SENTENCE, PHRASE, AND PIVOTAL WORD RECALL AS RELATED TO SENTENCE COMPLEXITY, RESPONSE MODE, AND PRACTICE

by

DIANA LEE MITCHELL

B.Ed. University of British Columbia, 1967
M.A. University of British Columbia, 1968

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF EDUCATION

in the Department

of

Educational Psychology

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

July, 1974

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Education Psychology

The University of British Columbia
Vancouver 8, Canada

Date Oct 9/74

ABSTRACT

The effects on sentence, phrase, and pivotal word recall of variations in mean depth, type, and length of sentences, response mode, and practice were studied with one hundred twenty-eight Grade Eleven students. The results provide substantial support for Martin and Roberts' (1966) listening-reproducing model of sentence processing. Verbatim recall of low mean depth sentences exceeded that of high mean depth sentences; pivotal words from phrases of the greatest mean depth were harder to recall than were their lower mean depth counterparts; and the analysis of phrase dependency measures indicated that left-to-right binary processing was taking place during the encoding and decoding of responses.

In addition, the results provided partial confirmation of the notion of coding by transformational tags. Verbatim sentence recall was affected by the stimulus sentence's transformational type, but not in the predicted direction. Errors in sentence recall, however, were as was predicted: K > N, P, Q > NP, NQ, PQ > NPQ. Further, declarative sentences were recalled correctly more often than were interrogative, as was predicted from the coding hypothesis, but active and passive sentences were equally well recalled, and negative sentences were recalled better than affirmative sentences, neither of which result can be accounted for by the coding hypothesis. Mood, voice, and modality changes were found to have various effects on the recall of phrases and pivotal items, but often these effects were related to the lexical density of specific phrases. Pivotal words were observed to have performed an important function in early recall particularly,

and this function was explained in terms of the memory for "gist" hypothesis (Reid, 1974).

Sentence length interacted with mean depth and lexical density, a result which was felt to be an artifact resulting from incomplete orthogonality of these variables.

The written response mode facilitated recall generally, as did practice, with practice resulting most noticeably in improved recall of all pivotal words, and of low mean depth and long sentences which were written rather than spoken.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION: STATEMENT OF PROBLEM

A. Scope of the Present Study

The present study is concerned with investigating some selected variables which might affect student performance in a task involving a speaker and a listener who is required to recall what he has heard. Of specific interest is the question of which characteristics of spoken sentences, and which response demands placed on the learner will affect his sentence recall performance. The characteristics of the sentences which are of particular interest in the present study are sentence complexity and sentence length. The response demands examined are the mode in which the student responds with what he has heard and repeated practice of the same task.

It seems conceivable that the amount and accuracy of the information recalled by the learner could vary according to the surface structure of the sentence he hears. For example, the more a subject-noun is pre-modified, and consequently, the further it is removed from the head noun, the harder it may be to correctly recall. On the other hand, combinations of mood (in this case declarative or interrogative), voice (active or passive) and modality (affirmative or negative), comprising the many sentence types may also differentially affect the student's recall. Certainly one might anticipate that a long sentence will be harder to recall than will a shorter version, and that, with practice, the student's recall of the message will improve. Also of interest is the question of whether saying what one has heard affects sentence recall performance differently from writing one's response.

Some of these variables have been examined in previous studies, both alone, and in combination (i.e. surface structure, length, and mood, voice, modality characteristics of the sentence stimuli), although their various roles, relative to influencing sentence recall performance, are still unclear. The roles played by differential response modes and repeated practice have not been examined within the context of a sentence recall task, either alone or relative to the other variables mentioned, but are of interest in the present study.

B. Practical and Theoretical Significance

There have been relatively few studies which have focussed specifically on the type of communication examined in the present study. Considering that students in school are frequently exposed to communication involving orally presented messages and are often expected to reiterate them orally or write what they have heard, this is surprising. It is hoped, therefore, that the present study will provide data, the implications of which might have some significance for practitioners such as those who design instructional materials and activities to be used in sentence recall tasks.

The possible theoretical significance of the present study was felt to be related to two hypotheses which deal with the way people remember sentences. Miller (1962) proposed a taxonomy of sentences based on an early version of the transformational grammar (Chomsky, 1957). Implied by the taxonomy is that a speaker-listener will differentially recall various sentence forms, dependent upon each sentence's transformational relationship to its simple, affirmative, declarative form (here called kernel, or K). It is hoped that the data from the present study might shed further light on the issue of the psychological reality of sentence complexity, as measured by transformational history, when sentence recall performance is demanded of the subject. Specifically, the significance of transformational sentence type might be expected to be reflected in the numbers of sentences of each form correctly recalled, and in the numbers and kinds of transformational errors made. Transformational errors are expected to be particularly revealing of some of the possible mechanisms involved in the sentence recall process.

Martin and Roberts' (1966) model of listening to and reproducing sentences, which was originally based on Yngve's (1960) speech production model, is, in many ways, pertinent to the situation examined in the present study. They propose that the greater the surface phrase structure complexity of a sentence, the less likely it is to be recalled correctly. Implied by their model is an equivalence of processes involved in both the listening and reproduction phases of the task. Upon closer examination, however, it would seem that Martin and Roberts actually focus on the listening aspect of the sentence memory process. The present study attempts, therefore, to examine what indications exist at three levels of analysis: sentence, phrase, and pivotal item, of phenomena which might define listening and reproduction phases of sentence recall. Nouns and verbs are classified as pivotal items in the present study.

The psychological significance of surface structure phenomena is expected to be reflected particularly in the numbers of sentences correctly recalled of low versus high indices of surface structure complexity. It is also expected to be reflected in the recall dependency measures between phrases, and, as with the transformational error measure, be revealing of some of the mechanisms involved in sentence recall processes.

Past studies of sentence recall performance have often encountered problems. Frequently these relate to the confounding of sentence stimulus length, surface structure complexity, and complexity as measured by transformational type. The variables are felt to be controlled in the present study.

In summary, the general objective of the present study is to investigate the relationship between sentence recall and sentence complexity. In particular, the concern is to examine selected variables which might affect the sentence recall performance of a speaker-listener: stimulus complexity and specified response variables. The specific aims are to:

1) evaluate the validity each of the sentence variables (mean depth, transformational or sentence type, and length of sentences) has as a predictor of sentence recall performance
2) assess the comparative effects on sentence recall of oral versus written response modes, and
3) investigate the effects of practice on recall of sentences.

C. Psychological Orientation

The orientation for the present study is a Process one, rather than Content or Associationist (cf. Reber, 1973), with certain assumptions being made about the speaker-listener who is involved in a sentence memory task. These are that:

1) he has an innate general cognitive mechanism for processing information in his environment (Slobin, 1966; 1971), which is assumed to facilitate his acquisition of a set of generative grammatical rules;

2) he possesses linguistic competence which far exceeds his performance, but which will be reflected in performance, to a greater or a lesser degree, depending upon learning conditions.

It follows from this that the emphasis in the thesis is primarily <u>psycho</u>linguistic rather than psycho<u>linguistic</u> (Reber, 1973).

CHAPTER II

ANALYSIS OF RESEARCH PROBLEMS

A. Background of Two Sentence Complexity Hypotheses Dealing with Sentence Recall

For the purpose of psychological studies of sentence complexity Martin and Roberts (1966) defined the mean depth of a sentence as an index of its complexity on the basis of Yngve's (1960) model for sentence production, which represents constituent structure analysis of language. Immediate constituent analysis involves subdivision of a linguistic construction into successively smaller constituents. For example, the sentence "The dogs tease the cat", would be subdivided into the immediate constituents, "the dogs" and "tease the cat", where the first is a noun phrase (fulfilling the function of subject of the sentence) and the second is a verb phrase (fulfilling the function of predicate). The noun phrase can be subdivided into its immediate constituents, a definite article "the" and a noun "dogs", while the verb phrase can be subdivided into its immediate constituents, a verb "tease" and another noun phrase "the cat" which is the direct object of the verb. Its immediate constituents are a definite article "the" and a noun "cat".

On the morphological level, the noun "dogs" can be further subdivided into "dog" and "s". Thus, immediate constituent analysis allows the sentence "The dogs tease the cat", to be analyzed as "The + dog + s + tease + the + cat" or as "article + noun (plural) + verb (present) + article + noun (singular)." Yngve (1960) uses symbols (such as "S", "NP", "VP", "N", "V", "Aux", "T") for form classes (sentence, noun phrase, verb phrase, noun, verb, auxiliary, article).

The various categories of constituents, as fefined by a phrase structure grammar and used in the previous analysis of "The dogs tease the cat" (i.e. its syntactic components), can be labelled as follows:

1.	S (Sentence)________	NP + VP
2.	NP________________	T + N
3.	VP________________	V + NP
4.	T_________________	the
5.	N_________________	dogs, cat, etc.
6.	V	tease, kick, etc.

By further applying the grammar we obtain the derivations as shown in Figure 1a. The derivation can also be represented in a tree diagram as shown in Figure 1b.

Yngve's model (1960) of speech production comprises a grammar, with all the rules of a language, and a mechanism which applies these rules and produces sentences of a language. The mechanism applies rules by expanding the constructions into their constituents (as in Figure 1) always beginning with the left-most constituent of a construction and working left-to-right.

One obvious characteristic of the surface structure of a sentence, as is demonstrated in Figure 1b, is the hierarchical nature of the structure. Implicit in Yngve's model is the notion that this hierarchical organization provides a kind of "plan of execution" (Miller, Galanter, and Pribram, 1960) for the production of a sentence. It follows from this that the more branches and levels there are to a surface phrase tree, the more involved is its organization, and the greater the amount of grammatical information that must be stored in memory.

S

NP + VP

T + N + VP

T + N + V + NP

T + N + V + T + N

the + N + V + T + N

the + dogs + V + T + N

the + dogs + tease + T + N

the + dogs + tease + the + N

the + dogs + tease + the + cat

Fig. 1a Steps in producing the sentence "The dogs tease the cat."

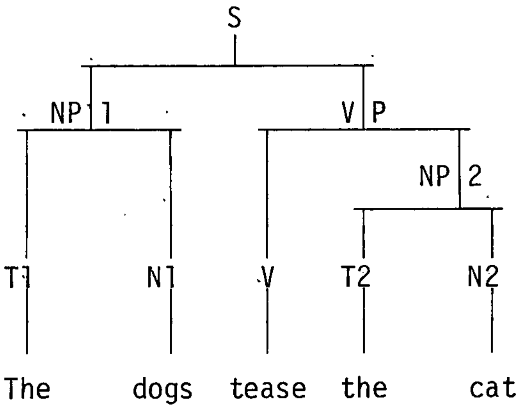


Fig. 1b Phrase marker of "The dogs tease the cat."

Yngve (1960) suggests that the load on the mechanism's temporary memory at any one time can be measured in terms of a metric defined on the basis of a sentence's phrase marker. First, the branches of each non-terminal node are numbered from right-to-left, beginning with 0. The depth of each word (denoted as d), viz., each terminal node, is then computed by adding up the numbers along the left-branches leading to the word. The depth of a sentence (denoted as D), is then defined as the maximum of the depths of its words, and indicates the amount of temporary storage needed for the sentence. This concept of sentence complexity is based on the premise that a speaker who wishes to communicate must do so within the restrictions of the grammar. For example, if he intends to say "When the president spoke, the people listened," he incurs certain commitments upon saying "when" which must be fulfilled in order to speak a grammatical sentence. These are, first to complete the dependent clause and second to provide an independent clause. Having next said "the", the speaker incurs three grammatical commitments; (1) to complete the independent clause ("When the president spoke"), (2) to complete the dependent clause ("the president spoke", and (3) to complete the sentence with a dependent clause ("the people listened"). Figure 2 shows a constituent structure analysis of this sentence "When the president spoke, the people listened." The number of grammatical commitments incurred by a speaker upon saying each word in the sentence defines the surface structure complexity of that word. The numbers under each word in Figure 2 express the complexity or depth (d) of the word in the sentence, "When the president spoke, the people listened." Beginning with Sentence (S), there are two "1's" on the left-branches leading to the word "when". The first

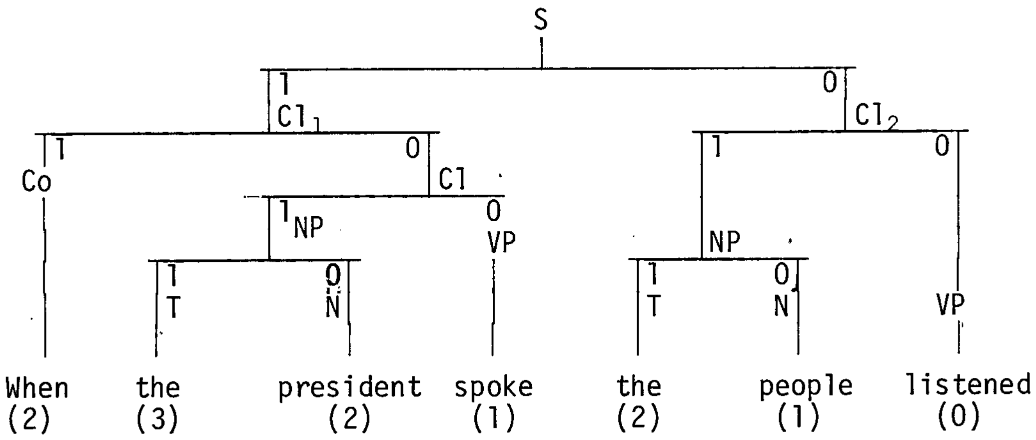


Fig. 2. Phrase marker of "When the president spoke the people listened."

of these occurs where S branches to the left for Clause 1 (Cl_1), and the second is where Cl_1 branches to the left to the coordinator (Co). Thus the word "when" has a depth of two.

The depths of the words in a sentence reflect the complexity of that sentence in terms of a phrase structure analysis of the immediate constituents. Yngve (1960) claims that a speaker produces the constituents of a sentence in the manner of a constituent structure tree, that is, from top to bottom and left to right. He further suggests that the number of grammatical commitments incurred at any one time by a speaker will generally not exceed 7 ± 2, the presumed upper limit of the short-term memory span (Miller, 1956). In support of this Yngve (1960) observes that when a speaker begins a sentence in which the depth exceeds 7 ± 2, he is likely to hesitate, stumble at the point of greatest depth, and perhaps forget what he was trying to say. Because of the memory span limitation, Yngve (1960) suggests that English speakers have developed preferred constructions which maintain the expressive power of the language, but which place a manageable load on memory. He contends that the effect of the depth limitation should be easily observable in a grammar of English. By implication from his model, the depth limitation should also be observable in the sentences produced by the speaker of English.

Martin and Roberts (1966, p. 211) attempted to extend Yngve's speech production model (1960) to include speech reproduction.[1] Since their adaptation of Yngve's model is of central importance to the present study, it merits a further examination.

---

1 Dr. Yngve concurs with the author on this point (personal communication, 1972).

1. Martin and Roberts' Mean Depth Hypothesis

Martin and Roberts (1966) examined a communication situation in which speaker A produces a sentence which is heard by listener B who then becomes speaker B, who in turn reiterates what he has heard. They propose that as speaker A utters a sentence, listener B encodes what he hears and forms expectations of what is to come. These expectations are determined by speaker A's choice of word class, manipulation of pitch, stress, and juncture, and the listener's own grammatical habit structures (p. 211-212). Martin and Roberts (1966) contend that the listener's expectations, or anticipatory responses are largely what is placed in memory. The more expectations or anticipatory responses the listener makes, the more load is placed on memory, and the less likely it is that he will correctly recall the sentence when required to reproduce it. They further contend that "... for each word in a given sentence, the number of expectations elicited in the listener is the same as the number of commitments incurred by the speaker" (p. 212). They suggest that the complexity of a sentence is more accurately given by the mean of the set of depths assigned to the words of that sentence following a constituent structure analysis (denoted as $\bar{D}$) of the sentence "When the president spoke, the people listened" (see Figure 2) is 2 + 3 + 2 + 1 + 1 + 0 = 11/7 = 1.57, whereas Yngve's depth (D) of the sentence is 3. They hypothesize that the likelihood of recall of a sentence is inversely related to the mean depth of that sentence.

To test the hypothesis, Martin and Roberts (1966) employed a 2 x 6 factorial design involving two levels of sentence mean depth as between-subjects (_Ss_) factor and six transformational types as within-_Ss_. All sentences were seven words long. The task set for Ss was a free recall task wherein the same six sentences (one of each type and all of one $\bar{D}$) were read to each S on each of six trials. The sentence types included were:

1. simple active declarative or kernel (K)
2. negative (N)
3. passive (P)
4. passive truncated (P_T - These are passive sentences in which the agent is no specified, i.e. "The bird was eaten.")
5. passive-negative (PN)
6. passive-truncated-negative (P_TN).

On each trial the S wrote his responses after having heard the set of six sentences. The number of sentences recalled verbatim was the dependent measure.

Martin and Roberts (1966, p. 216) reported that, structural complexity as indexed by the sentence mean depth is a definite factor in sentence retention; and that when sentence complexity and sentence length are controlled, the role of sentence kind in explaining recall behavior becomes marginal.

