of 56% freshmen, 19% sophomores, 12% juniors, and 13% seniors. Seventy-two percent of the subjects were males, and the mean age was 21.8 years.

Upon entering the classroom, the subjects were informed that they were to take a test which was a part of a Master's thesis experiment, and which was, in fact, a measure of their reading ability. They were told, "Since this experiment requires volunteer subjects, anyone who does not wish to take part may leave without penalty." None left. In eighty percent of the cases subjects were told by their instructors that the test scores would be used diagnostically to pinpoint any particular reading difficulty, but in all cases subjects were told that their scores would not count directly on their course grade.

The test packets were passed out by the experimenter, and the subjects were asked to read the first page of printed instructions.

Pencils were furnished by the experimenter. After everyone had sufficient time to read the instructions, the experimenter went over the instructions rapidly calling attention to the statement, "one word and only one word in each blank," and adding that this could be any word -- large or small. Subjects were told that they would have ten minutes, and that they would be given a five and a one minute warning. (The time warnings were given because it was felt that one's ability to estimate elapsed time should not be included as an experimental variable.) Subjects were then asked if they had questions about the test procedure. There were not more than two in any group, and as soon as all subjects were satisfied with the instructions they were told to begin on message number one, which was the same in all cases--the control message.

At the end of five minutes subjects were told that their time was half gone, and it was suggested that they try to get through the message to take advantage of the easy blanks. The one minute warning was given without comment. At the end of ten minutes the subjects were told to finish the word they were writing and stop. After the first test subjects were again asked for questions. There were none.

In preparation for the experimental message, subjects were cautioned not to be disturbed by the different arrangement of blanks.[28]

---

[28]Since the same words were deleted in all treatments, transposing the sentences gave some treatments a somewhat different appearance than the control message. It is held that this was not a confounding variable, since the within-sentence relationships remain constant.

After approximately one minute they were told to continue with the second message. The same timing procedure was followed as in the previous test.

After the second message had been completed the subjects were asked to turn to the last page of the packet and mark a point on the scales to represent their feelings toward the messages and the tests. A few subjects remarked to the experimenter in the discussion which followed that they had applied the Easy . . . Difficult scale to the test, the Interesting . . . Uninteresting scale to the subject matter, and the Useful . . . Worthless scale to the projected results of the study.

While the subjects marked the scales, the experimenter distributed red pencils. The subjects were then told that they could go back over the tests and make any additions or corrections they wished. They were further instructed not to erase any of their previous answers but to write second choices and fill out blanks that had not been completed during the timed interval. After about four minutes almost all subjects had stopped writing and indicated that they could do no more if given more time. One classroom group that had had previous experience with cloze procedure stopped writing before the end of the timed period and refused additional time. Since all treatments were equally represented, the scores of this group were retained.

An extraneous variable was inserted by chance with one group of twelve subjects. In the last ten seconds of the first timed period the lights went out. After a slightly prolonged rest interval,

the subjects indicated that they had adapted to the somewhat dim conditions and were anxious to go on. The experiment proceeded without further interruption. In the discussion which followed there was some comment about the value of a test of one's ability to read in the dark, but there were no apparent differences in the scores of this group and those of any other group. The unexpectedness of this situation did not seem sufficient cause for dropping these data from the experiment.

SCORING OF THE DATA

For the close score only exact replacement of the deleted word was counted as correct. Minor irregularities in spelling were not counted off, but changes in number or tense of verbs were considered incorrect. Since Taylor had indicated there was nothing to be gained, no attempt was made to evaluate or score synonyms.[29] For the measure of entropy, the frequency of each different word was recorded for each blank. Again, variations in spelling were not counted as different unless the variations produced another possible word; e.g., change in number or form. Completely unreadable words were scored as different "wrong" answers as were blanks. The timed scores were taken as the best estimate since the time limit minimized reorganization by the subject. In the "extra time" period subjects had been instructed to make corrections and fill any remaining blanks with a red pencil. Untimed scores were obtained by adding the correct red answers to the timed score and subtracting

---

[29] "Recent Developments in the Use of Close Procedure," 48.

the number of words that had been incorrectly changed. The very low proportion of the incorrect changes (less than 1% of total changes) suggests a high degree of certainty accompanying a correct closure.

CHAPTER III

This chapter reports results of a statistical analysis of the data.

HYPOTHESES

Empirical evidence for or against the theoretical hypothesis of this study is to be obtained by testing seven statistical hypotheses. Stated in null form they are:

1. The mean cloze scores on the control message for the seven treatment-groups are the same.

   (i.e., $\bar{M}_1^c = \bar{M}_2^c = \cdots = \bar{M}_7^c$)

2. The mean cloze scores on the seven timed experimental tests are the same.

   (i.e., $\bar{M}_1^e = \bar{M}_2^e = \cdots = \bar{M}_7^e$)

3. Interaction between individual ability and the structure variable is zero.

   (i.e., $\bar{M}_{ij} - \bar{M}_{i.} - \bar{M}_{.j} + \bar{M} = 0$)

4. The mean cloze scores on the untimed experimental tests are the same.

   (i.e., $\bar{M}_1 = \bar{M}_2 = \cdots = \bar{M}_7$)

5. The mean entropy scores for the seven experimental messages are the same.

   (i.e., $\bar{H}_1 = \bar{H}_2 = \cdots = \bar{H}_7$)

6. Rank order correlation between "transposition" ranks and "cloze" ranks equals zero.

7. Rank order correlation between "transposition" ranks and "entropy" ranks of the seven treatments equals zero.

The .05 level of significance was chosen to test these hypotheses.

The mean scores on which this analysis is based are shown in Table 2.

TABLE 2.--The Mean Scores on the Control and Experimental Messages for Seven Groups of Subjects

CLOZE SCORES	I	II	III	IV	V	VI	VII
Control							
timed	23.5	22.4	22.4	25.2	22.8	23.2	21.3
untimed	24.7	23.4	23.1	26.2	24.3	23.8	22.9
Experimental							
timed	21.5	19.2	17.8	18.8	18.3	18.4	15.3
untimed	21.7	19.7	18.2	19.4	18.4	19.2	15.8
CLOZE PROPORTION	.448	.401	.372	.393	.381	.383	.319
RELATIVE ENTROPY	.488	.586	.572	.562	.533	.520	.611
RELATIVE REDUNDANCY	.512	.414	.428	.438	.467	.480	.389
RELATIVE MISDIRECTION	.138	.100	.132	.118	.157	.166	.146

ANALYSIS

Before testing for differences among means with the analysis of variance it was necessary to test the assumption of homogeneity of variance. The Bartlett test for homogeneity was applied to the timed cloze scores for the control message. The chi-square obtained (Table 3) did not permit the rejection of the hypothesis that the variances are equal.