They also showed that the results reported earlier by Miller (1962) and Mehler (1963), which were at the time interpreted as being supportive of a transformational model of sentence reproduction, were also in accordance with the mean depth hypothesis. They interpreted the results as confirming what might be predicted on the basis of the model of their psychology of listening to and reproducing ordinary English sentences. A close examination of the study by Martin and Roberts (1966), however, leads to the observation that only a partial confirmation was provided, and that two assumptions which are implicit in their reformulation of Yngve's speech production model were not fully tested. First, in their application of Yngve's model to the case of sentence reproduction, Martin and Roberts (1966) proceeded on the assumption that the processes involved in listening to English sentences are equivalent to those involved in reproducing them. Implicit also in their reformulation is the assumption that sentence production, as Yngve describes it, involves similar processes to sentence reproduction.

The first of these assumptions does not appear to take into account what seems to be an important distinction between the speaker and the listener, that is: the speaker begins with a speech intention, or plan, which is unknown to the listener. The extent to which the listener is handicapped by not knowing the speaker's plan might be expected to be reflected differently during initial recall trials in a reproduction situation than in later trials. A way to evaluate the extent of this problem would be to compare verbatim recall of sentences and phrases in later trials as compared with their verbatim recall in early trials of the task. As familiarity with the material increases, the extent of the handicap of the listener ought to lessen.

A test of the second assumption might include not only verbatim recall of the sentence, and phrases, but also indicate the dependency between constituents and the location of specific kinds of errors. If sentence processing follows the binary rules, as was originally suggested by Yngve (1960) and subsequently by Martin and Roberts, one might expect that correct recall of a predicate verb phrase will be dependent upon correct recall of the subject noun phrase. One might also expect that correct recall of an object noun phrase would be dependent upon correct recall of the predicate verb phrase. It is assumed that, having correctly recalled the verb phrase, the S's memory load reduction would facilitate his correct recall of the object noun phrase. This dependency relationship is therefore expected to be stronger, because of the memory-load reduction, than is the first relationship mentioned. One would predict, however, that the dependency between subject and object noun phrases would not be so strong as the other two dependency relationships just described, but would be consistent with Martin and Roberts' reformulation of Yngve's model. If confirmed they would add strength and amplification to Martin and Roberts' model of psychology of listening to and reproducing ordinary English sentences. Verbatim recall of the sentence stimuli and constituents of the sentences, as well as phrase-dependency measures have all been included in the present study in the hope that they will be more revealing of the influence of surface structure on sentence reproduction performance than is verbatim recall of the sentence alone.

Included in the present study also, as in Martin and Roberts' study (1966), is the variable of sentence complexity as measured by transformational history (Miller, 1962). Martin and Roberts' finding of the significance of mean depth in sentence reproduction necessarily raised questions about the relative generality of Miller's hypothesis regarding the critical role of sentence type in sentence recall tasks.

2. Sentence Type

Miller (1962) suggested that sentences of different syntactic forms might differ also in the ease with which they can be processed and subsequently recalled. He suggested that relationships exist among sentences of a "family" and based his assumptions on Chomsky's early hypothesis regarding how one stores and retrieves a sentence, which puts central importance on the simple, affirmative, declarative sentence type (kernel). It was proposed (Chomsky, 1957) that underlying each kernel sentence is a string of symbols representing the basic structure of the sentence. By applying transformational rules to the underlying base structure, a variant of the kernel can be created, differing from the kernel in mood, voice, and modality. It was, Chomsky emphasized, from kernel sentences that "... more complex sentences of real life are formed by transformational development" (1957, p. 92). This would imply that, other things being equal, a sentence which is farther away from the K-terminal string, in simple terms of the number of transformational rules which have been applied, could be termed as being more complex than the kernel.

A sentence family, as Miller (1962) conceived of it, consists of the eight sentence types derived from the combinations of mood (in this case, declarative and interrogative), voice (active and passive), and modality (affirmative or negative) transformations of the K-terminal string. In other words, the number of transformations needed to derive the surface structures from the basic structures proposed to underlie them, determines the complexity of the outcome sentence (Chomsky, 1957; Miller, 1962; Miller and Chomsky, 1963). Figure 3 indicates the relationships Miller (1962) suggested exist between members of a sentence family.

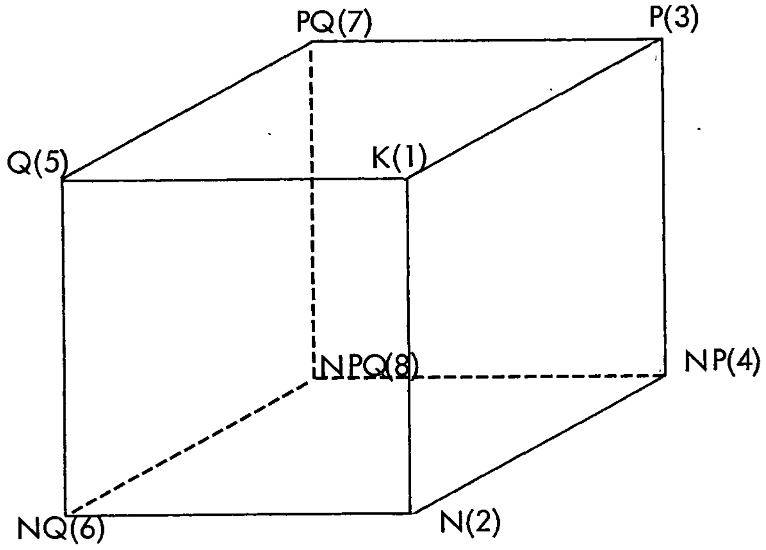


Fig. 3. A cubic representation of eight types of sentences formed by transforming a kernel form with respect to the mood, voice, and modality of a sentence.

Miller (1962) hypothesized "...that what people remember is the kernel sentence, but when you ask them to recite the original sentence exactly, they supplement their memory of the kernel with a footnote about the syntactic structure" (1962, p. 760). This hypothesis is generally referred to as the "coding hypothesis". From its early version it would be predicted that kernel sentences (K) would be easier to recall correctly than passives (P) negatives (N) and questions (Q). Negative passives (NP), passive questions (PQ), and negative questions (NQ) would be the next hardest to recall, and negative passive questions (NPQ) the hardest of all to recall correctly. Further, the implication is that when recall errors occur, they will be of a transformationally simpler type than the stimulus sentence, rather than of a more complex type.

Katz and Postal (1964) argue that N, Q (interrogative), and P sentences differ at the deep, basic structural level from the K, and not merely in terms of transformational history. The increased importance placed by them on deep structure is mirrored in Chomsky's (1965) amended theorizing and supported by others (Clark, 1969; Wanner, 1968). However, the notion of deep structure has become increasingly abstract, and it appears that linguists do not yet agree on an analysis of the underlying structures which will consistently provide an uncontested index of the deep structure complexity of any set of sentences (Bever, 1968; Dever, 1971). To date, it would appear that while the deep structure notion of sentence complexity has utility and linguistic appeal, its pragmatic value for psychology is still uncertain (Dever, 1971).

The two measures of sentence complexity which continue to be studied by psycholinguists with respect to the measures' predictive validity in a number of situations are mean depth (derived from surface structure) and sentence type (derived from transformational history). It is these two complexity measures which receive the major focus in the present study.

B. Examination of the Empirical Validity of the Mean Depth and Coding Hypotheses for Predicting Sentence Recall Performance

What follows is a critical review of the research literature bearing on these measures which leads to the following tentative propositions:

1. Surface structure phenomena influence sentence processing.
2. Mean depth is positively associated with sentence memory, but the nature of this association is not clear.
3. It is not clear yet whether or not sentence type (Miller, 1962) has a critical influence on sentence recall.
4. Some of the evidence used to support the notion that memory representation of sentences is not captured by surface phrase structure (Perfetti, 1969 a, b; Perfetti and Goodman, 1971; Fodor, Bever, and Garrett, 1974) has probably been a misinterpretation of the data. There is a need for continued, controlled study of the variables involved in sentence memory.

A number of studies have demonstrated the influence of surface structure on production and recall of sentences (Martin, 1967; Herriot, 1968) Wang, 1970; Levelt, 1970; James, Thompson, and Baldwin, 1973). They support Yngve's notion that the order-sensitive surface structure of a sentence serves as a memory-facilitating organizational framework for sentence processing. This concept is central to Martin and Roberts' (1966) reformulation of Yngve's model. The supportive literature is summarized here, therefore, as the background for a subsequent review of studies which deal with other specific aspects of Martin and Roberts' reformulation.

Martin (1967) reported a substantial influence of phrase structure on the decoding responses of Ss who were required to exactly repeat the original speaker's spontaneous, one-sentence descriptions of Thematic Aperception Test Pictures. It was clear that Ss paused most often in their reiteration of what they had heard at points where there were the greatest number of choices of lexical items. These points were usually where content words (noun, verb, adverb, adjactive) occurred, reflecting the Ss' use of major phrase structure constituents in their perceptual organization of the material.

Herriot (1968) provided evidence for the psychological reality of surface structure in a study requiring Ss to "translate" nonsense words into English words in grammatically structured nonsense sentences which were either phrase suffixed (PS) or phrase embedded (PE). By Yngve's analysis, PE constructions tax immediate memory more than do PS, as the former causes the predicate phrase to be postponed, thus adding to the depth of the first noun phrase. Subjects found it much harder to translate PE sentences (p .025) and made more overt errors in translation of PE ($p < .025$), which results support this aspect of Yngve's model.

Wang (1970) had Ss rate five 29-word sentences of varying complexity (according to transformational history, deep structure, surface structure) as to their comprehensibility. He provided few details in his paper regarding the sentence stimuli, but reported that Ss reliably judged sentences as being difficult to comprehend when the surface structure was complex, as measured by depth, rather than if base structure or transformational history were complex.

Levelt (1970), studying correct recall of noise-masked sentences, suggested that as one listens to sentences, one "chunks" the input into perceptual categories paralleling large constituents in surface phrase structure. Jarvella (1971) confirmed this in a study of response patterns when connected discourse was interrupted and prompted recall of immediately preceding clauses and sentences was compared with unprompted recall. Again, Ss clearly processed the verbal input into "chunks" similar to the large constituents in surface phrase structure.

James, Thompson, and Baldwin (1973) investigated the role of reconstructive processes in the recall of K and P sentences. They concluded that at least when Ss are instructed to recall the sentence stimuli verbatim, surface structure was an aid to memory.

From these studies, it can be seen quite clearly that surface structure has been demonstrated to play a significant role in the processing of sentences. Of particular interest in the present study are those studies which bear specifically on mean depth and sentence type and their roles in predicting responses in a sentence memory task.

1. Mean Depth, Length of Sentence, and the Depth Hypothesis

Support for the mean depth hypothesis comes from a number of sources. Martin, Roberts, and Collins (1968) by varying three variables, that is; their Ss recalled orally presented active and passive sentences of mean depth 1.00 and 1.86 at 0-, 10-, 20-, and 40-second retention intervals, found mean depth to reliably predict Ss' oral responses. In addition, they found the mean depth to interact with the sentence type, low mean depth passive and high mean depth active sentences being recalled more often than others. This differed from Martin and Roberts' finding (1966) that while K sentences were hardest to recall, there was no consistent effect on recall of their five other sentence types. This was previously interpreted by Martin and Roberts (1966) to be a strong indication that sentence complexity as defined by transformational history was inadequate for predicting recall performance in a sentence memory task.

Martin, Roberts, and Collins' study (1968) seems, however, to indicate that both mean depth and the interaction of mean depth and voice (active vs. passive) predict recall performance. A close examination of stimulus sentences used in Martin and Roberts (1966) and Martin, et al. (1968) reveals some uncontrolled aspects in them which seem to cast some doubts on their original conclusion about the predictive validity of the coding hypothesis. For example, with the K sentences of mean depth 1.29, "she" was the subject of sentences twice, "they" twice and "we" three times. In the P sentences of mean depth 1.29, the impersonal pronoun "it" was the grammatical subject four times, and for the remaining instances, subjects are "clues" and "assignments". In view of the uncontrolled nature of the grammatical subjects across the variations in voice, one is tempted to suggest an alternative way to account for the results which is contrary to what might have been predicted from the coding hypothesis. It could be that the Ss might have acted as if there were one less lexical item to learn, over the trials in "It was slowly lifted by the crane," with "It" occupying minimal memory space, whereas "she" in "She will certainly visit her two sisters" carries more information by itself as well as necessitating a specific form of the possessive pronoun in the object noun phrase. While the two exemplar sentences have the same surface structure complexity and the same mean depth, their derivations might be seen by a transformational grammarian to be sufficiently different as to have contributed to the depressed effect of the K sentences on Ss' recall in Martin and Roberts' study.

Mitchell (1968) also found a significant effect of mean depth in a study, replicating Martin and Roberts' finding with both young children and adult Ss. There was, however, no mean depth by type interaction, different from that found by Martin, et al. (1968), but Negative sentences and Negative Passive sentances were correctly recalled by all Ss more often than were K or P sentences.

Roberts (1968), employed seven-word active and passive sentences of the mean depth ($\bar{D}$) 1.14 and 1.86, and four degrees of association between logical subjects and objects. The results were that low $\bar{D}$ sentences were correctly recalled most often over-all and that there was no difference in recall pattern of active and passive sentences, except in the lower $\bar{D}$ condition where passives were recalled better than actives. This latter result partially supports the findings of Martin, et al. (1968).

In the first of two experiments examining the relative predictive validity of the mean depth and the coding hypothesis, Wright's (1969) data appeared to support the mean depth hypothesis. Unfortunately, however, the mean depth was confounded with sentence length. In addition, the questionable handling of the data leads to serious doubts about these results. Wright's analysis was of the numbers of words correctly recalled from a list which was presented after each stimulus sentence. If a S correctly recalled fewer than two-thirds of the stimulus sentences, his scores for the recall of the word lists were excluded from the study. Wright (1969) rationalized this by saying that the S who could correctly recall no more than two-thirds of the sentences had not understood the instructions. One-sixth of the original data was deleted using this technique. No analysis of the sentence recall data was provided, resulting in a loss of information from which some of the cognitive processes in this task might have been studied. The conclusion must be that this study can be seen to provide only weak support for the predictive validity of the mean depth hypothesis in a recall task.

Wearing (1970) examined correct recall of 11.8- to 13.5-word sentences. The Ss in the recall task more often correctly recalled sentences of $\bar{D}$ 1.40 than sentences of similar lengths and $\bar{D}$ 1.69. He found, as did Martin and Roberts (1966) and Mitchell (1968), that whether a sentence was active or passive made no difference in recall. Wearing (1972), using cued recall and the sentences from his previous study (1970) confirmed his earlier findings of a significant mean depth effect.

Bacharach and Kellas (1971) reported results from two experiments, the second of which involved a recall task in which Ss heard ten 6-word sentences of either K, P, or P_T type followed by 0-, 30-, or 60-second filled retention intervals, and wrote their responses after each interval. Responses were marked correct if the lexical items were recalled correctly and the voice was unchanged. The time taken to the completion of written responses was a second dependent variable. When the correct recall measure was examined, only retention interval was significant (p .001); whereas, when the time of response was examined, both retention interval and sentence type were significant (p .001 and .01, respectively).

However, Bacharach and Kellas (1971) incorrectly stated (p. 174) that their stimulus sentences were equated for d and $\bar{D}$. In fact, $\bar{D}$ of their K sentences is 1.00 (d = 2, 1, 1, 1, 1, 0) and $\bar{D}$ of their P and P_T sentences is 1.17 (d = 2, 1, 2, 1, 1, 0). Consequently the mean depth would have predicted the results measured by the time to process; that is, K $<$ P or P_T. One wonders whether the results, as measured by correct recall, would have differed if a more stringent correct recall definition had been used, viz., verbatim recall of the stimulus sentence. As it is, by Bacharach and Kellas' (1971) definition of the criterion measure, the written response, "Messages were dispatched by important governments," of $\bar{D}$ = 1.00 (d = 1, 2, 1, 1, 1, 0) woul have been marked as a correct response to the stimulus, "Important messages were dispatched by government," of d = 2, 1, 2, 1, 1, 0, and $\bar{D}$ = 1.17. While the lexical items are the same, as is the voice, this response would be classified as an error of the same type as the presented sentence by the more stringent definition of correct recall. Since the stringent score of verbatim recall, when evaluating the predictive validity of both surface and base structures, was not used as the criterion, the conclusions based on correct recall of the lexical items in Bacharach and Kellas' study (1972) can be taken as only weak support for the mean depth hypothesis.

Importantly, several other studies appear to have failed to confirm the mean depth hypothesis. Perfetti (1969a) first studied the effect on correct recall of high and low mean depth (average $\bar{D}$ = 1.78 and 1.35, respectively) twelve to fourteen-word active and passive sentences which had either a dependent clause located at the beginning or at the end. Subjects heard the sentences one at a time, followed by a list of ten digits, and were asked to write the sum of the digits and then the sentence. One trial per sentence was given, and the dependent measure was verbatim sentence recall of sentences. Only one-fourth of the responses were verbatim imitations and no report was made of the discarded, incorrect responses. Higher mean depth sentences with the dependent clauses at the end were recalled correctly much more often than were lower mean depth sentences with the dependent clauses in the initial position. High mean depth sentences were generally recalled better than were low mean depth sentences. Low mean depth passive sentences were recalled better than their low mean depth counterparts but the reverse effect was found when sentences were of a high mean depth.