TABLE 3.--Bartlett Test for Homogeneity of Variance Among Seven Groups of Cloze Scores on the Control Message

Group	N_i	n_i	$n_i s_i^2$	s_i^2	$\log_{10} s_i^2$	$\frac{1}{n_i}$
1	20	19	485.75	26.25	1.41913	.05263
2	20	19	795.34	41.86	1.62180	.05263
3	20	19	640.49	33.71	1.52763	.05263
4	20	19	311.89	16.42	1.21537	.05263
5	20	19	555.18	29.22	1.46553	.05263
6	20	19	638.02	33.58	1.52608	.05263
7	20	19	720.29	37.91	1.57864	.05263
Sum	140	133	4160.05		10.35418	.36841

$C = 1.02005$ $B' = 4.9218$ $B = 4.82$ $[\chi^2.95\ (6) = 12.6]$

The next step was to test hypothesis number one--the mean cloze scores on the control message for the seven treatment-groups are the same--with an analysis of variance, single variable design. The results are reported in Table 4. The obtained F fails to reject the hypothesis that the means are equal.

TABLE 4.--Analysis of Variance of the Control Message

	Sum of Squares	df	Mean Square	F	p
Between Means	169	6	28.17	1.09	--
Within	3425	133	25.75		
Total	3594	139			

As far as the overall purpose of the experiment is concerned, the failure to reject these hypotheses indicates that the groups were effectively matched in ability.

In the analysis of the experimental scores it was again necessary to test for homogeneity of variance. The Bartlett test of homogeneity was applied to the experimental timed scores. Although the range of variance is somewhat larger than on the control message the obtained chi-square is not significant at the .05 level and again fails to reject the hypothesis that the variances are equal (Table 5).

Given this evidence of homogeneity it is possible to test the second hypothesis--the mean close scores on the seven timed experimental tests are the same--with an analysis of variance design. A single variable design yields an F of 2.67 and the hypothesis is rejected at the .05 level of confidence (Table 6).

TABLE 5--Bartlett Test for Homogeneity of Variance Among Seven Groups of Close Scores on the Experimental Timed Message

Group	N_i	n_i	$n_i s_i^2$	s_i^2	$\log_{10} s_i^2$	$\frac{1}{n}$
1	20	19	295.00	15.526	1.19106	.05263
2	20	19	773.75	40.723	1.61090	.05263
3	20	19	590.55	31.081	1.49249	.05263
4	20	19	452.55	23.818	1.37690	.05263
5	20	19	240.20	12.642	1.10181	.05263
6	20	19	420.80	22.147	1.34531	.05263
7	20	19	260.20	32.642	1.51378	.05263
Sum	140	133	3393.07		9.63225	.36841

$C = 1.02005$ $B' = 9.37243$ $B = 9.19$ $[\chi^2 .95\ (6) = 12.6]$

TABLE 6--Analysis of Variance -- Experimental - Timed Scores

	Sum of Squares	df	Mean Square	F	p
TREATMENT	407.95	6	67.99	2.67	<.05
Within	3393.05	133	25.51		
Total	3801.00	139			

Since close scores on the timed control message were found to correlate .58 (< .01) with timed experimental scores, it was thought advisable to remove that portion of variance which could be attributed to individual ability to produce close scores. In the two way classification used in this test it was possible to test three separate hypotheses:

(a) The row means are equal (i.e., there is no difference attributable to the control variable).

(b) The column means are equal (i.e., there is no difference among treatments).

(c) Interaction is zero (i.e., high ability and low ability subjects are equally affected by the experimental variable).

The experimental-timed close scores were divided into two sections. One section was composed of those individuals who scored more than 23 on the control test ("high" group) and the other of those who scored 23 or less ("low"). This division gave approximately equal frequencies in the fourteen sub-classes. The computation used the means of the sub-classes as scores. The results of the test (Table 7) reject hypotheses (a) and (b) but fail to reject (c). The conclusions are: (1) The means of the high group are different from the means of the low group. (2) Means produced by the seven experimental treatments are different. (3) There is no significant interaction between individual ability and message structure.

TABLE 7.--Analysis of Variance of Experimental-Timed Scores Divided on the Basis of Scores Made on the Control Message (23 and below / more than 23)

	Sum of Squares	df	Mean Square	F	p
Individual Differences	82.09	1	82.09	40.43	<.01
Treatment	39.62	6	6.60	3.25	<.01
Interaction	4.56	6	.76	.37	--
Total	126.27				
Error	256.47	126	2.03		

The same operations were performed on the untimed experimental scores with almost identical results. The groupings were made on the basis of the untimed control scores, one group being those who scored 25 or less on the untimed control and the other those who scored more than 25.

TABLE 8.--Analysis of Variance -- Experimental-Untimed Scores

	Sum of Squares	df	Mean Square	F	p
Treatment	277.14	6	62.86	2.43	< .05
Within	3438.00	133	28.85		
Total	3815.14	139			

TABLE 9.--Analysis of Variance of Experimental-Untimed Scores Divided on the Basis of Scores Made on the Untimed Control Message (25 and below/ more than 25)

	Sum of Squares	df	Mean Square	F	p
Individual Differences	59.57	1	59.57	41.37	< .01
Treatment	36.02	6	6.00	4.17	< .01
Interaction	7.18	6	1.19	.83	--
Total	102.77				
Error	181.36	126	1.44		

In an effort to pinpoint the source of the differences among means, *t*-tests were applied to the adjacent means of the experimental scores. The results of the *t*-tests appear in Tables 10 and 11. The total group and the two subgroups were tested individually. None of the *t*'s are significant at the .05 level.

TABLE 10.--*t*-Tests Between Adjacent Means--Experimental Timed Close Scores

	Total Group			High Group			Low Group		
Between	*t*	df	p	*t*	df	p	*t*	df	p
1 & 2	1.34	38	--	.49	17	--	.87	19	--
2 & 3	.74	38	--	.52	18	--	.57	18	--
3 & 4	- .60	38	--	- .87	20	--	.48	16	--
4 & 5	.41	38	--	1.07	20	--	-1.14	16	--
5 & 6	- .75	38	--	- .33	17	--	- .56	19	--
6 & 7	1.86	38	--	.74	16	--	1.72	20	--

TABLE 11.--*t*-Tests Between Adjacent Means--Experimental Untimed Close Scores

	Total Group			High Group			Low Group		
Between	*t*	df	p	*t*	df	p	*t*	df	p
1 & 2	.85	38	--	.60	19	--	.96	17	--
2 & 3	.45	38	--	.08	17	--	.84	19	--
3 & 4	- .39	38	--	.22	20	--	.33	16	--
4 & 5	.51	38	--	.44	20	--	.97	16	--
5 & 6	- .38	38	--	- .55	16	--	- .34	20	--
6 & 7	1.34	38	--	1.68	17	--	1.36	19	--

One of the primary hypotheses of this study was that mean entropy (freedom of choice) scores for the seven experimental orders would increase in direct relation to the number of transpositions from a "because" order. (The null version of this hypothesis is that the mean entropy scores for the seven experimental treatments are the same.) The analysis of variance was again used to test for differences among means. A two-way classification, by blanks and by treatments, was used. There were forty-eight blanks and seven treatments. Each blank is the same for all seven treatments. Each blank in each treatment represents the responses of twenty subjects. The results reported in Table 12 indicate that there were differences among blanks and among treatments which would occur less than 1% of the time by chance. The null hypothesis is rejected.

TABLE 12.--Analysis of Variance of Entropy Scores -- Two-Way Classification by Blanks and Treatments

	Sum of Squares	df	Mean Square	*F*	*p*
Blanks	17.5631	47	.3737	40.18	< .01
Treatment	.5088	6	.0848	9.12	< .01
Residual	2.6213	282	.0093		
Total	20.6932	335			

As to the direction of these differences, Fig. 2 shows the mean redundancy scores for the seven orders and illustrates the relationship between relative redundancy (RR), close proportion (CP), and misdirection (M). (Relative redundancy is 1 - relative entropy.)

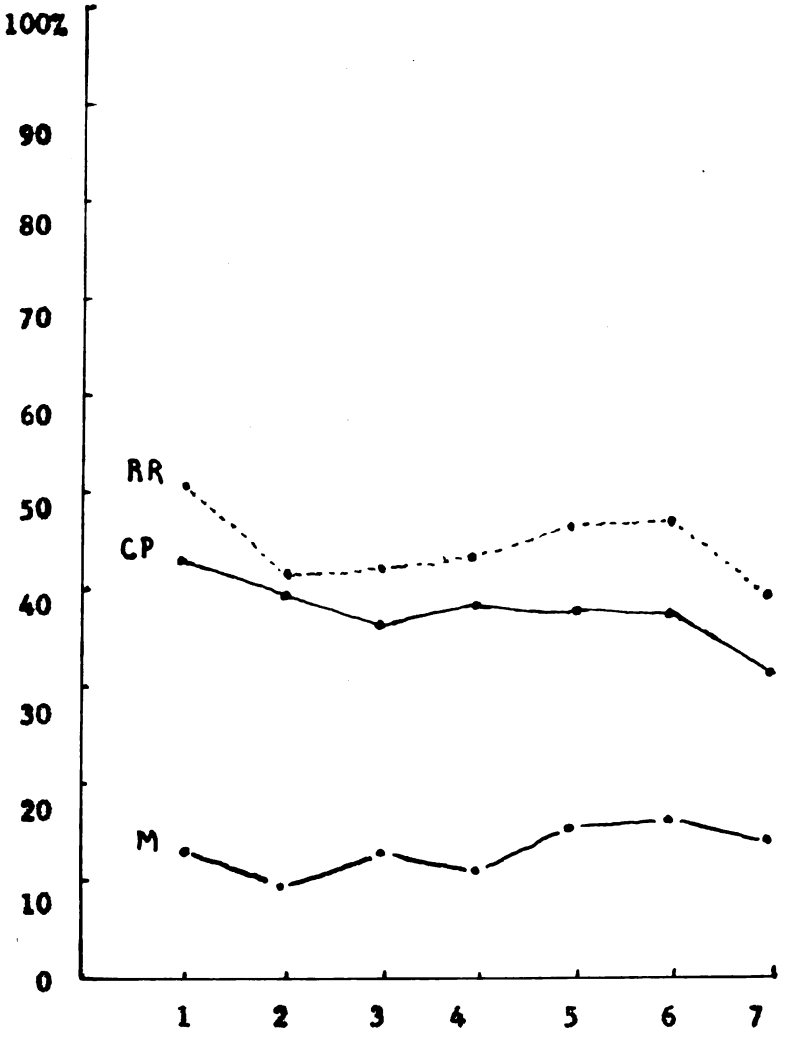

Fig. 2.- Comparison of mean relative redundancy (RR), close proportion (CP), and misdirection (M) for the seven treatments

The message treatments can be ranked according to difficulty by (1) number of transpositions from a "because" order, (2) obtained cloze scores, and (3) relative redundancy scores. Computation of rank order correlation coefficients shows a significant relation between (1) and (2) ($\rho = .75$, $p < .05$). There is no evidence of a relation between (1) and (3) ($\rho = .29$) or (2) and (3) ($\rho = .54$).

To complete the analysis of the data, a Pearson r was computed to determine the relation between a subject's cloze score and his score on each of the five-point evaluation scales. The scales were coded 5 for the most favorable response--Easy, Interesting, and Useful--and 1 for the most unfavorable response. The results reported in Table 13 provide no evidence of a relation between a subject's evaluation of the message on these three dimensions and the cloze score which he produced.

TABLE 13.-- Correlation between Cloze Scores and Scales of Ease, Interest, and Utility

Scale	r	p
Easy . . . Difficult	.063	--
Interesting . . . Uninteresting	.041	--
Useful . . . Worthless	.151	--

As reported in Table 14, there were no significant differences among the seven treatment-groups on mean ratings of easiness, interest, or utility.