From Yngve's model the predicted outcome would be the opposite of these results, since dependent clauses which occur at the beginning of the sentence are presumed to load up short-term memory. By the time the S has correctly decoded the initial dependent clause, particularly in such long sentences as were used by Perfetti (1969a), he may have forgotten all or part of the remainder of the sentence. Perfetti felt (p. 102) that this reversal may have resulted from large sampling errors consequent upon the initial formulation of experimental groups with "better performing" Ss being overly represented in the group which received high mean depth sentences. He also proposed that the poor overall recall of sentences might have been due to the length of the sentences.

A second experiment was performed with high and low mean depth (average $\bar{D}$ = 1.90 and 1.07 respectively) ten word sentences, no examples of which were reported in Perfetti's (1969a) paper. Perfetti found no mean depth effect and that recall was "highly negatively related to the number of lexical ... words and the number of words per sentence ..." (p. 104), but no statistical support for this statement was reported. By virtue of his own assessment of Experiment I, and the lack of supportive data in Experiment II, both would seem, at best, to be ambiguous with respect to their evaluation of the mean depth hypothesis.

In a follow-up study Perfetti (1969b) appeared to have attempted to test independently two possible factors in sentence retention: mean phrase structure depth and lexical density of a sentence. Lexical words were defined as being: nouns, verbs, adjactives, adverbs, and adjective and adverb verbals. Grammatical words comprised: determiners, copulas, pronouns, conjunctions, and verb auxiliaries. Lexical density was defined as $\frac{\text{number of lexical items}}{\text{number of words in sentence}}$. Ten word sentences averaging $\bar{D}$ = 1.17 and 1.78, and of lexical densities of 0.50 and 0.70 were used. No mean depth effect was found, but the author again felt that this could have been attributed to uncontrolled variations, this time in the stimulus sentences. In an attempt to correct for this, and using twelve word sentences of average $\bar{D}$ = 1.87 and 1.23 and lexical densities averaging 0.70 and 0.44, he found _Ss_' correct recall of sentences was best predicted by lexical density.

In a third set of studies, Perfetti and Goodman (1971) attempted to test the validity of sentence depth (D= 7 and 2) for predicting the correct recall of long sentences (14 to 19 words). Apparently, the sentences included a considerable amount of uncontrolled variability resulting in a significant depth by sentence list interaction. One-half of the sentences was recalled best in their high depth version, and the other half in their low depth version. Again, because of methodological inconsistencies, these results must be viewed cautiously.

An overview of Perfetti's (1969 and b) and Perfetti and Goodman's (1971) studies leads to the conclusion, however, that the relationship between sentence mean depth and sentence length merits further study. From the mean depth hypothesis one would predict that low mean depth sentences would be correctly recalled more often than high mean depth sentences, irrespective of their length. This prediction has not yet been empirically substantiated.

An overview of all the studies reviewed here provides the following conclusions:

a. There seems to be some evidence which supports the mean depth hypothesis (Martin, 1967; Herriot, 1968; Martin, Roberts, and Collins, 1968; Mitchell, 1968; Roberts, 1968; Wright, 1969; Wang, 1970; Wearing, 1970; Bacharach and Kellas, 1971; Wearing, 1972; James, Thompson, and Baldwin, 1973).

b. The major study evaluating the mean depth hypothesis (Martin and Roberts, 1966) suffers from certain methodological difficulties and therefore requires replication.

c. Without exception, the published studies offer no explicit information about recall error patterns and consequently eliminate the opportunity for insight into many of the processes which must be going on during the recall task.

d. The evidence from the few studies which appear to disconfirm the depth or mean depth hypotheses must be viewed as tentative in that they are methodologically weak (Perfetti, 1969a and b; Perfetti and Goodman, 1971).

e. The relationship of mean depth to sentence length requires relatively unambiguous examination.

2. Types of Sentences and the Coding Hypothesis

The studies reviewed here reflect the impact on recall of varying sentence types, and exclude data derived from studies of recognition memory, verification time, sentence completion, reaction time in motor response task, etc.

Fillenbaum (1970, p. 232) makes the point that the "...way in which a sentence is analyzed and stored in a memorial task of the sort employed ..." by Clifton and Odom (1966), that is, a recognition task, "...may be different in important respects from the way in which it is processed in a task which requires full understanding of the sentence, and discrimination between it and other closely related sentences" (p. 232). The same can be presumed about the ways in which sentences are analyzed, stored, and decoded in the various tasks mentioned above, although there is some disagreement about this. Tulving and Thomson (1971, p. 123) assert that there is little to support the notion that "...there is something inherently different about processes of recall and recognition." Underwood (1972) takes the opposite view, in a sense, supporting the "different task, different process" approach presented by Fillenbaum.

There is no cohesive theory of memory which describes the processes and variables involved in all these various tasks. Postman (1972) states that "...I see as the most pressing problems for the organization theorist the identification of functional units of memory and the identification of the retrieval cues that make these units accessible in recall." Much of what is known about memory for verbal information comes from studies which employ word lists, CVC trigrams, nonsense syllables, etc. Only recently has some progress been made toward developing theories about sentence memory (Cermak, 1972, Melton and Martin, 1972; Tulving and Donaldson, 1972). In order to keep a clear focus, then, on the topic of recall of sentences, only those studies dealing with this aspect of memory are reviewed here. Curiously, many of these studies also privide a considerable amount of information, either directly or indirectly regarding the effect of mean depth.

Miller (1962) and Mehler (1963) performed similar tests of the coding hypothesis except that Miller's sentences were presented auditorally in a prompted recall situation. Ss memorized lists of sentences, each representing a Kernel (K) and its seven transformations. Correct recall of the sentences was the dependent variable. The results were taken as support for the coding hypothesis, that is, K sentences were recalled best followed by sentences of one step removed, in terms of transformational "tags" (N, Q, P), then by sentences of two steps and by those of three steps removed.

The coding hypothesis appears to have received some support from Mehler's (1962) study also when only K, P, N, and NP sentences are considered, and correct recall is the dependent variable. (The author defined correct recall by including the errors of the same type as the presented sentence with verbatim sentence recall.) The results from this analysis were K P N NP. These results, as mentioned above, are predictable also from the mean depth hypothesis, since the mean depth of the K sentences is 1.17; P, 1.38; N, 1.43; NP, 1.67 (Martin and Roberts, 1966, p. 217). The analysis of the transformational errors clearly failed to support the coding hypothesis since approximately as many sentences, when recalled incorrectly, were recalled in a more complex form than the presented form as were recalled as simplifications. It could be said, when appraising Mehler's study (1963) as a whole, that correct sentence recall can most parsimoniously be described in terms of the mean depth hypothesis, while transformational history of the sentences seemed to affect recall errors, but not in the manner expected in light of the coding hypothesis. In addition, Mehler's presentation of the recall errors data was useful in that this provided the means for making some guesses as to the process variables which might be involved in a sentence recall task.

One way to measure the amount of space that is left in immediate memory by various sentence types is to require *S*s to listen to a sentence, then a list of unrelated words, have them repeat the sentence, and then the words. Savin and Perchonock (1965) used this technique and the eight sentence types employed by Miller (1962) and by Mehler (1963) (they also used emphatics and wh-questions, which will not be discussed here). The dependent variable used was the number of words correctly recalled after an absolute correct recall of a sentence. Ss heard 10 Ks, and 5 instances of each of the other sentence types. Each sentence was heard once, and, if imperfectly recalled, data for that trial were apparently excluded, although some mention is made about sentence errors, so that this is not entirely clear. It is interesting that 5% of the sentence errors were transformations of the presented sentence and were scored according to the form of the response. This, of course , may give quite different results from what might be found if a more stringent analysis were performed. Savin and Perchonock also state that, while 42% of these transformation errors were deletions of the emphatic, 58% were other errors (unspecified) which "...were distributed quite uniformly over all other sentence-types" (p. 351).

The data of recalled words supported the coding hypothesis, and K was overwhelmingly easier than all other types (always $p < .02$ or better). It should be recalled, however, that there were twice as many K sentences per set than there were any of the other types. From this one could conjecture that a response set had been established which favoured guessing in the K form, or perhaps established an anticipatory response bias which favoured K's. If either of these hypotheses is valid, the results would have been similar to Savin and Perchonock's (1965). One wonders whether it was either of these factors, or whether it was the coding phenomenon which affected memory and these response patterns .

Mathews (1968), used a similar technique to that used by Savin and Perchonock (1965), but with written responses, and qualified as well as unqualified sentences. He found no difference in the numbers of words recalled after the various sentence types, and reported that sentence recall errors were of a more complex syntactic structure than the presented sentences. He controlled for sentence length, which Savin and Perchonock (1965) did not. The author points out that their results could as easily have been predicted by stimulus sentence length as by the coding hypothesis. Again, data were excluded if the sentences were recalled incorrectly. The number of words correctly recalled after a sentence was the dependent variable. Unfortunately, kinds of recall errors were not reported, so no insight into the Ss' attempts to master the task could be gained.

Mathews (1968, p. 128) argues that in a given situation the subject is attempting to process an input larger than the normal memory span and where the most efficient ways of encoding and storing material must be used, there is very little support for the notion that this method is by encoding on transformational principles. Instead, he takes a view that the unit which is encoded and stored is little affected by syntactic complexity. He proposes that the apparently less relevant qualifiers are discarded at presentation and the S attempts to encode and store the syntactically formed unit of whatever complexity.

Howe (1970), in a more carefully controlled but otherwise rather similar study to Mehler's (1963), examined the relationships among transformations, semantic errors, and the associative uncertainty (H) of sentences in a free recall situation. He also used the eight sentence types of a "family" and allowed Ss 16 trials on an 8-sentence word list with orally-presented stimuli and written responses. Declarative sentences were much easier to recall than were interrogative sentences ($p < .001$), and there was no difference between actives and passives. There were weak interactions ($p < .05$) between declarative and interrogative, and affirmative and negative, with a tendency for more transformation errors (and correct recalls) to occur for queries than declaratives and an increase in this trend when sentences were affirmative. As in Mehler's study (1963), Q and NQ were often confused at recall as were PQ and NPQ. It is worth noting here that Q and PQ are simplifications of NQ and NPQ, and as such are errors predictable according to the coding hypothesis. NPQ errors when PQ was the stimulus sentence type, and NQ for Q are not predictable from the coding hypothesis. Howe (1970) offers no explanation for these results. This, and the fact that, as in Mehler's study (1963), sentence length and mean depth are confounded seem to support the conclusion that Howe's results (1970) are suggestive only.

Bacharach and Kellas' (1971) first experiment involved K, P, and P_T sentences (their second study dealt with mean depth and was previously discussed). Sentence type was a between-S variable as was filled retention interval (0, 15, 30 seconds). The Ss heard ten examples of the 6-word sentences of any one type and wrote their responses. Time in seconds to record and whole sentence correct recalls were the dependent variables. No difference between active and passive sentences was found in this study, and was suspected by the experimenters to have been a result of sentence type's having been a between-subjects variable. The result was that very little can be gathered from these data regarding the relative difficulty of K, P, and P_T sentence types in a recall situation.

In another study attempting to assess the relative difficulty of active versus passive sentences in a recall situation, André and Kulhavy (1971) also found no differences between correct recall of K and P sentences. Their task was one in which Ss heard active or passive sentences with categorized "acted-upon" nouns, belonging to one of three categories: animal, occupation, or natural earth formation. After 5 study and 5 free-recall trials, the Ss were asked to sort the sentences according to categories. They overwhelmingly sorted according to the "acted upon" noun as if the category characteristics of the nouns were so strong as to have caused the Ss to focus on these to the exclusion of syntax. André and Kulhavy (1971) interpreted their findings in the light of Chomsky's (1965) notion that active and passive sentences have similar deep structures, and therefore the two types of sentence should be equally easy to recall. In this respect, it seems highly likely that the results are equally well explained in terms of an induced response bias due to obvious category associations between nouns of the various sentence stimuli.

Active and passive sentences were correctly recalled equally well by the Ss in James, Thompson, and Baldwin's (1973) study also. The Ss were asked to recall active and passive sentences, with either high or low imagery subjects or object nouns, 2, 3, or 4 days after hearing them. As in several of the studies reviewed here, while sentence type, as measured by transformational history, did not predict verbatim sentence recall, surface structure complexity did.

It would seem, in the light of the evaluation of sentence recall data, that Bever's (1968) conclusion that "...sentences are memorized in terms of their underlying syntactic structures..." needs re-evaluation. When free recall is the task under study, the literature seems to support the following generalizations concerning the coding hypothesis:

a. Many of the studies which have been interpreted as supporting the coding hypothesis have been methodologically flawed in that sentence type has been confounded with other variables such as mean depth and sentence length, or induced some kinds of response set the effect of which is not certain (e.g., Miller, 1962; Mehler, 1963; Savin and Perchonock, 1965).

b. Furthermore, the reported studies often fail to analyze recall errors, thus depriving one of the opportunity to trace through the possible processes involved during the recall trials.

c. There seems to be some indication that voice differences do not affect sentence recall (Martin and Roberts, 1966; Roberts, 1968; Mathews, 1968; Perfetti, 1969a; Wearing, 1969; Howe, 1970, André and Kulhavy, 1971; Bacharach and Kellas, 1971; James, Thompson, and Baldwin, 1973), although several of the studies which conclude this are suggestive only.

d. There also seems to be lack of agreement regarding the relative impact on correct recall of modality and mood differences (Mehler, 1963; Martin and Roberts, 1966; Mitchell, 1968; Howe, 1970).

C. Response Mode and Practice as Related to Sentence Characteristics

Finally, there are two variables which have so far been excluded from sentence recall studies attempting to test the predictive validity of both the coding and the mean depth hypotheses. One of these variables is response mode. Studies dealing with either hypothesis have overwhelmingly favoured oral presentation of sentence stimuli and written responses. It should be pointed out that some of the conflicting findings associated with the two major hypotheses previously discussed, may well be due to different response modes used. (For example, refer to Martin and Roberts, 1966; Martin, et al., 1968.) As part of the function of the present study is to attempt to re-assess Martin and Roberts' listening-reproducing psychology, oral presentation of sentence stimuli is used for comparison with Martin and Roberts' study. While a plea has been made (Slobin, 1968) for investigation of possible differences in performance which might occur in conditions of written versus oral presentation of sentence stimuli, no such argument has been advanced regarding response mode in keeping with the listening aspect of their model. It seems useful, however, to include both oral and written response modes in the present study: to examine the first since this is the mode in which one most frequently responds when asked to recall what a speaker has just said, and the second for a more direct comparison with the results from the studies reviewed in the present study. There appear to be no studies of sentence recall which compare recall performance under these two response conditions.

Intuition would suggest that during writing three cues are available to the Ss: tactile, visual, and oral (vocal or sub-vocal rehearsal), and only the latter is available to the Ss who speak their responses. From this it is predicted, therefore, that Ss who write their responses will correctly recall more sentences, phrases and words than will Ss who simply speak their responses. Further, it is suggested in the present study that the expected response mode differences might increase over trials with Ss in both groups performing poorly initially in terms of verbatim recall as a result of the large load placed on all Ss' memories by the general task demands. Both oral and written response groups will undoubtedly demonstrate equally poor recall performance in the first trial, after which the facilitative effects of writing the response are expected to show up and compound over trials.

The second element which has not been included in other studies of mean depth and coding hypotheses is the variable of practice. It is likely that Ss begin the recall task with strategies and problem solving hypotheses (Reber, 1967; Reber and Millward, 1968). It may very well be that at the beginning of the recall task these will not match the speaker's intentions or plans for each sentence. It is expected, however, that the early trials will reveal which aspects of each sentence are most memorable, and will perhaps reflect the strategies employed. Alternatively, only after the Ss have heard several repetitions of the sentences, and have several times attempted their recall, might the responses be expected to reflect the Ss' knowledge of the original speaker's speech intentions for each sentence.

D. Summary of Hypotheses

The purpose of the present study was to attempt to examine the functions of some specific variables which might affect performance on a sentence recall task. Two models of sentence production provide the major hypotheses for the present study. The sentence reproduction model reformulated by Martin and Roberts (1966) and based upon Yngve's production model proposes a "...psychology of listening to and reproducing ordinary English sentences" (p. 211). Martin and Roberts suggest that the likelihood of recall of a sentence is inversely related to the mean depth of that sentence. The preceding review of the literature generally supported this assertion but not unequivocally. First, there are more stringent, and important, ways to test this hypothesis. Second, much of the supportive data was derived from ex post facto analyses of the coding hypothesis rather than from direct empirical tests.

The coding hypothesis (Miller, 1962) suggests that sentences are encoded as the string of symbols underlying the kernel form, plus transformational tags. A prediction that can be made from this hypothesis (Schlesinger, 1968) is that kernel sentences will be easiest to recall, while sentences which involve one-tag transformations (e.g. N, P, Q) will be next easiest and easier to recall than those which involve two tags (e.g. NP, NQ, PQ), and two-tag sentences will be easier than a three-tag sentence (NPQ). A second prediction is that recall errors will be in the direction of the kernel sentence type. The results of studies examining the coding hypothesis have not been clear, largely because, as is the case with studies of mean depth, many have been methodologically flawed.