TABLE 14.--Analysis of Variance among Treatments for Three Judgment Scales

Scale	1	2	3	4	5	6	7	*F*	*p*
E - D	3.0	2.2	2.3	2.1	2.7	2.8	2.6	2.20	--
I - U	3.9	2.9	3.6	3.6	3.4	3.8	4.0	1.54	--
U - W	3.5	3.6	3.5	3.5	3.6	3.6	3.0	.91	--

CHAPTER IV

This chapter includes a summary of the research, conclusions, discussion, some observations on Information Theory, and suggestions for further research.

SUMMARY

The present study is an attempt to analyse the relationship between sentence order and message comprehension. More exactly, it is an investigation of the relationship between structural variation and organization. A message was prepared which met all the requirements of a "because" order outline. Six alternative messages were prepared which were assigned predicted ranks of difficulty in terms of transpositions of sentences. For example, the most difficult, or seventh ranking message, was 43 transpositions from the first ranking message. The seven treatments were then mutilated and prepared according to the requirements of close procedure. Close scores were subsequently obtained for the seven treatments from seven randomly selected groups of subjects. An estimate of the relative entropy of the seven treatments was obtained by a procedure outlined by Wilson L. Taylor. Judgments of the perceived difficulty, interest, and utility of the treatments were also obtained from the subjects. Mean close and entropy scores were compared by means of analysis of variance, and correlation coefficients computed for close scores and the judgments of difficulty, interest, and utility. Rank order correlations were also computed between the

mean cloze and entropy scores and the predicted ranks of difficulty.

CONCLUSIONS

1. There is a significant relation between cloze scores and transpositions from the "because" order of sentences. This conclusion is based on the observations of significant differences in mean cloze scores among order-treatments and significant rank order correlation between cloze ranks and transposition ranks. However, the relation does leave something to be desired in that t-tests between adjacent means are not significant and the rank order correlation coefficient is only .75.

2. There are significant differences among mean relative entropy scores for the seven treatments, but the rank-order correlation between relative entropy and transpositions from a "because" order is not significant.

3. There is no evidence of a relation between a subject's cloze score and his judgments of difficulty, interest or utility, nor is there a treatment difference among mean judgments on any of the scales.

DISCUSSION

The relation between cloze scores and the transpositions from the "because" order, though statistically significant, does not permit the conclusion that any transposition of sentences will have the same effect on the comprehension of a message. Such a conclusion would necessarily assume that there is no possibility of two orders of the same sentences being equally well structured, and that the same degree of dependency obtains between all pairs of adjacent

sentences. The first of these assumptions is denied by the existence of 128 possible "because" orders of the sentences used in this experiment. The second is denied by definition, in that the "because" order permits co-ordinate sentences which may be transposed without modifying the structure of the message. It would seem that the technique of counting transpositions from any deductive or inductive order would in fact control the above-mentioned factors; however, the particular orders developed for this research specified the consequences of the transpositions and did not permit alternative orders to occur.

The transpositional idea of varying structure was selected from a number of alternative methods because: (1) it had the potential of an ordinal scale of structure and (2) it seemed to be the most likely to be of some use outside the experimental frame. On the basis of further thinking, and an analysis of the results of this research, an alternate method is proposed as a tool for further research which would require much less effort on the part of the experimenter and eliminate the bias of transposing selected sentences.

Successive approximations to a given order could be obtained in this way. Starting with a given order of message units, every n^{th} unit (letter, word, sentence, paragraph) is deleted from the set and reinserted in random order. The second approximation is obtained by randomizing every $(n-1)^{th}$ unit, etc. The size of n would be determined by the number of units in the set and the number of approximations desired. The final order would of course be a random arrangement of all units, of size x_0, in the message. The chance of

reproducing the original order is considered negligible, and this procedure has the added advantage of being applicable to many kinds of existing messages. Such a procedure would, by definition, provide an ordinal scale of structure providing only that some structure obtained in the original message.

SOME OBSERVATIONS ON INFORMATION THEORY

During the process of analyzing and reporting the present research the author has made some observations about Information Theory, some of which are directly related to the present study and some of which are of a more general nature. These observations are reported here for whatever benefit they may be to other communication researchers.

The first of these observations is that Taylor's application of Information Theory obtains a biased estimate of the relative entropy in a system, in that his estimate of maximum entropy is dependent on the number of subjects giving responses. The bias is related to the fact that, in Taylor's calculations, "choice" is defined differently for the computations of absolute and maximum entropy. Choice is defined for absolute entropy as a different word; i.e., the number of choices is the number of different words actually supplied for a given blank. For maximum entropy, however, choice is defined as the number of subjects; i.e., the number of choices is the number of different words that hypothetically could occur IF all the subjects gave different answers.[30]

---

[30] Wilson Lewis Taylor, "Application of 'Close' and Entropy Measures to the Study of Contextual Constraint in Samples of Continuous Prose" (unpublished Ph.D. dissertation, The University of Illinois, 1954) pp. 17-18.

In an example, using an n of 100, Taylor points out his own error.

> If 96 out of 100 subjects choose the same word, a comparatively large amount of organization [redundancy] must be inferred even if the other four all choose different other words and the number of alternatives becomes five. However, if each of the five kinds were chosen by twenty subjects, absolute entropy would reach its maximum value for that number of kinds.[31]

This analysis seems to agree with Weaver's statement.

> If one reckons, for this case, [a two choice situation] the numerical value of H, it turns out that H has its largest value, namely one, when the two messages are equally probable; that is to say when $p_1 = p_2 = \frac{1}{2}$ . . . .[32]

In application, however, Taylor ignores another statement of Weaver's,

> When the number of cases is fixed, we have just seen that then the information is greater, the more nearly equal are the probabilities of the various cases. There is another important way of increasing H, namely by increasing the number of cases. More accurately, if all the choices are equally likely, the more choices there are, the larger H will be. (Underline mine.)[33]

The practice of defining "choice" differently for the numerator and denominator of the proportion "relative entropy" results, then, in a low estimate of the relative entropy of a system.

To demonstrate the extent of this bias the following graphs have been prepared. In these figures, the proportion of relative entropy is computed for a set of hypothetical cases. The cases

---

[31] Ibid., p. 19.

[32] Claude E. Shannon and Warren Weaver, The Mathematical Theory of Communication (Urbana: The University of Illinois Press, 1949), p. 105.

[33] Ibid., p. 106

are the same in both figures (3 and 4). Figure 3 is obtained using Taylor's formula, and figure 4 using the corrected formula which defines *choice* in all cases as the number of different words supplied by the subjects. The vertical dimension of the graphs is the proportion of relative entropy. The horizontal dimension is a hypothetical close score. Lines A, B, and C in figure 3 are computed for three sizes of sample, *n* equals 10, *n* equals 20, and *n* equals 100. For these lines the assumption is made that only two different answers are given, one of which is "right" and one "wrong". Line A in figure 4 makes the same assumptions and is the same for all *n*'s.[34] Lines X, Y, and Z, in both cases, assume a given close proportion and that all the wrong answers are different ones; i.e., as the close proportion increases the number of choices decreases. The principle difference is to be observed between lines A, B, and C (Fig. 3) and line A (Fig.4), where the close proportion equals 50%.

Another observation that can be made is that what Taylor calls *misdirection* is just another source of entropy. In fact, if remaining relative entropy is computed according to Taylor's formula, using the corrected figure for maximum entropy, the sum of remaining relative entropy and the close proportion may exceed unity. What might be considered misdirection, using the corrected formulas, is the proportion of relative redundancy minus the close proportion. In the case where close exceeds redundancy, misdirection would,

---

[34]This is the same curve which Shannon presents for a two-choice situation with probabilities of (p) and (1-p). *Ibid.*, p. 20.

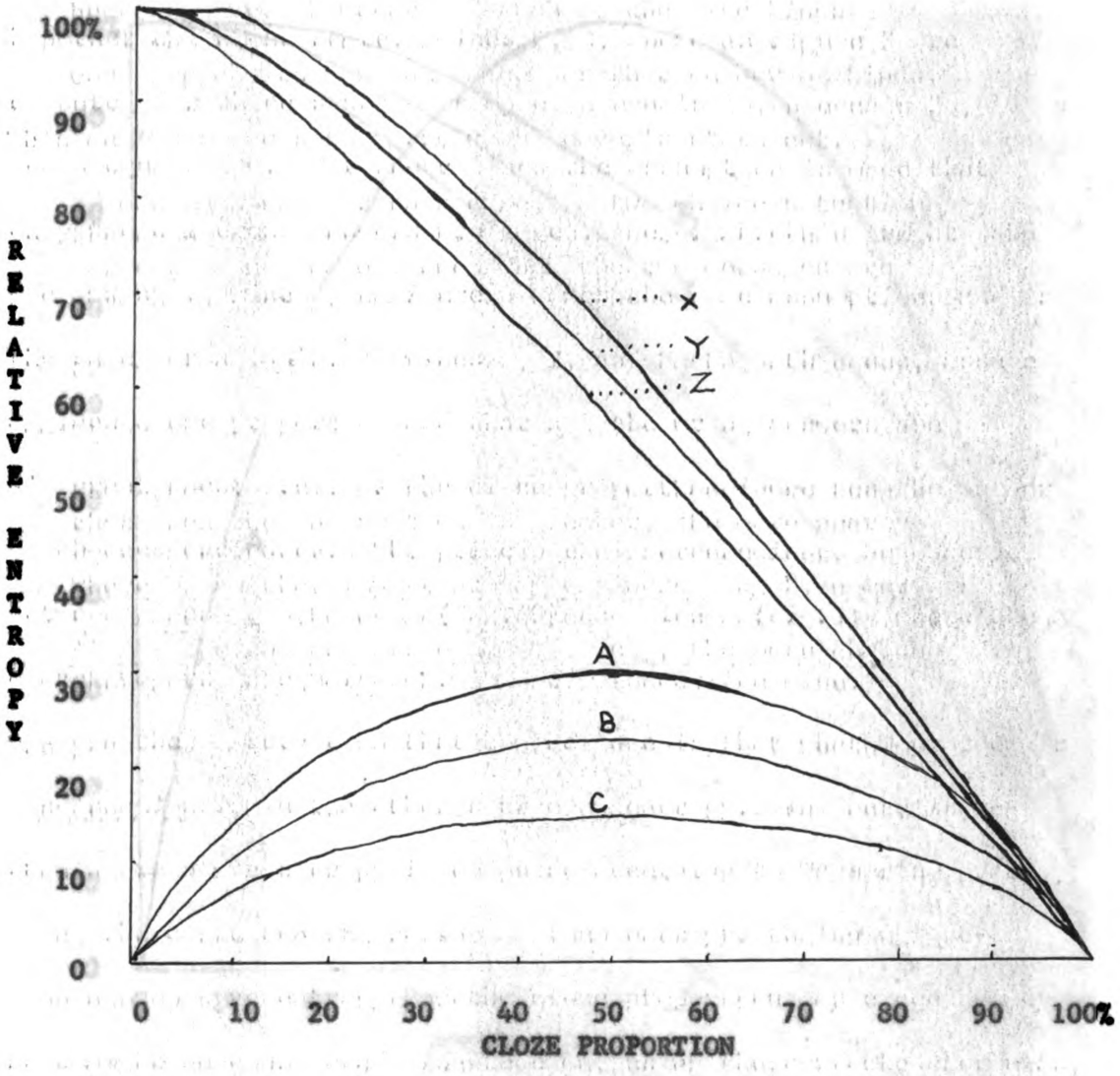

Fig. 3.--Taylor's calculation of Entropy for two ideal cases and n's of 10, 20, and 100