Finally, the relegation of stimulus sentence length, differential response modes, and the variable of repeated practice to minor roles by the majority of the studies reviewed is seen as a serious deficiency in terms of restricting the examination of variables which might affect sentence memory processes.

Many have operated here as though the critical variables were all linguistic, hence these other variables have been neglected. As recent work in the field has indicated, performance variables interact strongly with so-called "competence" factors and the nature of the interaction is poorly understood. As cases in point, when studies have failed to conform to predictions of linguistic theories, post hoc appeals to these performance factors are made to "rescue" the interpretations (cf. Perfetti, 1969a).

The major variables examined are shown in Figure 4, which also indicates that a focus of the study is to attempt to ascertain relationships which might exist between the stimulus and response variables and the recall performance and process variables studied.

If Martin and Roberts' reformulation of Yngve's speech production model is correct, and if the mean depth hypothesis is psychologically real, the following predictions would be expected to be upheld:

a. Verbatim recall of sentences and phrases will be inversely related to their mean depth.

b. The conditional error probability of a verb phrase error, given the subject noun phrase is recalled correctly (VP_e / NP_{1c}), should be greater than (NP_{2e} / VP_c), and both of these should be less than (NP_{2e} / NP_{1c}). This

Relationship Studied

Stimulus and Response Variables

1. Surface complexity of sentence
2. Transformation type of sentence
3. Length of sentence
4. Output response mode.
5. Practice

Processes

1. Mean depth hypothesis
2. Coding hypothesis

Recall Performance Variables

1. Verbatim recall of sentence
2. Verbatim recall of phrase units
3. Verbatim recall of major message units

Process Variables

1. Conditional dependency between phrases
2. Transformational errors

Fig. 4. A Schematic Illustration of Experimental and Dependent Variables Studied

should be the case particularly in the later trials of recall if such a binary process of encoding and decoding sentences as implied in the tree diagram shown in Figure 1b is going on in Ss. For example, in the sentence "The dogs tease the cat", one would expect, given binary left-to-right processing, that once a S has correctly recalled the NP_1 and the verb, memory load is greatly reduced and all that remains to be recalled (NP_2) is recalled relatively easily, because there is nothing left to interfere with it. One would expect the S to have somewhat greater difficulty in recalling the VP, given he has correctly recalled NP_1, since the VP may comprise a substantial amount of what is stored in memory. The most remote relationship should be between NP_1 and NP_2 since, once NP_1 is correctly recalled, both VP and NP_2 remain in memory and there are potentially many opportunities for memory failures during their recall.

Furthermore, if Miller's notion of sentence families is correct, and the coding hypothesis is psychologically real, the following predictions would be expected to be upheld:

a. Verbatim recall will be greater for K sentences than for the other seven types, followed by N, P, and Q, followed by NP, NQ, PQ, and finally by NPQ. Alternately, the mood, voice, and modality of sentences as the complexity variables should show up as significantly effective variables for predicting recall performance.

b. Transformation errors will be simplifications toward the K sentence type.

It is expected that there will be fewer verbatim recalls of sentences, phrases, and major message units when sentences are long than when they are short. If sentence length is a significant factor in the sentence recall process, intuitively one would predict that it would interact with sentence mean depth. The seven-word sentences of the present study are believed to fall well within the limits of short-term memory, so that, if recall performance differences occur within this length, they are expected to be wholly due to differences in mean depth of the sentence stimuli. The mean depth effect might be expected, however, to be compounded by increasing the length of the sentences to ten words, since this would be expected to increase the load on memory . It would be expected that the same effects might occur when verbatim recall of phrases is the dependent measure.

When practice trials are blocked, a blocks by length interaction effect is expected to occur, with significantly fewer verbatim sentence, phrase, and major message unit responses during the first half of the trials when stimulus sentences are long, with the length differences diminishing during the latter half of the trials.

CHAPTER III

METHOD

A. Experimental Design

The present experimental study was carried out in a basic 2 x 2 x 2 x 2 x 8 x 8 factorial design. Sentence Length (7 and 10 words), mean Depth (low, 1.40-1.43; high, 1.85-1.90), and Response mode (oral and written) were between subjects factors. Each *S* received eight trials on the same eight sentences, one sentence of each of eight Types. (For the definitions of mean depth and types of sentences, see Figure 1b (p. 8), Figure 3 (p. 17), and Appendices A and B as well as the section of Stimulus Materials.)

Two Blocks of four trials (obtained by collapsing across the first four trials and the last four trials) and sentence Type were within subject variables. Since the eight sentence types were formed by factorially combining Mood, Voice, and Modality, orthogonal comparisons were made to assess effects on correct recall of these latter three variables as well. The order of presentation of the sentences was counterbalanced across trials (Winer, 1971, p.691). There were two concomitant variables, STEP Listening Test performance and Grade Point Average. The GPA was associated only with the verbatim recall of sentences and verb phrases with both regression coefficients = .20, *F*s (1,118) = 12.99, 12.01, respectively $P < .001$, and the STEP Listening Test score was related to the recall of subject nouns, verbs, object nouns (regression coefficients = .026, .036, .037; *F*s (1,118) = 3.39, 5.15, 2.15; *P* $<$ .068 , .025, .009, respectively).

B. Subjects

The Ss were 128 Grade Eleven students attending Delta Senior Secondary School in Delta, B.C. during the school year 1972-73. All were at least second-generation Canadians who spoke English as a first language. Their parents were all either professionals, semi-professionals or skilled labourers.

No Ss, according to school and Public Health records, had a hearing, visual or other physical impediment which might interfere with his functioning in the task. No child was classified by the school personnel as being in need of remedial help in any subject.

The mean STEP Test score was 54.62 and the mean GPA was 2.97. Current I.Q. scores were available for only some of the Ss, so were not recorded.

C. Stimulus Materials

The orthogonal variations of the length (2), mean depth (2), and types (8) of sentences required the construction of 32 different stimulus sentences. The basic patterns of the sentence materials are shown in Table 1. The mean depth of sentences, as defined by Martin and Roberts (1966), was varied by adding or deleting qualifiers and thereby employing discontinuous constituents, while keeping the same sentence patterns based on the use of transitive verbs across all sentences.

More specifically mean depth was increased by:

1. replacing an adjective from the NP_1 phrase by an adverb in the Predicate Phrase which then became the last word in the sentence, thus exaggerating the regressive structure

Table 1

Basic Constituent Patterns of Stimulus Sentences Used

Gr.	Type	Noun Phrase$_1$, Subject	Predicate Verb Phrase	Noun Phrase$_2$, Object
	K	Adj. N	Aux. Adv. V	Adj. N
	N	Adj. N	Aux. Neg. V	Adj. N
1,	P	Adj. N.	Aux. Adv. V	Prep. N
2,	NP	Adj. N	Aux. Neg. V	Prep. N
3, 4,		Sentences KQ, NQ, PQ, NPQ are the same as above except "were....." or "are....."		
5,	K	N	Aux. Adv. V ...Adv.	Adj. N
6,	N	N	Aux. Nev. V ...Adv.	Adj. N
7,	P	N	Aux. Adv. V ...Adv.	Prep. N
8,	NP	N	Aux. Neg. V ...Adv.	Prep. N
		Sentences KQ, NQ, PQ, NPQ are the same as above except "were....." or "are....."		
9,	K	Adj. Adj. N	Aux. Adv. V	Adj. Adj. Adj. N
10,	N	Adj. Adj. N	Aux. Neg. V	Adj. Adj. Adj. N
11,	P	Adj. Adj. N	Aux. Adv. V	Prep. Adj. Adj. N
12,	NP	Adj. Adj. N	Aux. Neg. V	Prep. Adj. Adj. N
		Sentences KQ, NQ, PQ, NPQ are the same as above except "were....." or "are....."		
13,	K	Adj. Adj. Adj. N	Aux. Adv. V ....Adv.	Adj. N
14,	N	Adj. Adj. Adj. N	Aux. Neg. V ....Adv.	Adj. N
15,	P	Adj. Adj. Adj. N	Aux. Adv. V ....Adv.	Prep. N
16,	NP	Adj. Adj. Adj. N	Aux. Neg. V ....Adv.	Prep. N
		Sentences KQ, NQ, PQ, NPQ are the same as above except "were....." or "are....."		

of the Predicate Phrase (see Table 1, sentences for Groups 1-4 vs. sentences for Groups 5-8); or

2. adding another adjective to the Noun $Phrase_1$, deleting an adjective from the Noun $Phrase_2$, and adding an adverb as the final constituent of the Predicate Phrase (see Table 1, sentences for Groups 9-12 vs. sentences for Groups 13-16).

Surface structure phrase markers of the stimulus materials, indicating Yngve depths, as well as mean depth, appear in Appendix B. Attempts were made to control for unwanted possible sources of variation at the various levels of constituents of sentences.

First of all, it can be seen that Martin and Roberts (1966)[2] used sentence stimuli whose constituents varied, apparently somewhat randomly, within a sentence type. For example, of their six lower mean depth K sentences, the final Noun Phrase of three followed the pattern; Pronoun, Adjactive, Noun. Two sentences followed the pattern: Article, Adjective, Noun. One sentence followed the pattern: Adjective, Pronoun, Noun. They also used a mixture of transitive and intransitive verbs. To reduce the possibility of introducing deep structure differences and unknown "noise" which might result as a consequence of this mixture, only transitive verbs were used to keep the patterns of sentences constant.

Secondly, at the lexical level, only plural concrete nouns were used, consulting Paivio, Yuille, and Madigan (1968). This attempt was motivated by the observation that Martin and Roberts (1966) used a mixture

---

2 The author appreciates her personal communication with Dr. E. Martin who has supplied the set of sentence materials used in Martin and Roberts (1966).

of pronouns with singular and plural, concrete and abstract nouns in their sentence stimuli. The extent to which this mixture detracted from the clarity of their results is unknown. In order to more clearly determine the influence of the independent variables in this study, and to reduce the possibility of unwanted "noise" which might be caused by mixing pronouns with nouns, only nouns were used. Recall of nouns seems usually to be high (Mandler and Mandler, 1964; Martin, Roberts, and Collins, 1968; Martin and Walter, 1969; Howe, 1970; James, 1972), but seems particularly high for concrete nouns (Paivio, 1967; Beatty and Borree, 1971; James, 1972).

No derived modifiers were used to exert some control over possible deep structure differences among sentences, except in cases such as "broken dikes" and "willingly" where "broken" and "willingly" were judged to be used so commonly as modifiers that they are as 'natural' in this form as in their verb forms.

The eight types of sentences were those described by Miller (1962, p. 760) and are shown in Figure 3 (p. 17):

1. Kernel (K)
2. Negative (N)
3. Passive (P)
4. Negative-Passive (NP)
5. Kernel Question (KQ)
6. Negative Question (NQ)
7. Passive Question (PQ)
8. Negative-Passive Question (NPQ)

While the simple active assertive or declarative sentence type is not currently called Kernel, this label is used here so as to avoid confusion when comparisons are made with previous studies.

Because of the requirements of the basic experimental design, that is, to present the eight types of sentences describing the eight different semantic situations which were thematically independent to different experimental Ss, an attempt was initially made to demonstrate that the set of different semantic situations does not introduce biases in evaluating the effects of other experimental variables. This was achieved by constructing an alternative control set using:

(1) two sentences describing separate semantic situations;

(2) these same two situations occurring within each type, across length and mean depth, with modifiers being added to increase sentence length from seven to ten words (see Appendix A).

The eight sentence types were presented to all Ss. Ss who heard Set One K, N, P and NP (of any one mean depth and length combination) also received the KQ, NQ, PQ and NPQ transforms of the Set Two K, N, P and NP. Ss who heard Set Two K, N, P and NP (of any one mean depth and length combination) also heard the KQ, NQ, PQ and NPQ transforms of the Set One K, N, P and NP. In this way the sixteen groups of sentence stimuli were created and the sixteen experimental groups established.

D. Apparatus and Procedure

Stimulus sentences were tape recorded for standardization, particularly to ensure against variations in such things as stress, pitch and juncture, and further to ensure that these reflected the phrase markers for the sentences (Appendix B). Seven-word sentences were recorded at a rate of one sentence per 50 seconds with 2-1/2 seconds between sentences. Ten-word sentences were recorded at a rate of one sentence per 72 seconds with 2-1/2 seconds between sentences.

The taped experimental stimulus materials were presented on transistorized Sony TC-110 cassette recorders with built-in, remote-controlled microphones. Ss in the Oral Response groups spoke their responses into similar Sony tape recorders, and these responses were later transcribed into booklets for scoring. Ss in the Written Response groups wrote their responses in eight-page booklets.

A table screen separated the S from the experimenter (E) so as to reduce distractions which might have interfered with the S's concentration in the task.

Response booklets and IBM scoring sheets provided by the Educational Testing Service (Princeton, N.J.) were used for the STEP Listening Test, Form 2B (1957). All four hundred Grade Eleven students of Delta Senior Secondary School were administered the STEP Listening Test, Form 2B (1957) in one sitting with a five-minute break between the two 45-minute halves.

The students with the top one hundred twenty-eight scores were chosen to be Ss and each was assigned to one of the sixteen experimental groups initially established. The Ss with the top sixteen scores were assigned, at random, one to each group. This procedure was followed eight times until the sixteen (see Appendix A for details of the design) groups each comprised eight Ss.

Only students with the top scores on the STEP were chosen as it was felt that in this way there was some assurance that they had demonstrated skills necessary for the experimental task: hearing, processing and acting upon information listened to. Having assured the presence of these skills in the Ss' repertoires, it was felt that any difficulty the Ss had with the experimental task might more safely be attributed to the task itself and not to the S's deficiency with critical skills.

Ss were tested individually in the presence of the E, in a quiet room empty of all else but the experimental equipment. After being seated at the table, the S was shown how the recording equipment would operate during the task, and told that the E would operate all equipment.

Subjects in the Oral Response groups (1, 2, 5, 6, 9, 10, 13, 14; see Appendix A) heard the instructions (see Appendix C for details). The Ss were told to listen carefully to the playing of tape recorded sentences and to write or say as many of the sentences or parts and fragments as they could remember in any order. The E removed each page after the S indicated that he had completed his or her responses for each trial. Upon completion of all eight trials, every second S was asked to describe any technique or strategy used in recall sentences, and was requested not to discuss the experiment with any other student. The S was then released with thanks.

CHAPTER IV

RESULTS

A. Recall Performance Variables

1. Analysis of Verbatim Recall of Sentences

The majority of reported studies evaluated the effects of the variables under investigation in terms of verbatim recall of sentences as presented to Ss. Since the present study was also concerned with the predictive validities of the hypotheses in terms of this measure, Ss' recall recponses were first scored with respect to the number of sentences recalled absolutely verbatim over two blocks of four trials. The observed number of sentences recalled is presented in Table 2 for each treatment of combination. Initially an analysis of covariance was performed in an attempt to reduce error variance using the GPA and STEP Listening Scores as covariates. The result of this analysis was briefly mentioned elsewhere (see Chapter III). For the reasons already given there, only the results of an analysis of variance performed on the data in Table 2 will be presented here. Any result of statistical tests of p .01 will not be considered in the following presentation.

The results of the analysis of variance of the verbatim recall of sentences are presented in summary Table Form in Appendix F. Inspection of the table reveals that all the main effects are significant. As expected, short sentences are, in general, recalled better than long sentences; F (1,, 120) = 19.724, $p < .001$. The mean numbers of correct verbatim sentence recalls are 1.041 and 0.617 for short and long sentences, respectively. Similarly, sentences of low mean depth were recalled better than sentences with high

Table 2. Mean Number of Sentences Recalled Verbatim over Two Blocks of Four Trials as a Function of Experimental Treatments

											MARGINAL
	LENGTH =		1.000	1.000	1.000	1.000	2.000	2.000	2.000	2.000	
	DEPTH =		1.000	1.000	2.000	2.000	1.000	1.000	2.000	2.000	
	RESPONSE=		1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000	
	BLOC	TYPE									
K	1	1	.37500	1.62500	.50000	.37500	0.0	.18750	.06250	.31250	.49219
N	1	2	.56250	.68750	.12500	.56250	0.0	.43750	.12500	.12500	.32813
P	1	3	.31250	.81250	.31250	.50000	.12500	.25000	.06250	.12500	.31250
NP	1	4	.25000	1.75000	.37500	1.00000	0.0	.31250	.12500	.37500	.52344
Q	1	5	.43750	1.00000	.12500	.56250	.06250	.25000	0.0	.06250	.31250
NQ	1	6	.25000	.68750	.12500	.87500	0.0	.18750	.12500	.18750	.30469
PQ	1	7	.37500	.56250	.31250	.68750	0.0	.31250	0.0	.06250	.28906
NPQ	1	8	.56250	1.31250	.18750	.81250	0.0	.37500	0.0	.25000	.43750
K	2	1	.93750	2.93750	.68750	1.87500	.93750	2.25000	0.0	1.18750	1.35156
N	2	2	1.56250	2.37500	.87500	1.68750	.43750	1.93750	.31250	1.37500	1.32031
P	2	3	.68750	2.56250	.87500	1.93750	1.25000	2.31250	.06250	1.25000	1.35719
NP	2	4	1.12500	2.93750	.87500	2.12500	.62500	2.12500	.68750	1.81250	1.53906
Q	2	5	.81250	2.31250	.43750	1.75000	.31250	1.68750	.25000	.81250	1.04688
NQ	2	6	.81250	2.43750	1.00000	2.31250	.18750	2.00000	.12500	1.62500	1.31250
PQ	2	7	.56250	1.87500	.12500	1.43750	.81250	2.25000	.18750	1.00000	1.03125
NPQ	2	8	.50000	2.56250	.12500	2.00000	.18750	2.25000	.56250	2.18750	1.29688
	MARGINAL		.63281	1.77734	.44141	1.31250	.30859	1.19531	.16797	.79688	.82910
	COUNT		16	16	16	16	16	16	16	16	128

mean depth with F (1, 120) = 0.805; $p < 0.001$, and respective verbatim recall means of 0.979 and 0.680. Subjects required to write their responses performed better than did subjects who were required to say their responses, with respective means of 1.271 and 0.388; $F (1, 120) = 85.587$; $p < .001$. A significant effect of blocks, $F (1, 120) = 259.388$, $p < .001$, resulted because recall during the second block of presentations (mean = 1.283) was better than recall during the first block of presentations (mean = 0.375). This effect can be seen to indicate that the subjects were learning the sentences over blocks of trials. The type of sentence being recalled also had a significant effect, $F (7, 840) = 4.491$; $p < .001$, with negative-passive sentences being recalled best (mean = 1.031), kernel sentences next (mean = 0.921), negative-passive-questions next (mean = 0.867), followed by simple passive sentences, simple negative sentences, negative-questions, simple questions, and passive questions with respective means of 0.840, 0.824, 0.809, 0.680, and 0.660.

The significant main effect of Type of Sentences was decomposed into three sources of contrasts each associated with the ordered prediction of K > Types with 1 tag (N, P, Q) Types with 2 tags > Types with 3 tags from K. In order to ascertain the ordinality of correct recall of sentences by the types of sentences as would be predicted by the coding hypothesis, three contrasts using the Bonferroni t-tests with the overall type error of .05 were made between (1) K and Types of 1-tag transform, (2) Types of 1-tag and Types of 2-tag transform, and (3) Types of 2-tag and Types of 3-tag transform. None of them was found to be significant.

In order to determine the effects of sentence mood, voice, and modality, appropriate orthogonal comparisons among the eight sentence types were made. Declarative sentences were recalled better than interrogative sentences (mood); F (1, 120) = 11.891, $p < .001$, with respective means at 0.904 and 0.754, and negative sentences were recalled better than affirmative sentences (modality); F (1, 120) = 5.784, $p < .018$, with respective means of 0.883 and 0.776. Voice had no effect on verbatim sentence recall; F (1, 120) < 1.

The first order interaction between blocks and response mode was significant; F (1, 120) = 78.618, $p < .001$. It appears that improvement in recall performance was much greater when responses were written rather than spoken. The first order interaction between blocks and mean depth was also significant; F (1, 120) = 9.080, $p < .003$.

The first order interaction between the sentence type and response mode was also found to be significant; F (9, 840) = 2.804, $p < .007$. The source of the interaction was primarily due to a three-way interaction involving voice, modality, and response mode such that negative passive and negative passive question forms of sentences were recalled more often than all other types given the written response mode; whereas given the oral response mode the voice and modality of sentences do not make any difference.

There is a highly significant second order interaction involving trial blocks, types, and lengths of sentences; F (7, 840) = 3.078, $p < .004$; which was found to be primarily due to a three-way interaction involving trial blocks, length and voice of sentences. That is to say, given active sentences, where is no interaction between blocks and length of the sentences, whereas longer sentences were recalled more often than any other only in the later trial block when they were passive sentences, $\underline{F}$ (1, 120) = 10.418, $\underline{p} < .002$. All other first, second, and higher order interactions were foun d to be nonsignificant and thus are not mentioned here (for specific details, see Appendix F).

2. Analyses of Verbatim Recall of Phrases (NP_1, VP, NP_2)

Even if one fails to recall presented sentences verbatim, he could certainly recall parts or segments of them as $\underline{S}$s were initially instructed to try to do so. The basic question to be answered here using verbatim recall of phrases is whether there would be any effects on phrase recall of length, mean depth, and types of sentences as contextual effects and those of response mode and practice block. The observed mean numbers of subject noun, predicate verb, and object noun phrases recalled are shown in Tables 3a, 3b, and 3c. Three separate analyses of variance were performed on data in the tables of means, the results of which are presented in Appendices G-1, G-2, and G-3, for NP_1, VP, and NP_2, respectively.

Examination of the tables indicates that the main effect of the length of sentences (with the mean differences 1.323, .644) is significant as expected, only in terms of NP_1 and NP_2 recall; $\underline{F}$s (1, 120) = 141.616, 40.931, $\underline{p}$s $< .0001$; but not in terms of VP recall with mean difference of .134 $\underline{F}$ (1, 120) = 1.820, $\underline{p} < .180$. The main effect of mean depth is also significant, as expected, now only in terms of VP with the mean difference of .528 recall; $\underline{F}$ (1, 120) = 28.399, $\underline{p} < .00001$; but not in terms of NP_1 and NP_2 recall; $\underline{F}$s (1, 120) = 3.603, 3.512, $\underline{p}$s $< .063$. However, the main effect of type of sentence is significant in terms of all three phrase recalls;

Table 3a. Mean Number of Subject-noun Phrases Recalled Verbatim Over Two Blocks of Four Trials as a Function of Experimental Treatments

	LENGTH =		1.000	1.000	1.000	1.000	2.000	2.000	2.000	2.000	MARGINAL
	DEPTH =		1.000	1.000	2.000	2.000	1.000	1.000	2.000	2.000	
	RESPONSE=		1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000	
	TYPE	BLOC									
NP1	1	1	1.68750	3.37500	2.18750	2.62500	.50000	1.18750	.37500	.93750	1.60938
NP1	1	2	2.50000	3.87500	2.50000	3.81250	2.31250	3.43750	1.06250	1.75000	2.65625
NP1	2	1	1.37500	2.56250	2.06250	2.75000	.50000	1.37500	.31250	.62500	1.50781
NP1	2	2	3.31250	4.00000	3.25000	3.62500	2.25000	3.18750	.87500	2.50000	2.87500
NP1	3	1	2.18750	2.62500	2.50000	2.75000	1.25000	2.06250	.56250	.75000	1.83594
NP1	3	2	3.00000	3.87500	3.00000	3.81250	2.56250	3.12500	1.31250	2.93750	2.95313
NP1	4	1	1.68750	2.87500	2.43750	2.43750	.62500	1.18750	.68750	1.12500	1.63281
NP1	4	2	2.62500	3.87500	3.06250	3.75000	1.43750	2.93750	1.62500	2.75000	2.75781
NP1	5	1	1.37500	2.50000	1.93750	2.68750	.81250	1.06250	.12500	.25000	1.40625
NP1	5	2	2.06250	3.68750	2.87500	3.68750	1.93750	3.12500	.81250	2.18750	2.54688
NP1	6	1	1.25000	2.00000	2.12500	2.93750	.31250	1.37500	.25000	.56250	1.35156
NP1	6	2	2.56250	3.93750	3.62500	3.93750	1.18750	2.62500	.81250	2.62500	2.66406
NP1	7	1	1.68750	2.56250	1.93750	2.68750	.50000	1.43750	.18750	.62500	1.45313
NP1	7	2	2.18750	3.62500	2.37500	3.62500	2.25000	3.68750	.93750	2.56250	2.65625
NP1	8	1	1.68750	2.68750	1.93750	2.81250	.68750	1.00000	.37500	.56250	1.46875
NP1	8	2	2.31250	3.68750	3.18750	3.50000	1.50000	3.43750	1.37500	2.75000	2.71875
	MARGINAL		2.15625	3.23438	2.56250	3.21484	1.28906	2.26563	.73047	1.59375	2.13086
	COUNT		16	16	16	16	16	16	16	16	128

Table 3b. Mean Number of Predicate-verb Phrases Recalled Verbatim over Two Blocks of Four Trials as Function of Experimental Treatments

											MARGINAL
	LENGTH =		1.000	1.000	1.000	1.000	2.000	2.000	2.000	2.000	
	DEPTH =		1.000	1.000	2.000	2.000	1.000	1.000	2.000	2.000	
	RESPONSE =		1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000	
	TYPE	BLOC									
VP	1	1	.87500	2.37500	.50000	1.50000	.62500	.62500	.18750	.75000	.92969
VP	1	2	1.43750	3.12500	1.00000	2.25000	1.93750	3.18750	.87500	2.62500	2.05469
VP	2	1	1.06250	1.25000	.37500	.75000	1.31250	1.43750	.68750	.81250	.96094
VP	2	2	2.12500	2.68750	1.25000	2.00000	2.12500	3.25000	.93750	1.87500	2.03125
VP	3	1	.56250	1.56250	.37500	.75000	.31250	.68750	.06250	.18750	.56250
VP	3	2	1.18750	2.93750	1.12500	1.87500	1.62500	2.93750	.18750	1.75000	1.70313
VP	4	1	.62500	2.00000	.62500	1.25000	.43750	1.18750	.43750	.87500	.92969
VP	4	2	1.37500	3.06250	1.37500	2.50000	1.31250	2.56250	.87500	2.43750	1.93750
VP	5	1	.68750	1.75000	.43750	.93750	.75000	1.25000	.37500	.87500	.88281
VP	5	2	1.43750	3.18750	.62500	2.56250	1.62500	3.00000	1.06250	2.56250	2.00781
VP	6	1	1.06250	1.31250	.43750	1.12500	1.25000	1.25000	.31250	.50000	.90625
VP	6	2	2.12500	2.56250	1.37500	2.56250	2.12500	3.06250	1.00000	2.18750	2.12500
VP	7	1	.50000	1.18750	.50000	.81250	.06250	.43750	.12500	.12500	.46875
VP	7	2	.81250	2.43750	.37500	1.75000	1.25000	2.56250	.56250	1.68750	1.42969
VP	8	1	.87500	1.87500	.31250	1.12500	.93750	.75000	.25000	1.25000	.92188
VP	8	2	.81250	2.87500	.56250	2.81250	1.87500	2.68750	1.31250	2.81250	1.96875
	MARGINAL		1.09766	2.26172	.70313	1.66016	1.22266	1.92969	.57813	1.45703	1.36377
	COUNT		16	16	16	16	16	16	16	16	128

Table 3c. Mean Number of Object-noun Phrases Recalled Verbatim over Two Blocks of Four Trials as a Function of Experimental Treatments

7

CELL MEANS FOR 3-RD DEPENDENT VARIABLE

	LENGTH =	1.000	1.000	1.000	1.000	2.000	2.000	2.000	2.000	MARGINAL
	DEPTH =	1.000	1.000	2.000	2.000	1.000	1.000	2.000	2.000	
	RESPONSE=	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000	
	TYPE BLOC									
NP2	1 1	.50000	2.25000	1.50000	1.25000	.18750	.50000	.31250	.87500	.92188
NP2	1 2	1.93750	3.75000	1.68750	3.37500	1.25000	3.06250	1.12500	2.06250	2.28125
NP2	2 1	1.93750	2.37500	1.68750	2.00000	.37500	.87500	1.68750	1.75000	1.58594
NP2	2 2	2.87500	3.43750	2.50000	3.50000	1.37500	2.93750	2.56250	3.37500	2.82031
NP2	3 1	1.43750	1.62500	1.56250	2.50000	.75000	1.06250	.93750	1.56250	1.42969
NP2	3 2	2.12500	3.37500	2.43750	3.68750	2.12500	3.43750	2.50000	3.31250	2.87500
NP2	4 1	1.81250	2.81250	1.93750	2.37500	.37500	1.12500	1.31250	1.62500	1.67188
NP2	4 2	3.43750	4.00000	2.56250	3.50000	1.87500	3.75000	2.75000	3.56250	3.17969
NP2	5 1	1.06250	1.75000	1.25000	1.18750	.37500	.43750	.31250	.62500	.87500
NP2	5 2	1.68750	3.68750	1.50000	3.31250	.87500	2.00000	.75000	1.68750	1.93750
NP2	6 1	1.75000	2.25000	1.43750	2.56250	.18750	.81250	1.50000	2.12500	1.57813
NP2	6 2	2.25000	3.87500	2.87500	3.75000	1.31250	3.18750	2.62500	3.56250	2.92969
NP2	7 1	1.25000	2.06250	.75000	2.50000	.56250	1.56250	1.18750	1.12500	1.37500
NP2	7 2	2.06250	3.43750	1.50000	3.62500	1.62500	3.18750	2.43750	3.00000	2.60938
NP2	8 1	1.62500	2.62500	1.18750	2.06250	.18750	1.06250	1.75000	2.68750	1.64844
NP2	8 2	3.00000	4.00000	2.68750	3.62500	1.62500	3.18750	3.00000	3.81250	3.11719
	MARGINAL	1.92188	2.95703	1.81641	2.80078	.94141	2.01172	1.67188	2.29688	2.05225
	COUNT	16	16	16	16	16	16	16	16	128

Fs (7, 840) = 3.685, 8.118, and 22.696; $p < 001$, .0001, and .0001 for NP_1, VP and NP_2, respectively. The significant sources of variation for the type effects were partitioned into those associated with the predicted order of the varying degrees of simplification transformations toward the K-type sentences, namely K > types of 1 tag > types of 2 tags > type of 3 tags away from the K-type. As was done with the data of verbatim recall of sentences, three sets of three contrasts were made in terms of the three verbatim phrase recall using the Bonferroni t. None of each set of contrasts in terms of the NP_1 and VP phrase recall was significant at the overall type I error level of .05. Only two contrasts in terms of the NP_2 recall were found to be significant; $t\ (\alpha) = 3.572$ and 4.822; $p < .05$. This means that object-noun phrases in 1-tag sentences (N, P, Q) are recalled more often than when they are in K-type sentences, and object-noun phrases in 2-tag transform sentences (NP, NQ, PQ) are correctly recalled more often than are object-noun phrases in 1-tag sentences. Both results are in the opposite direction to that predicted by the coding hypothesis.

Alternatively, an orthogonal decomposition of the type effects according to the Mood, Voice, and Modality of sentences was carried out. The type effect in terms of the NP_1 recall is primarily attributable to variations in the Mood, that is NP_1 phrases in declarative sentences were recalled better than those in interrogative sentences with a mean difference of .196; $F\ (1, 120) = 15.902$, $p < .00001$ with means of 2.229 and 2.03. The type effect in terms of VP recall is found attributable to variations in the voice and modality of sentences as contexts. That is, verb phrases in the active sentences were recalled better than in the passive with the mean difference of .247 and those in the affirmative are recalled better than in the negative sentences with the mean difference being .217; $\underline{F}s\ (1, 120) = 17.401$, respectively and 13.930; $\underline{p}s < .0001$. The type effect in terms of NP_2 recall is also traceable to variations in the voice and modality of sentences. That is, NP_2 phrases in the passive sentences were recalled better than in the active with the mean difference of .364; and those in the negative sentences were recalled better than in the affirmative with the mean difference of .528; $\underline{F}s\ (1, 120) = 50.847$ and 97.848, respectively; $\underline{p}s < .00001$. The main effect of the response mode is significant as expected in terms of all three phrase recall measures with the written response mode being superior to the oral with the mean difference of .893, .927 and .929; $\underline{F}s\ (1, 120) =$ 64.538, 87.393, and 85.258 NP_1, VP and NP_2, respectively; $\underline{p}s < .00001$. The main effect of practice blocks is also significant as expected in terms of all three phrase recall measures with the mean differences of 1.195, 1.087, and 1.333; $\underline{F}s\ (1, 120) = 482.173$, 381.584, and 750.075 for NP_1, VP and NP_2, respectively; $\underline{p}s < .00001$.

The first order interaction effect between the length and depth of sentences as contexts on the recall of NP_1 phrases was significant; $\underline{F}\ (1, 120) = 13.242$, $\underline{p} < .0001$; and another significant one was the interaction effect between the type and depth of sentence as contexts on the NP_1 recall; $\underline{F}\ (1, 840) = 3.223$, $\underline{p} < .002$. The former means that, given a shorter sentence context, the low mean depth of the sentence depressed the recall of NP_1 relative to the high mean depth of them, whereas the high mean depth depressed the recall of NP_1 relative to the low mean depth given a longer sentence, as one might expect (see Table 1, p. 46, for constituents of phrases). The original source of the latter first order interaction was traced down to the interaction effect between the modality and depth of the sentence context, such that low depth in the affirmative context most enhanced and high depth in the negative context most depressed the recall of NP_1 phrases.

The recall of verb phrases was not affected at all in the same way by the two interactions described above. Instead, it was affected by the first order interaction between type and response mode, F (1, 840) = 2.957, $p < .005$. The main source of the interaction was traced down to the observation of the interaction of response mode, voice, and modality such that, given the oral response mode, the active-negative sentence context enhanced the VP recall most; whereas given the written response mode, the active-affirmative context or otherwise the passive-negative contaxt enhanced the VP recall; F (1, 120) = 15.356, $p < .0001$.