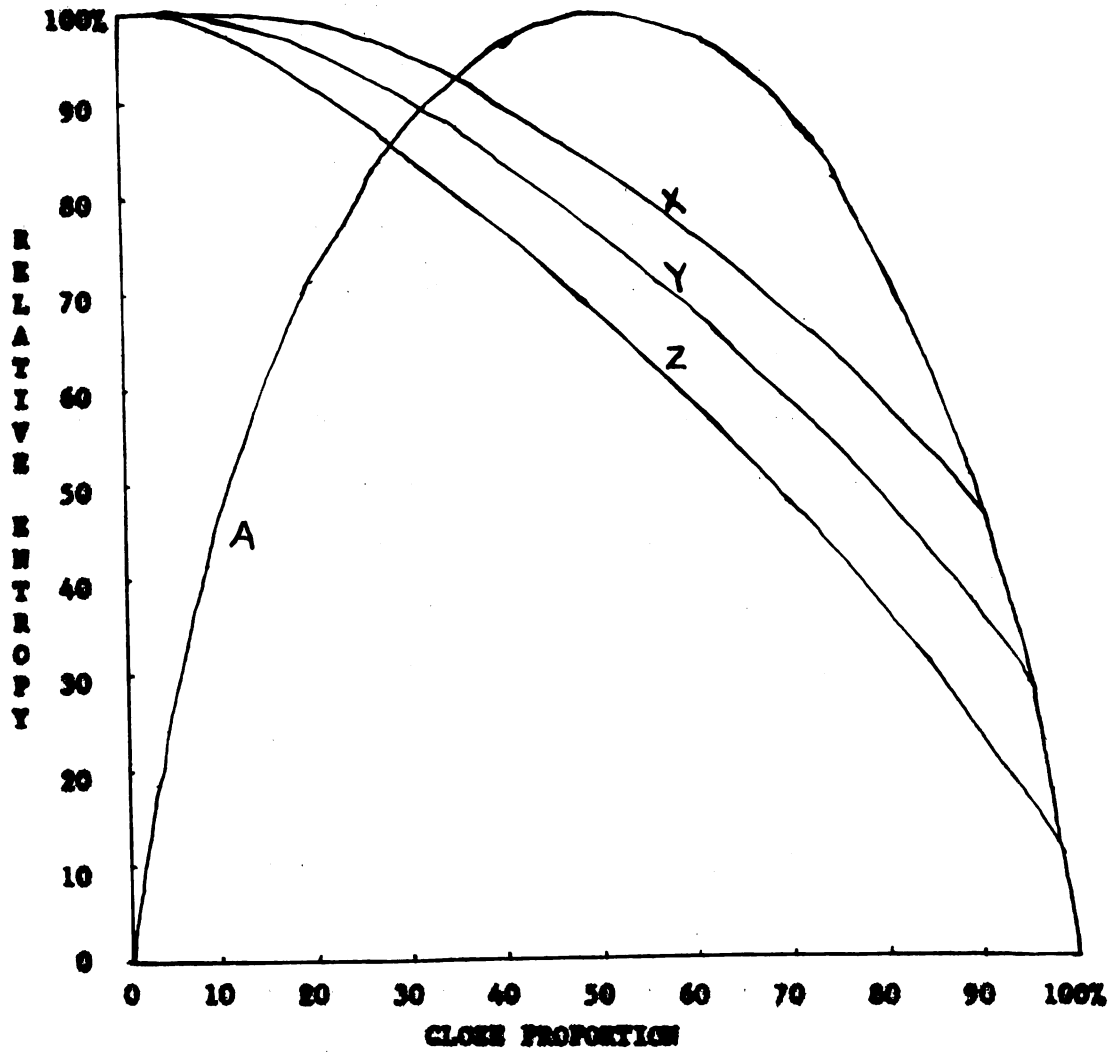

Fig. 4.--Corrected calculation of Entropy for two ideal cases and $\underline{n}$'s of 10, 20, and 100

then, have a negative value; i.e., direction. In the 336 computations in this study, the cloze proportion did exceed the redundancy figure in a majority of the cases when the corrected formulas were used. Fig. 5 shows three lines, the mean scores for the seven treatments, showing the relationship between the cloze proportion and the average relative redundancy as computed by the two methods previously described. It will be noted that the average misdirection, as defined above, has disappeared in the revised computations.

Fig. 5 also shows that there is much closer agreement on the relationship among treatments between cloze proportion and the revised estimate of relative redundancy. Rank correlation between cloze and redundancy measures is now significant at the .05 level. The correlation between redundancy ranks and predicted ranks, however, still does not attain significance.

The two-way classification analysis of variance was repeated for the revised entropy scores, and again differences were significant between blanks and treatments ($p < .05$). There is, therefore, no change to be made in the conclusions of the study as stated on page 45 of this paper. If, however, one wishes to attribute some absolute value to the proportion of relative entropy or redundancy in a situation, the differences in the results of the two methods of computation could become extremely significant.

The final observation to be made has a purely theoretical value (if any) and may seem inconsistent with the preceding discussion. Entropy, or information, as defined by Shannon has

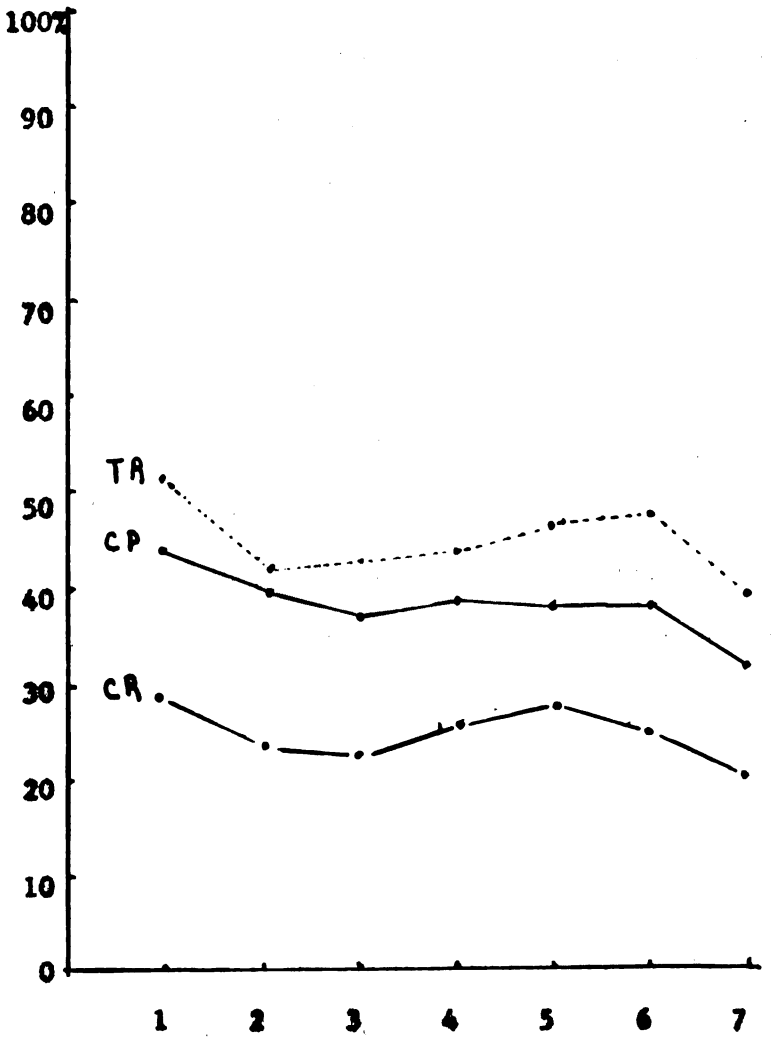

Fig. 5.--Comparison of cloze proportion (CP) and relative redundancy as computed by Taylor's method (TR) and the corrected method (CR)

nothing to do with denotative meaning, but is concerned only with the uncertainty of the source as to what the next symbol is going to be.[35] Now, if it can be assumed that the source has a purpose, a meaning to communicate (we speak now of a human source), this purpose will in many cases limit the choices available to him, thus reducing the entropy of the source. What about the case, however, in which the source is faced with two equally probable synonyms? This is a situation in which it seems a useful distinction could be made between redundancy and organization. That is, in the situation just described, redundancy has a value of zero, while it could be said that organization is unity. There is maximum uncertainty about the symbol and no uncertainty about the meaning. If the viewpoint is switched to that of the receiver, the receiver in the process of receiving a message may be quite sure of the meaning intended but maximally uncertain about the symbol to be used. In fact, it is on the basis of this discrepancy that a receiver is able to "correct" a source. Another situation in which this distinction holds--between redundancy and organization--is after a message is completed. Once the symbols are selected (for the source) or received without interruption (by the receiver) there is no uncertainty about what the symbols are (no entropy), but there may be a great deal of uncertainty about their intended meanings (low organization). For these reasons, it is questionable whether the figures obtained in this study, and called entropy and redundancy, are in any exact

---

[35] Ibid., pp. 99-100.

sense measures of information in the meaning intended by the developers of Information Theory. They are apparently, however, measures of some psychological phenomenon which warrants careful scrutiny and extensive investigation.

SUGGESTIONS FOR FURTHER RESEARCH

1. The relationship between "entropy" and "organization" needs to be thoroughly explored both theoretically and experimentally.

2. A series of comparative studies of the various methods of arranging messages could be conducted along the lines of the present study. Such a series is needed to answer the question, "What is the best way to arrange a message?"

3. A series of studies is needed to establish the relation between--and interaction of--order effects at the various levels of message construction, where levels are defined in terms of size of unit under investigation.

4. Order effects could be investigated across languages, or among subjects of different cultural backgrounds who speak the "same" language.

The present study has shown that the deletion-completion method (cloze procedure) is sensitive to the order variable. Combination of this tool with the randomizing-every-n^{th}-unit method of obtaining approximations to structure suggests a frame within which the suggested investigations can be conducted.

BIBLIOGRAPHY

Aborn, Murray, and Rubenstein, Herbert. "Information Theory and Immediate Recall," Journal of Experimental Psychology, XLIV (September, 1952), 260-266.

Beighley, K. C. "An Experimental Study of the Effects of Four Speech Variables on Listener Comprehension," Speech Monographs, XIX (November, 1952), 249-258.

Bryant, Donald C., and Wallace, Karl R. Fundamentals of Public Speaking, New York: Appleton-Century Company, Inc., 1947.

Carter, Richard F. "Writing Controversial Stories for Comprehension," Journalism Quarterly, XXXII (Summer, 1955), 319-328.

Dale, Edgar, and Chall, Jeanne S. "A Formula for Predicting Readability," Educational Research Bulletin, XXVII (January, 1948), 11-20, 28.

_____. "A Formula for Predicting Readability: Instructions," Educational Research Bulletin, XXXVII (February, 1948), 37-54.

Flesch, Rudolph. "A New Readability Yardstick," Journal of Applied Psychology, XXXII (June, 1948), 221-233.

Gilkinson, Howard, Paulson, S.F., and Sikkink, D.E. "Effect of Order and Authority in an Argumentative Speech," Quarterly Journal of Speech, XL (April, 1954), 183-192.

Gray, William S., and Leary, Bernice E. What Makes a Book Readable. Chicago: The University of Chicago Press, 1935.

Gulley, Halbert E., and Berlo, David K. "Effects of Intercellular and Intracellular Speech Structure on Attitude Change and Learning," Speech Monographs, XXIII (November, 1956), 288-297.

Hamilton, Francis M. The Perceptual Factors in Reading. "Archives of Psychology," Vol. I, No. 9; New York: The Science Press, 1907.

Hockett, Charles F. Review of A Mathematical Theory of Communication, by Claude E. Shannon and Warren Weaver. Language, XXIX (1953), 69-92.

Hovland, Carl I., et al. The Order of Presentation in Persuasion. New York: Yale University Press, 1957.

Howes, Davis H. "The Definition and Measurement of Word Probability." Unpublished Ph.D dissertation, Dept. of Psychology, Harvard University, 1950.

_____, and Solomon, Richard L. "Visual Duration Threshold as a Function of Word-Probability," Journal of Experimental Psychology, XLI (June, 1951), 401-410.

Miller, George A. Language and Communication. New York: McGraw-Hill Company, Inc., 1951.

_____, "What is Information Measurement?" American Psychologist, VIII (January, 1953), 3-11.

_____, and Friedman, Elizabeth A. "The Reconstruction of Mutilated English Texts," Information and Control, I (September, 1957), 38-55.

Mills, Glen E. Composing the Speech. New York: Prentice-Hall, Inc., 1952.

Newman, Edwin B. "Computational Methods Useful in Analyzing a Series of Binary Data," American Journal of Psychology, LXIV (April, 1951), 252-262.

Osgood, Charles E., and Sebeok, Thomas A. (eds.). Psycholinguistics; A Summary of Theory and Research Problems. Supplement to Journal of Abnormal and Social Psychology, XLIX (October, 1954).