In contrast to the observations of the first order interaction made above in terms of the VP recall, the first order interaction effect between the length and depth of sentence contexts on the NP_2 recall was significant; F (1, 120) = 10.079, $p < .0001$, as was the interaction effect between type and mean depth of the sentence on the NP_2 recall; F (7, 840) = 3.327, $p < .002$. The former means that, given a short sentence context, high mean depth of the sentence depressed the recall of NP_2 relative to low mean depth, whereas low mean depth depressed the NP_1 recall relative to high mean depth given a long sentence, unlike what was observed for the NP_1 recall (see Table 1, p. 46, for constituents of phrases). The source of the latter first order interaction was traced down to the interaction of depth with modality, F (1, 120) = 10.712, $p < .001$; and with voice and modality, F (1, 120) = 6.980, $p < .009$, which means that NP_2 is far better recalled in negative than in affirmative sentences of high mean depth relative to that of low mean depth in general, further that such a differential effect is greather when the sentences are active rather than passive.

Another significant first order interaction effect was found between practice blocks and response mode; F (1, 120) = 24.874, 53.565, and 44.642, for NP_1, VP and NP_2, respectively; all $ps < .00001$, all indicating that oral recall of phrases does not improve over blocks, whereas written recall of phrases does improve over blocks of trials. A further significant first order interaction effect was also found between practice blocks and length of sentence; F (1, 120) = 18.231, 11.238, and 8.233 for NP_1, VP and NP_2 recall, respectively; all $ps < .005$, all indicating that, given a short sentence, recall does not improve over practice blocks, whereas it does given a long sentence.

There appear to be two significant second order interactions, one of which involves practice blocks, length of sentence, and response mode, F (1, 120) = 7.05, $p < .009$. This means that, given a short sentence, oral recall of NP_1 phrases improves over blocks but written recall of NP_1 does not, and that, given a long sentence, the converse is the case, as one would predict. However, this interaction in terms of VP and NP_2 recall came out as non-significant. Although another second order interaction effect involving the type, blocks, and length of sentence contexts on the NP_1, VP and NP_2 recall appeared as consistently significant; F (7, 840) = 3.354, 3.203, and 2.51; $p < .002$, .002, and .015 for the NP_1, VP and

NP_2 measures, respectively, no further tracing of the sources of variation was conducted here because of lack of usefulness of the finding.

3. Analyses of Pivotal Word Correct Recall (N_1, V, N_2)

Since the test response instructions given to Ss encouraged the recall of any parts or segments of the sentences presented if they fail to recall as given, it was expected that the recall of key words without any surface grammatical requirements would index something more than verbatim recall of sentences and phrases, as defined and scored in the present study. The pivotal (or key) words comprise the three key words in a given sentence; subject noun (N_1) predicate verb (V) and object noun (N_2). Subject and object nouns which were recalled verbatim, as well as those to which an article or determiner had been added, or which had been recalled in the singular form, or in a synonymous form, were scored as correct responses. Predicate verbs which were recalled verbatim and those recalled with optional tense changes, auxiliary changes, or in synonymous form were scored as correct verb responses. All other responses were scored as incorrect.

The number of correctly recalled pivotal words was expected to reflect some degree of comprehension during the processing of the given sentences (cf. Reid, 1974). Presented in Tables 4a, 4b, and 4c are the mean numbers of correctly recalled pivotal words: subject noun (N_1) predicate verb (V), and object noun (N_2), respectively. Three separate analyses of variance appropriate to the designs implied in the tables were performed on the data, the results of which are also presented in Appendices H-1, H-2, and H-3 for N_1, V, and N_2, respectively. A general question to be answered is whether or not the experimental variables serve as

Table 4a. Mean Number of Correctly Recalled Subject-nouns over Two Blocks of Four Trials as a Function of Experimental Treatments

CELL MEANS FOR 4-TH DEPENDENT VARIABLE

	LENGTH =		1.000	1.000	1.000	1.000	2.000	2.000	2.000	2.000	MARGINAL
	DEPTH =		1.000	1.000	2.000	2.000	1.000	1.000	2.000	2.000	
	RESPONSE =		1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000	
	TYPE	BLOC									
N1	1	1	2.12500	3.43750	2.87500	3.18750	.87500	1.25000	.43750	1.12500	1.91406
N1	1	2	3.06250	3.93750	3.00000	3.87500	2.81250	3.50000	1.37500	2.37500	2.99219
N1	2	1	2.06250	2.56250	2.37500	2.87500	.75000	1.43750	.37500	.62500	1.63281
N1	2	2	3.31250	4.00000	3.31250	3.87500	2.43750	3.18750	.93750	2.56250	2.95313
N1	3	1	2.43750	2.81250	2.81250	3.06250	1.81250	2.25000	.68750	.75000	2.07813
N1	3	2	3.12500	4.00000	3.50000	3.87500	3.25000	3.56250	1.37500	2.93750	3.20313
N1	4	1	2.62500	3.43750	2.75000	2.75000	1.87500	2.06250	1.50000	1.68750	2.33594
N1	4	2	3.68750	4.00000	3.37500	4.00000	3.31250	3.00000	2.25000	2.93750	3.32031
N1	5	1	2.06250	2.75000	2.25000	3.12500	1.18750	1.25000	.18750	.75000	1.69531
N1	5	2	2.62500	3.68750	3.43750	3.93750	2.93750	3.12500	1.06250	2.43750	2.90625
N1	6	1	1.31250	2.18750	2.31250	3.00000	.68750	1.56250	.37500	.75000	1.52344
N1	6	2	2.56250	4.00000	3.68750	3.93750	1.62500	2.62500	.81250	2.75000	2.75000
N1	7	1	1.87500	2.75000	2.06250	3.06250	1.12500	1.50000	.43750	.87500	1.71094
N1	7	2	2.75000	3.87500	2.50000	3.81250	3.12500	3.68750	1.37500	2.81250	2.99219
N1	8	1	2.31250	2.93750	2.43750	3.18750	1.81250	1.87500	1.12500	1.50000	2.14844
N1	8	2	3.18750	3.81250	3.56250	3.87500	2.56250	3.81250	2.00000	3.31250	3.26563
	MARGINAL		2.57031	3.38672	2.89063	3.46484	2.01172	2.48047	1.01953	1.88672	2.46387
	COUNT		16	16	16	16	16	16	16	16	128

Table 4b. Mean Number of Correctly Recalled Predicate Verbs over Two Blocks of Four Trials as a Function Experimental Treatments

9

CELL MEANS FOR 5-TH DEPENDENT VARIABLE

											MARGINAL
	LENGTH =		1.000	1.000	1.000	1.000	2.000	2.000	2.000	2.000	
	DEPTH =		1.000	1.000	2.000	2.000	1.000	1.000	2.000	2.000	
	RESPONSE=		1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000	
	TYPE	BLOC									
V	1	1	1.87500	3.12500	1.75000	2.31250	1.43750	1.50000	1.12500	1.75000	1.85938
V	1	2	2.56250	3.81250	2.06250	3.43750	2.93750	3.75000	2.06250	3.56250	3.02344
V	2	1	2.18750	3.25000	1.37500	2.62500	1.87500	2.37500	1.56250	2.25000	2.18750
V	2	2	3.62500	3.93750	2.37500	3.50000	3.50000	3.62500	2.43750	3.43750	3.30469
V	3	1	2.12500	2.62500	1.93750	2.37500	2.12500	2.31250	1.81250	1.68750	2.12500
V	3	2	2.75000	4.00000	2.87500	3.75000	3.25000	3.56250	2.75000	3.06250	3.25000
V	4	1	2.06250	2.68750	1.37500	1.93750	1.93750	2.18750	1.06250	1.56250	1.85156
V	4	2	3.31250	3.87500	2.93750	3.50000	3.50000	3.75000	3.00000	3.50000	3.42188
V	5	1	1.37500	2.43750	1.18750	2.18750	1.75000	1.93750	1.18750	1.68750	1.71875
V	5	2	2.31250	3.75000	1.75000	3.43750	2.62500	3.68750	2.12500	3.06250	2.84375
V	6	1	1.87500	2.68750	1.25000	2.50000	2.43750	2.18750	.81250	2.12500	1.98438
V	6	2	2.87500	3.81250	2.62500	3.37500	3.00000	3.43750	1.87500	3.62500	3.07813
V	7	1	1.31250	2.06250	1.56250	2.50000	1.56250	1.81250	1.75000	2.12500	1.83594
V	7	2	2.62500	3.25000	2.43750	3.12500	2.93750	3.31250	2.81250	2.81250	2.91406
V	8	1	2.00000	2.56250	1.43750	2.12500	2.12500	1.93750	1.50000	2.56250	2.03125
V	8	2	3.00000	3.93750	2.18750	3.37500	3.37500	3.56250	2.50000	3.31250	3.15625
	MARGINAL		2.35719	3.23828	1.94531	2.87891	2.52344	2.80859	1.89844	2.63281	2.53662
	CCUNT		16	16	16	16	16	16	16	16	128

Table 4c. Mean Number of Correctly Recalled Object Nouns over Two Blocks of Four Trials as a Function of Experimental Treatments

10

CELL MEANS FOR 6-TH DEPENDENT VARIABLE

	LENGTH =		1.000	1.000	1.000	1.000	2.000	2.000	2.000	2.000	MARGINAL
	DEPTH =		1.000	1.000	2.000	2.000	1.000	1.000	2.000	2.000	
	RESPONSE=		1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000	
	TYPE	BLOC									
N2	1	1	1.81250	2.87500	2.25000	2.18750	1.31250	1.00000	1.18750	1.93750	1.82031
N2	1	2	2.68750	3.93750	2.87500	3.75000	2.93750	3.37500	2.56250	3.31250	3.17969
N2	2	1	1.93750	2.37500	1.68750	2.18750	.37500	.87500	1.68750	1.93750	1.63281
N2	2	2	3.00000	3.43750	2.50000	3.50000	1.62500	2.93750	2.56250	3.43750	2.87500
N2	3	1	2.06250	2.18750	2.00000	2.68750	1.12500	1.43750	1.31250	1.87500	1.83594
N2	3	2	3.00000	3.93750	3.12500	3.81250	2.56250	3.56250	3.00000	3.62500	3.32813
N2	4	1	2.06250	3.00000	2.00000	2.56250	1.25000	1.31250	1.50000	1.87500	1.94531
N2	4	2	3.56250	4.00000	2.93750	3.62500	2.43750	3.81250	3.00000	3.75000	3.39063
N2	5	1	1.81250	2.56250	1.62500	2.00000	1.31250	1.31250	1.00000	1.81250	1.67969
N2	5	2	2.62500	3.87500	2.43750	3.68750	2.12500	2.37500	1.93750	3.00000	2.75781
N2	6	1	1.81250	2.31250	1.50000	2.62500	.50000	.81250	1.50000	2.25000	1.66406
N2	6	2	2.37500	3.93750	2.87500	3.87500	1.37500	3.31250	2.62500	3.75000	3.01563
N2	7	1	1.56250	2.62500	1.31250	3.12500	1.18750	2.06250	1.56250	1.50000	1.86719
N2	7	2	2.75000	3.87500	2.43750	3.81250	2.37500	3.43750	3.00000	3.68750	3.17188
N2	8	1	1.81250	2.75000	1.68750	2.62500	.37500	1.12500	2.00000	2.81250	1.89844
N2	8	2	3.43750	4.00000	3.12500	3.81250	2.25000	3.18750	3.50000	3.93750	3.40625
	MARGINAL		2.39453	3.23047	2.27344	3.11719	1.57031	2.24609	2.12109	2.78125	2.46680
	COUNT		16	16	16	16	16	16	16	16	128

inhibiting (or facilitating) context variables on the recall of pivotal lexical items.

Examination of the tables shows that the main effect of mean depth of the sentences on the recall of pivotal words does not remain the same across the recall of N_1, V, and N_2; so it is also with the main effect of length. The recall of the subject and object nouns of short sentences is much greater than that of the longer sentences with the mean differences of 1.229 and .574; $\underline{F}s\ (1, 120) = 160.258$ and 34.505 respectively; $\underline{p}s < .0001$; but there is no effect of length on the recall of the predicate verbs. The recall of the N_1 and V words in the sentences of high mean depth is much greater than that in those of low mean depth with the mean differences of .297 and .396; $\underline{F}\ (1, 120) = 9.358$ and 13.271; $\underline{p}s < .009$ and $.0001$, respectively; but there is no such significant effect on the recall of the N_2 words. The recall of the N_1 and N_2 words is affected by the interaction between the length and depth of the sentences where they were imbedded; $\underline{F}\ (1, 120) = 26.137$ and 11.403; $\underline{p}s < .0001$ and $.001$, respectively; whereas there is no such interaction effect on the recall of the verbs. The interactions appear to be what might be predicted on the basis of the number of lexical items that constitute the subject and object noun phrases (see Table 1, p. 46). That is, the N_1 items in the long sentences of high mean depth were least recalled and those in the short sentences of low mean depth most often correctly recalled; and the N_2 items in the long sentences of low mean depth were least well recalled. It looks as if the recall of nouns is depressed by the qualifiers surrounding them within each phrase unit.

All subject and object nouns, and verbs alike were recalled more in the written than oral responses with the mean differences of .682, .754, and .706; $\underline{F}s$ (1, 120) = 49.343, 59.493, and 42.302 respectively; all $\underline{p}s < .00001$. The recall of all three pivotal words improved over the practice blocks alike with the mean differences of 1.168, 1.793, and 1.175; $\underline{F}$ (1, 120) = 585.295, 979.852, and 528.509, respectively; all $\underline{p}s < .00001$, as might be expected. The first order interaction effect between the practice block and response mode on the recall of the N_1 and N_2 items is significant; $\underline{F}s$ (1, 120) = 12.055 and 16.672, respectively; $\underline{p}s < .001$; meaning that the effect of practice block is even greater in the written than in the oral response mode. Another first order interaction effect between the practice block and length of sentence context on the recall of the N_1 and N_2 is also significant; $\underline{F}$ (1, 120) = 26.817 and 19.768, respectively; $\underline{p} < .0001$; meaning that the effect of practice block is even greater in the long than in the short sentences. Both of the first order interaction effects described above are not observed on the recall of verbs. It is suggestive here that the depressing effect of the added qualifiers of the N_1 and N_2 diminishes over the practice blocks differentially more in the written than oral mode and more in the long sentences than in the short.

The recall of the three pivotal items is significantly affected by the type of sentence context where each occurs; $\underline{F}$ (7, 840) = 13.519, 6.046, and 5.996 for the N_1, V, and N_2 items, respectively; all $\underline{p}s < .0001$. The significant source of the type effect was sought to examine whether correct recall of pivotal items is affected by the types of sentences in the order predicted from the notion of simplification transformations, that is, K $>$ types of 1-tag transforms > types of 2-tag > type of 3-tag away from the K-type. Three sets of the same three contrasts as made in the corresponding analysis of the verbatim recall of sentences were made of the recall of pivotal items. None of them was significant except for two contrasts but in the opposite direction; that is, the N_1 items were recalled more often in the type of 3-tag transform than that of 2-tag; $t\ (\alpha) = 3.578$; $p < .05$ and the N_2 items recalled more often in the type of 2-tag than in the type of 1-tag transform; $t\ (\alpha) = 2.695$; $p < .05$.

Alternatively, an orthogonal decomposition of the type effects was carried out into the components by mood, voice, and modality of sentences. The N_1 and N_2 items are recalled more often in the passive sentences than in the active (voice); $F\ (1,\ 120) = 53.008$ and 29.181, respectively; $ps < .0001$. Particularly, the recall of the N_1 items is greatest in the negative-passive sentence as compared to the other combinations of voice and modality; $F\ (1,\ 120) = 27.261$, $p < .0001$. The N_1 and verb items were recalled more often in the declarative than in the interrogative sentences (mood); $F\ (1,\ 120) = 12.716$ and 15.304; $p < .001$ and .0001, respectively. The verbs were recalled more often in the negative than in the active sentences (voice) $F\ (1,\ 120) = 13.495$; $p < .0001$.

B. Process Variables Underlying Recall

1. Analyses of Conditional Error Probabilities Based on the Dependency Between Phrases

The mean depth hypothesis is formulated on the basis of the assumption that the constituents of the sentences are stored and retrieved via binary left-to-right processes involved in each constituent or phrase leading to another following constituent or phrase. An attempt was made in the present study to obtain a logically plausible index of the dependency of recalling a phrase on the recall of the preceding one. In so doing three conditional error probabilities were defined: (a) probability (C1 = VPE : NP_{1c} / NP_{1c} verb phrase errors given subject noun phrase correct), (b) probability (C2 - NP_2E :' VP_c / VP_c object noun phrase error given verb phrase correct), and (c) probability (C2 - NP_2E_1 :' NP_{1c} / NP_{1c} object noun phrase error given subject noun phrase correct) within each block of 4 trials over 8 types of sentences. The probabilities were determined by dividing the conditional error frequency by the total number of correctly recalled preceding items (e.g. VP_e :' NP_1 correct / No. of NP_1 correct). The smaller the number is, the less degree of dependency the error has on the recall of the preceding phrase and thus the greater the degree to which correct recall is dependent on the preceding phrase's correct recall. The mean conditional error probabilities are presented in Table 5.