Peterson, Eleanor M. Aspects of Readability in the Social Studies. New York: Bureau of Publications, Teachers College, Columbia University, 1954.

Selfridge, Jeniffer. "Investigations into the Structure of Verbal Content." Unpublished Honors thesis, Dept. of Psychology, Radcliffe College, 1949.

Shannon, Claude E., and Weaver, Warren. The Mathematical Theory of Communication. Urbana: The University of Illinois Press, 1949.

Sponberg, Harold. "The Relative Effectiveness of Climax and Anti-Climax Order in an Argumentative Speech," Speech Monographs, XIII (No. 1, 1946) 35-44.

Taylor, Wilson L. "Application of 'Cloze' and Entropy Measures to the Study of Contextual Constraint in Samples of Continuous Prose." Unpublished Ph.D. dissertation, The University of Illinois, 1954. Summarized in Dissertation Abstracts, XV (March 1955) 464-465.

_____. "'Cloze' Readability Scores as Indices of Individual Differences in Comprehension and Aptitude," Journal of Applied Psychology, XLI (No. 1, 1957), 19-26.

_____. "Recent Developments in the Use of Cloze Procedure," Journalism Quarterly, XXXIII (Winter, 1956), 42-48.

APPENDIX A (Experimental message in outline form)

READABILITY

Something should be done to inform more people about the research that has been done on readability.

I. Many of those who could make the best use of readability research are unaware of its existence.

   A. Much of the research on readability is effectively hidden from teachers in the traditional disciplines.

      1. Much of the research on readability has been published in psychology journals.

      2. Many of the reports on readability research have been published in a very unreadable form and cluttered with mathematical calculations.

   B. Readers with limited ability are not encouraged to learn about readability.

      1. They cannot read the technical reports which are generally filled with syntactic S ynnastics and fifty cent words.

      2. Readability has generally been presented as a secret weapon designed especially for writers.

II. The knowledge now available about readability can be used to facilitate communication.

   A. Writers who apply the principles of readable writing find their writing to be more effective.

      1. A reading audience understands better, learns faster, and retains more from writing that is rather easy to understand.

      2. People generally enjoy reading more if it is not too difficult.

   B. Readers who understand the factors that influence readability are able to read more efficiently.

      1. Given a choice of materials, a reader who knows the principles of readability can, quickly and easily, select books within the range of his reading ability.

      2. If required to read a difficult book or article, readers find that knowing the source of the difficulty is a step toward overcoming it.

APPENDIX B

NAME_____ AGE _____ SEX_____

Year in School _____ Grade Point Average _____

Your presence here indicates that you are willing to act as a subject for an experiment. This experiment is part of a thesis that is being done in the College of Communication Arts. Your scores will not count on your grade, but the usefulness of this research is dependent on your doing your best.

In the following messages you are asked to fill in the blanks with the word that seems most appropriate to you. These messages have been carefully checked for mistakes, and if the right word is filled in, every sentence will be "good English" complete with punctuation. Each blank requires one word and only one word.

Start work on the next page and work on "Impromptu Speaking" until time is called. You will have ten minutes on message one. Do not go beyond the stop sign. Work as rapidly as possible, but pay close attention to every clue that might suggest the "right word" or eliminate a "wrong word." You should try to get through the whole message in the time allotted to take advantage of any easy blanks. If no word seems exactly right, GUESS.

Ten minutes will also be allowed on the second message.

IMPROMPTU SPEAKING

Impromptu speaking _____ often confused with extemporaneous _____, not only by students _____ also by dictionaries. It _____ inadequate to state that _____ speaking is done without _____. It is more accurate _____ say that it is _____ without specific preparation. An _____ speaker relies exclusively upon _____ general preparation. A speech _____ is given without either _____ or specific preparation, if _____ be possible, is not _____ listening to. Some learning _____ in speaking without specific _____ is extremely valuable. It _____ us for the otherwise _____ occasions when we are _____ to "make a few _____" without recourse to a _____ speech. There are undoubtedly _____ situations which lend themselves _____ well to impromptu speaking _____ even an inexperienced speaker _____ do quite well. Perhaps _____ remarks of another speaker _____ serve as a challenge _____ must be answered. For _____ less favorable situations, however, _____ is well to be _____ for the short notice _____. Look and listen carefully _____ that you can adapt _____ what is going on. _____ facility in the use _____ some simple patterns of _____, and

practice the control _____ stage fright. Despite all _____, the practice of impromptu _____ has limited usefulness. Any _____ who succumbs to the _____ rationalization that he can _____ his best without specific _____ is deluding himself. It _____ well to learn to _____ on one's feet, but _____ is more important to _____ to think before getting _____ to speak.

S T O P

(1) READABILITY

Something should _____ done to inform more _____ about the research that _____ been done on readability. _____ of those who could _____ the best use of _____ research are unaware of _____ existence. Much of the _____ on readability is effectively _____ from teachers in the _____ disciplines. Much of the _____ on readability has been _____ in psychology journals. Many _____ the reports on readability _____ have been published in _____ very unreadable form and _____ with mathematical calculations. Readers _____ limited ability are not _____ to learn about readability. _____ cannot read the technical _____ which are generally filled _____ syntactic gymnastics and fifty _____ words. Readability has generally _____ presented as a secret _____ designed especially for writers. _____ knowledge now available about _____ can be used to _____ communication. Writers who apply _____ principles of readable writing _____ their writing to be _____ effective. A reading audience _____ better, learns faster, and _____ more from writing that _____ rather easy to understand. _____ generally enjoy reading more _____ it is not too _____. Readers who understand

the _____ that influence readability are _____ to read more efficiently. _____ a choice of materials, _____ reader who knows the _____ of readability can, quickly _____ easily, select books within _____ range of his reading _____. If required to read _____ difficult book or article, _____ find that knowing the _____ of the difficulty is _____ step toward overcoming it.

For the first message, "Impromptu Speaking," mark an x on the following scales to represent your feeling about this experimental test.

Interesting ____:____:____:____:____Uninteresting

Easy ____:____:____:____:____Difficult

Useful ____:____:____:____:____Worthless

For the second message, "Readability," make an x on the following scales to represent your feeling about this experimental test.

Interesting ____:____:____:____:____ Uninteresting

Easy ____:____:____:____:____ Difficult

Useful ____:____:____:____:____ Worthless