In order to render the data amenable to parametric analyses, the three dependency measures were then transformed using the formula, $y' = 2 \arcsin \sqrt{y}$ in radian and subjected to an analysis of variance, the results of which are given in Appendix I.

It is interesting to observe that there were significant main effects of the length, response mode, and practice blocks; Fs (1, 120) = 29.749, 39.883, and 159.278, respectively; ps $< .0001$; but no main effect of the mean depth. This means that, in terms of three phrase dependency measures defined, there are stronger relations between phrases of the shorter sentences than in the longer, when recalling in the written mode rather than oral, and in the later trials rather than in the earlier ones in general. There was also

Table 5. Mean Proportions of Conditional Error Probabilities for Phrases as a Function of Experimental Treatments

Length		7 Words				10 Words			
Depth		Low		High		Low		High	
Response Mode		Oral	Written	Oral	Written	Oral	Written	Oral	Written
Block 1	VPE	0.628	0.462	0.811	0.681	0.654	0.567	0.810	0.748
	NP_2E	0.380	0.222	0.214	0.267	0.843	0.619	0.491	0.438
	NP_2E_1	0.460	0.288	0.434	0.382	0.797	0.505	0.603	0.432
Block 2	VPE	0.575	0.267	0.698	0.413	0.403	0.242	0.669	0.368
	NP_2E	0.229	0.045	0.154	0.087	0.539	0.178	0.282	0.200
	NP_2E_1	0.269	0.071	0.338	0.101	0.505	0.222	0.294	0.167

an interaction effect between the length of sentence and the trial block; $\underline{F}$ (1, 120) = 10.729, $\underline{p} < .001$, meaning that such a block effect of strengthening recall is greater for the longer sentences than for the shorter ones.

In examining the validity of assumptions underlying mean depth, the data on the effects of interaction between the dependency measures previously defined and such experimental variables as the mean depth, length of sentences, and response mode are the most relevant ones to examine. First of all, there is the main effect of the three dependency-between-phrase units, with the mean probabilities of .562, .324, and .367 for C_1-VPE, C_2-NP_2E, and C_3-NP_2E_1; E (2, 240) = 88.96, $\underline{p} < .0001$. The Bonferroni t-tests of the contrasts of interest indicate that the correct recall of NP_2 depends more strongly on correct recall of the immediately preceding verb phrase than does correct recall of VP depend on correct recall of NP_1; $\underline{t}$ (240) = 12.706, $\underline{p} < .01$; that the correct recall of NP_2 depends more on that of NP_1 than does recall of VP depend on NP_2, $\underline{t}$ (240) = 9.882, $\underline{p} < .01$; and that there is no difference in dependency between the C_2-NP_2E and C_3-NP_2E_1, $\underline{t}$ (240 = 2.824, $\underline{p} < .01$. In view of the way in which the major source of mean depth was loaded on the verb phrase by introducing discontinuous constituents, this great reduction in recall error can be expected by the time $\underline{S}$s recall the NP_2 phrases.

Second, the dependency of the correct recall of VP on the NP_1 is greater in the low mean depth sentences than in those of high mean depth (i.e., .475 vs. 650 for C_1); whereas dependency of the correct recall of NP_2 on the VP or NP_1 is relatively greater in the high mean depth than in the low mean depth sentences (i.e., .267 and .344 vs. .382 and .390 for C_2 and C_3); $\underline{F}$ (2, 240) = 38.341, $\underline{p} < .0001$. This finding was expected considering the fact that high mean depth was heavily loaded on the VP due to the presence of the discontinuous constituents.

Third, the same pattern of dependencies as described just above exists given the long sentence context (i.e., .483 vs. .651 for C_1, .219 vs. .180 for C_2, and .272 vs. .314 for C_2); whereas, given the short sentence context, the dependency of the correct recall of NP_2 on the VP is much more pronounced in the high mean depth conditions while the dependency of the correct recall of VP on the NP_1 is greater in the low mean depth than in the high mean depth sentences (i.e., .467 vs. .649 for C_1, .545 vs. .353 for C_2, .507 vs. .374 for C_3); $\underline{F}$ (2, 240) = 5.062, $\underline{p} < .007$.

Finally, there was a significant first order interaction between the conditional error probabilities and the length of sentence context; $\underline{F}$ (2, 240) = 22.903, $\underline{p} < .0001$; such that the dependencies of VP correct recall on NP_1 correct recall are just about the same chance level for both the short and long sentences (i.e., .567 vs. .558) whereas increase in the dependency of correct recall of NP_2 on the preceding phrase is greater in the short than in the long sentences (i.e., .200 and .293 vs. .449 and .441). A significant second order interaction involving the conditional error probabilities, mean depth, and the response mode was found; $\underline{F}$ (2, 240) = 5.367, $\underline{p} < .005$. This appears to show that, given the oral response mode, the general pattern of the interaction between the conditional probabilities and the mean depth remains as before; but that, when $\underline{S}$s write their responses, correct recall of VP is greatly enhanced by virtue of the correct recall of NP_1 in the low mean depth sentences as compared to that in the higher mean depth (i.e., .384 vs. .553), and the increased dependencies of the correct recall of NP_2 on the preceding phrases are about the same for both depths (i.e., .266 and .271 vs. .248 and .270).

2. Analyses of Transformational Errors Which Are Simplifications Toward the Kernel

The coding hypothesis is formulated on the basis of the assumed process of coding a given sentence into its base form plus the number of "footnotes" or "tags" that cue the number of steps away from it with respect to three dimensions, mood, voice, and modality. Ss' response errors in verbatim recall of sentences were tabulated in a 10 (8 response types plus omission and unmarkable errors) x 8 (stimulus types as presented) confusion matrices by eight treatment combinations over two blocks of four trials which are presented in Appendix E. Two simplification transformation error measures were derived from the confusion matrices; (a) one based on the paired comparisons of the upper and lower off-diagonal entries of the confusion matrices and (b) one based on grouping conditional error proportions by varying degrees of simplification toward the K-type sentence responses.

If in fact recall errors occur during the assumed coding process, one of the ways in which they would occur should be in the form of simplification. However, this index would be valid provided that the frequency of counter instances of simplification (i.e., lower off-diagonal) is less than that of instances of simplification (i.e., upper off-diagonal), given 28 pairs of entries of 16 confusion matrices. Therefore, the coding hypothesis should be first examined in terms of the proportion of simplification instances, and the conditional error proportions.

A proportion for each treatment combination was determined over two blocks of 4 trials by dividing the number of counter-instances of simplification by the total number of pairs of upper and lower off-diagonal entries which was further subtracted from the expected value of no difference, .50. The conditional error proplrtion was determined by dividing the number of K-type response errors pooled over types of 1-tag transform, types of 2-tag transform, and types of 3-tag transform by respective total number of errors pooled as above. The scores on the two variables are tabulated in Tables 6a and 6b. In order to employ the desired parametric analysis of variance, the same kind of arcsine transformation as used for the conditional error probability of phrase recall was performed on both measures. The results of the analyses of variance are ;presented in Appendices J-1 and J-2.

The results of the analysis clearly shows that the grand mean number of the counter instances of the simplification transformations is significantly far less than .50; $\underline{F}$ (1, 4) = 1846.678, $\underline{p} < .001$. There are no other significant sources of variation in the proportions of counter instances of simplifications, as can be seen in Appendix J-1.

Examination of the analysis of variance table on the conditional error proportion shows that the simplification errors toward the K-type are made more often in oral than in written response, ($\underline{F}$ (1, 4) = 27.977, $\underline{p} < .01$, and in the first block than in the second block of 4 recall trials, $\underline{F}$ (1, 7) = 15.691, $\underline{p} < .01$. The Bonferroni t-tests were performed on the main effect of the three simplification transformational error proportions towards the K-type. The results show that the types of N, P, and Q sentences (the type of 1-tag transform) are more often recalled as the K-type than the types of

Table 6a. Proportions of Counter-simplification Errors Based on Upper Off-diagonal to Lower Off-diagonal Paired Comparisons by Experimental Treatments

Length		7 Words				10 Words			
Depth		Low		High		Low		High	
Response Mode		Oral	Written	Oral	Written	Oral	Written	Oral	Written
Block 1	Number (of 28)	2	3	2	3	2	3	3	7
	Proportion	0.071	0.107	0.107	0.071	0.107	0.107	0.107	0.250
Block 2	Number (of 28)	2	3	6	2	3	1	3	4
	Proportion	0.071	0.107	0.214	0.071	0.107	0.036	0.107	0.143

Table 6b. Conditional Error Probabilities for Varying Degrees of Simplification Toward K-type Sentences by Experimental Treatments

Length		7 Words				10 Words			
Depth		Low		High		Low		High	
Response Mode		Oral	Written	Oral	Written	Oral	Written	Oral	Written
Block 1	1-tag	22/124	21/147	22/118	18/154	28/123	17/132	28/113	17/139
	2-tag	2/112	3/145	7/107	15/147	5/126	9/133	4/104	0/185
	3-tag	0/30	0/46	3/32	2/51	0.43	0.39	2/39	0 63
Block 2	1-tag	16/161	14/191	18/164	20/188	15/180	6/180	24/162	13/190
	2-tag	6/162	1/191	5/160	3/189	2/174	0/184	4/159	0/185
	3-tag	3/56	0/64	2/57	0/62	0/58	0/61	0/57	0/63

NP, PQ, and NQ sentences (the type of 2-tag transform) are as the K-type; $\underline{t}$ (14) = 8.573, $\underline{p} < .01$. The types of 2-tag transform are recalled more often as the K-type than the NPQ type (the type of 3-tag transform), $\underline{t}$ (14) = 3.952, $\underline{p} < .01$; which is less often recalled than the types of 1-tag transform, $\underline{t}$ (14) = 12.525, $\underline{p} < .01$. The results clearly support what is implied with respect to the simplification transformation phenomenon by the coding hypothesis.

CHAPTER V

SUMMARY OF FINDINGS AND DISCUSSION

The analyses of the present data indicate that the validity of assumptions underlying the two major hypotheses, the mean depth and the coding hypothesis seems to be positively conformed, at least in terms of the measures derived in the present study. The outcomes of analyses of recall performance are also shown to be as predicted from the mean depth hypothesis, but are not consistent with predictions made from the coding hypothesis. Since there are a number of findings related to the outcomes that could be predicted from the major hypotheses, details of the findings need to be further examined in the light of some theoretical conceptions related to the major questions.

A. Findings and Assumptions Associated with the Mean Depth Hypothesis

It was predicted that the verbatim recall of low mean depth sentences and phrases should be superior to that of their high mean depth counterparts. The analysis of the data shows that the above prediction is clearly upheld in terms of the recall of full sentences and of verb phrases; it was not upheld in terms of the recall of subject and object noun phrases. This study's finding of superior recall of low versus high mean depth sentences then, can be added to the cumulative findings from a number of studies (Mehler, 1963; Martin and Roberts, 1966; Herriot, 1968; Martin, et al., 1968; Mitchell, 1968, Roberts, 1968; Wright, 1969; Wearing, 1970a, 1972; Bacharach and Kellas, 1971).

The interactive effect on sentence recall of practice block and mean depth was predicted such that the observed depth effect would be disproportionately greater in the second block of four practice trials. This prediction is clearly confirmed and can be taken as strong evidence of the predictive validity of the mean depth hypotheses and verification of this specific feature of the reproduction model of sentence memory.

The nonsignificant mean depth effect for recall of noun phrases indicates that the loci of the mean depth effect on sentence recall can be specified. The fact of verb phrases being inferior in recallability (i.e., $NP_1 = 2.131$, $VP = 1.364$, $NP_2 = 2.052$, $F (2, 240) = 229.509$, $p < .0001$), appears to be responsible for the main effect of mean depth in terms of sentence recall. This interpretation seems to have a certain merit in view of the observation that there is a significant interaction between three phrase recall measures and two mean depths ($F (2, 240) = 41.563$, $p < .0001$), as can be seen in Figure 5a, and that another interaction involving three phrase recall measures, mean depth and length of sentences appeared to be significant ($F (2, 240) = 42.191$, $p < .0001$), as can be seen in Figure 5b. These figures show that the interior recall of verb phrases is accelerated by the sentences being long and of high mean depth. In this connection, it is noteworthy here that the interaction between the three phrase recall measures and the length of sentences was observed to be significant ($F (2, 240) = 114.529$, $p < .00001$), meaning that the inferiority of verb phrases in recall is much greater in the long sentence context. It is also interesting to note that the inferior recallability of verb phrases is, as might be predicted, according to the way in which the mean depth was

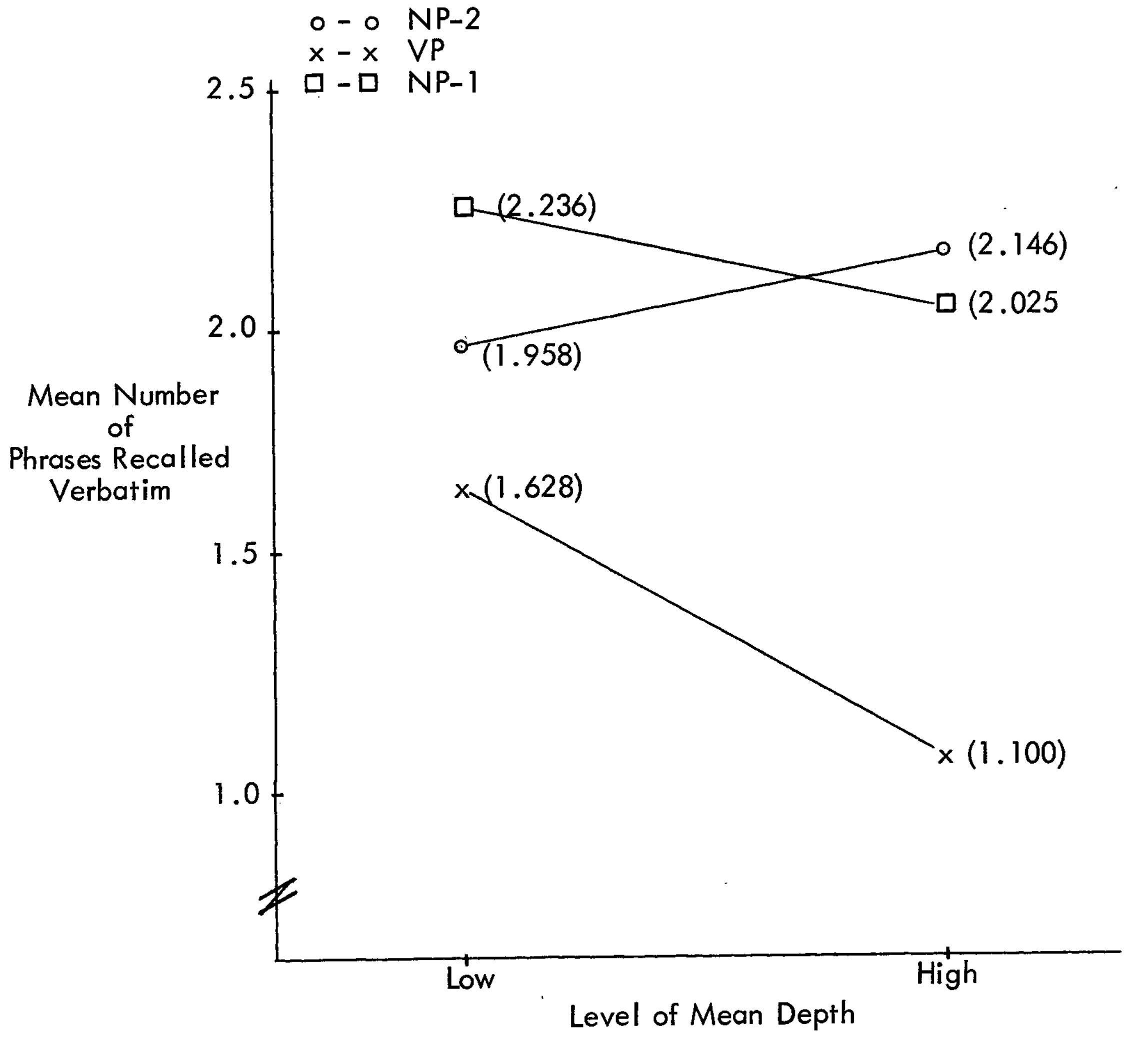


Figure 5a. Mean number of phrases recalled verbatim as a function of mean depth of sentence.

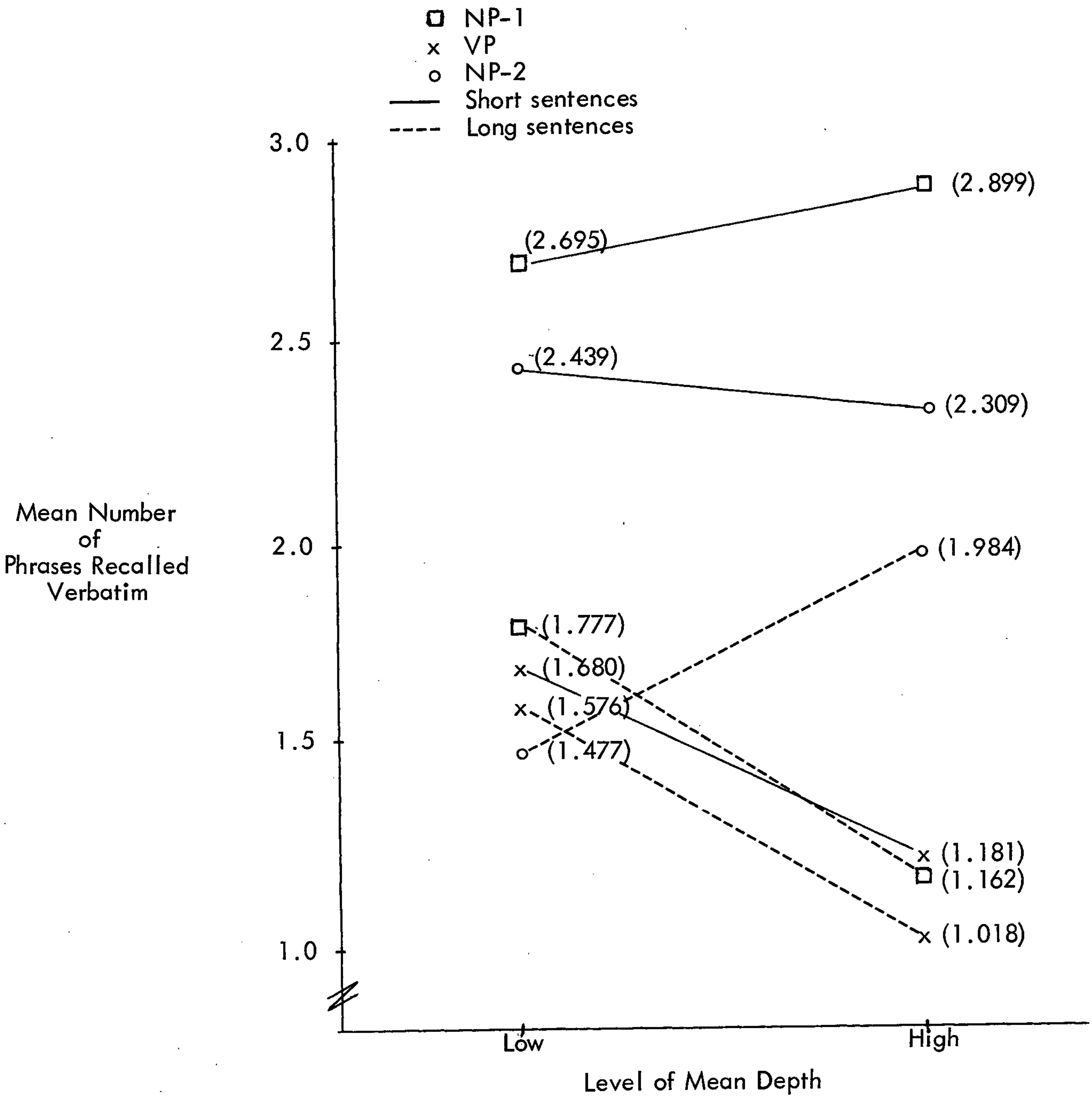


Figure 5b. Mean number of phrases recalled verbatim as a function of mean depth and sentence length.

physically manipulated by introducing an additional discontinuous constituent, as can be seen in Table 1 (p. 46).

In order to examine the validity of the notion of left-to-right binary storage and retrieval processes assumed in the mean depth hypothesis, three conditional error probabilities as measures of dependency between phrases were defined, as mentioned above. The assumption was that, if the mean depth hypothesis is correct, the load placed on memory (cf. Martin, Roberts, and Collins, 1968, p. 560) by having to store all the constituents would be greatest during reproduction of the subject noun phrase. Having correctly recalled the subject noun phrase, there still presumably is a large load on memory which may result in errors in recall of the verb phrase. Having correctly recalled the verb phrase, however, the S's memory load is presumably vastly reduced, so that there should be an increased probability of his correctly recalling the object noun phrase. Since Martin and Roberts' adaptation of Yngve's model is based largely on an assumption of left-to-right binary processing, no such noticeable reduction in error dependency from the subject noun to the object noun phrase as compared to that of the verb to the object noun phrase would be predicted.

The results of the Bonferroni t-tests confirm that both predicted outcomes were observed. The observed ordinality of the outcomes, namely, the recall error dependency of verb phrase being greater than that of the object noun phrase on the immediately preceding phrase, is much more pronounced in the low mean depth than in the high mean depth sentences. This was expected to be the case since it was assumed that the predicate verb phrases of high mean depth sentences impose substantially greater loads on memory because of the discontinuous constituent, than do their low mean depth counterparts. It was felt that this would result in more errors in recall when a S attempts to recall the high mean depth predicate phrases. In order to substantiate that the significant interaction was not due to the artifact of differential recallabilities of the subject noun phrase to begin with, it seems necessary to show that the subject noun phrases are recalled equally well from high and low mean depth sentences. In fact, it was shown that this is the case.

This interaction was further accentuated by the influence of sentence length, as previously mentioned. Thus, the analyses of conditional error probabilities in recalling phrase units seem to indicate that the assumed binary processing in storing and retrieving (or encoding and decoding) sentence materials is a reasonable model of Ss' performance, particularly during the oral recall of longer sentences. However, it should be pointed out that the dependency measure derived in the present study is defined at the phrase unit level and is not the only one to employ. It is perhaps because of the gross nature of the derived measure, that the expected interaction effect on the dependency measures between practice blocks and mean depth did not show up.

B. Findings and Assumptions Associated with the Coding Hypothesis

The logical derivation of sentence recall performance based on the coding hypothesis led to the prediction of a recall hierarchy of sentences because of the fewer transformations ostansibly required in recalling sentences closer to the K-type, as can be seen in Figure 3 (p. 17). Given the three-dimensional representation of a family of sentences in terms of mood, voice, and modality of sentences, two specific predictions were made, that is,

(1) the recallability of K-type > (N-, P-, Q-type) > (NP-, PQ-, NQ-type) > NPQ-type in an ascending order of the numbers of STEPS (or tags) away from the string underlying the declarative, active, and affirmative K-type; and (2) in terms of mood, voice, and modality variations, the recallability of declarative > interrogative, active > passive, and affirmative > negative sentences. It should be mentioned here that the second alternative could well be predicted in terms of the notion of transformational STEPS or tags as well; that is, declarative (a total of 4 tags: 0, 1, 1, and 2 tags for K, N, P, and NP) vs. interrogative (a total of 8 tags: 1, 2, 3, and 3, for Q, NQ, PQ, and NPQ), active (similarly 4 tags) vs. passive (similarly 8 tags), and affirmative (similarly 4 tags) vs. negative (similarly 8 tags).

None of the Bonferroni t-tests of the contrasts made to ascertain the predicted ordinality of the sentence recall hierarchy was significant, though they were partially in the predicted direction (i.e., 1.493, 1.358, 1.300, 1.445). However, the alternative analyses showed that there was a positive finding of a significant mood variation effect, declarative sentences being recalled more often than interrogative; but that the significant effects of modality variation were in the direction contrary to the predictions and that the voice made no difference in sentence recall.

The superior recall of declarative sentences is consonant with Howe's (1970) similar finding, while the lack of difference in numbers of active versus passive sentences recalled mirrors the findings of a number of studies (Perfetti, 1969b; Roberts, 1970; Howe 1970b; Wearing 1970, 1972; Bacharach and Kellas, 1971; André and Kulhavy, 1971; Baker, Prideaux, and Derwing, 1973; James, Thompson, and Baldwin, 1973). The superiority of negative sentences supports Mitchell's (1968) finding. A possible explanation might be that "not" serves a powerful cueing function during encoding and decoding. The negative signalizes a change of meaning, is relatively unambiguous, and seems to be well learned in our culture, all of which might logically contribute to superior recall of negative sentences relative to their affirmative counterparts.

Examination of the interaction effects of type with practice block and response mode variation did not yield any evidence consistent with the predictions from the coding hypothesis. When phrase recall was examined with respect to the recall hierarchy as applied to phrase units, there was also no finding consistent with what might be predicted as being general contextual effects of sentences.

Considering the fact that the locus of mood, voice, and modality variations in types of sentences focusses largely on variations in verb phrases, it should also be possible to make some predictions about verbatim recall of phrases based on mood, voice, and modality differences among the stimulus sentences from whose contexts the phrases are recalled. More specifically the string of symbols underlying declarative sentences, as was discussed earlier, is on the average, more transformational tags away from the K-terminal string than are interrogative sentences. Because of this, and since mood changes result in changes within the predicate verb phrase, one would predict, from the coding hypothesis that verb phrases from declarative sentences will be correctly recalled more often than will verb phrases from interrogative sentences.

Second, voice changes from active to passive will affect both subject and object noun phrases as well as predicate verb phrases; the consistency of grammatical subject and object with logical subject and object ceases to exist, resulting in changes of their grammatical functions. One would therefore predict, from the coding hypothesis, that at least verb phrases would be recalled verbatim more often from active than from passive sentences since the presumed coding operation has to be carried out on the verb phrase in changing active to passive.

Third, affirmative sentences are presumably transformationally closer to the K-terminal string than are negative sentences and changes in modality necessarily affect the predicate verb phrase. One would therefore predict, from the coding hypothesis, that the verb phrase will be recalled verbatim more often from affirmative than from negative sentences, due to the similarity presumed to exist between strings underlying affirmative sentences and the K-terminal string.

The orthogonal contrasts made indicate that the second and third predictions regarding the effects of voice and modality on the recall of the verb phrase are clearly confirmed as expected, while the first prediction on the mood effect is not, though in the predicted direction, due to apparent lack of difference between the effects of the declarative and interrogative contexts. It was worth noting in this connection that the subject noun phrases from declarative sentences are recalled better than from the interrogative, as would be predicted from the coding hypothesis, while the object noun phrases are recalled better from the passive and the negative than from the active and the affirmative sentences. The finding regarding passive sentences appears to suggest that the recallability of grammatical object noun phrases is greater when they are logical subject noun phrases rather than logical object noun phrases. This would be expected on the basis of the assumption that the passive voice tends to topocalize the logical subject noun phrase whereas the topicalized noun phrase in active voice sentences is the logical and grammatical subject. Negation seems to change this, however, which does not follow from the hypothesis. What seems to be the case is that, along with the verb phrase, the negative signalizes or focusses attention on the change in gist of the sentence. It is conjectured that it is this obvious meaning-changing function of the negative which makes the verb and object noun phrases from the negative sentences memorable.

The present recall data of sentences and phrases under examination clearly show that information concerning processes underlying the coding hypothesis can be best obtained not in terms of the verbatim recall of sentence, but rather in terms of a finer measure, such as the most relevant verb phrase recall.

Having evaluated sentence and phrase recall performance as predicted from the coding hypothesis, examination of the assumptions made in the hypothesis is in order. It was shown already that the proportion of counter simplification errors in the verbatim recall of sentences was observed to be phenomenally small as compared to the chance level of .50 irrespective of any other aspects of stimulus sentence variation. It was also shown that the conditional transformational error, given that K-type responses were made most in the sentence types of 1-tag transformation, next most in those of 2-tag transformations, and least in those of 3-tag transformation. In terms of the two measures derived and analyzed, the assumptions underlying the coding hypothesis can be said to be valid ones to make, especially in view of the fact that the type of sentences was manipulated completely independent of other experimental variables.

C. Findings of Other Control Variables

The observation that the longer sentences were recalled less often than the shorter ones can be traced down to the fact that the subject and object noun phrases from the longer sentences were recalled less often than those from the shorter ones, and that the recall of verb phrases is the same for both the longer and shorter sentences. This fact is not unexpected considering the number of constituents in the NP_1 and NP_2 as compared to that of VP as can be seen in Table 1 (p. 46) and the number of lexical items and density of sentences and the three phrases, as can be seen in Table 7. However, it seems important to emphasize that the observation of a non-significant length effect on the VP recall was expected because of no physical difference in the number of lexical items between the 7- and 10-word sentences. Yet the mean depth effect on the VP recall was significant, which might well have been confounded with the variation in the number of lexical items. However, if one examines the unsystematically varying lexical densities of the VP in Table 7, the possibility of accounting for the mean depth effect in terms of the number of lexical items can be ruled out (see Perfetti, 1969b for relevance of lexical density rather than the number of lexical items).

Table 7. Number of Lexical Items and Lexical Density of Sentences and Phrases by Experimental Variables

			NP_1	VP	NP_2	# Lexical Items	Sentence Lexical Density	$\bar{x}$ Lexical Density A+P Sentences
7 Words	Low $\bar{D}$	Active	2* / .29**	2 / .29	2 / .29	6	.87	.79
		Passive	2 / .29	2 / .29	1 / .14	5	.71	
	High $\bar{D}$	Active	1 / .14	3 / .43	2 / .29	6	.87	.79
		Passive	1 / .14	3 / .43	1 / .14	5	.71	
10 Words	Low $\bar{D}$	Active	3 / .30	2 / .20	4 / .40	9	.90	.85
		Passive	3 / .30	2 / .20	3 / .30	8	.80	
	High $\bar{D}$	Active	4 / .40	3 / .30	2 / .20	9	.90	.85
		Passive	4 / .40	3 / .30	1 / .10	8	.80	

* Number of Lexical words

** Lexical density of phrase $\left(\frac{\text{\# of lexical words in phrase}}{\text{\# of words in sentence}}\right)$

The sentences and the subject and object noun and predicate verb phrases are recalled more often in the written mode (particularly in the last four trials) than in the oral mode and in the early trials in general. Furthermore, response mode and practice blocks interact in such a way that the written recall of sentences and NP_1, VP, and NP_2 improves more markedly than the oral recall of these.

D. Pivotal Word Recall and the "Gist" Hypothesis

Unlike what might be tapped by the verbatim recall of sentences and phrases, the recall of pivotal words of a sentence would be expected to bring out more information of unput sentence materials stored and processed by Ss in addition to those that can be retrieved in accordance with the grammatical requirements. That is to say, the scoring criteria used for pivotal words recall lowered the recall threshold for obtaining a probably deeper (or conceptual) level of what is stored. This scoring of Ss' response protocols was further encouraged by the investigator's informal probe of the strategies for recall that might be used by the Ss. There emerged two classes of Ss out of diversely different groups of Ss who reacted to the probe questions. For example, the first of two Ss whose recall performance was the best of all Ss, said that he tried to draw a mental picture (image) of what was described by the set of eight sentences, and then to use it in remembering or to decode it back to verbal sentences; the second said that he drew eight different mental pictures instead of one. Some others mentioned that they tried to use recency or primacy strategies.

It is quite conceivable that what Ss remember is the "gist" (or image) of a sentence rather than a grammatically well formed sentence and Ss try to encode and decode sentence constituents in terms of paradigmatic features of subject and object nouns and syntagmatic relations between them. If this is close to the psychological reality (as is suggested by Bobrow, 1970; Reid, 1974), then the recall of the pivotal words should be predicted on the basis of the number of factors that determine the complexity of various features entering the image representing what is described in the sentence. Two of the factors might well be the number of lexical items or density and the mean depth of the sentence.

Mor specifically, according to the "gist" hypothesis, if one comprehended the messages (or meanings) of the input sentences, he should be able to recall all three pivotal words equally well; however, if one's comprehension is not quite complete, this would be indicated by the recall level of verbs, which complete the contextual meaning of sentences provided that subject and object nouns are remembered (Reid, 1974). In the present study, the critical role of verbs in determining and comprehending the meaning of sentences is indicated by the fact that the STEP listening score presumably measuring the comprehension of what Ss hear is significaly associated with the correct recall of verbs only. The present data under study shows the equal recallability of the three pivotal words. That is to say, the overall means of pivotal words recalled were 2.464, 2.537, and 2.467 for the subject noun, verb, and object noun, respectively, the difference between them being nonsignificant, $F\ (2, 240) = 2.705$, $p < .069$.

Second, it would be expected according to the "gist" hypothesis (a) that pivotal words from the shorter sentences lexical density in the present study, would be recalled more often than those from the longer sentences and (b) that those from the low mean depth sentences would be recalled more often than those from the high mean depth. Of these two, the first one is confirmed, $\underline{F}$ (1, 120) = 48.956, $\underline{p} < .0001$, although the second is not, $\underline{F}$ (1, 120) = 2.977, $\underline{p} < .087$. However, it is clearly shown that the number of lexical items or density interacts with the mean depth of sentences in affecting the recall of pivotal words.

The subject nouns were most often recalled given short sentences (i.e., 3.078, 2.607, 2.754 for N_1, V, N_2), while the verbs were most often recalled given long sentences (i.e., 1.850, 2.467, 2.180 for N_1, V, N_2); $\underline{F}$ (2, 240) = 119.364, $\underline{p} < .0001$. These findings from the present study can be better understood if one examines some further interaction phenomena. The ordinality in the recall of subject nouns, verbs, and object nouns is different depending upon the level of mean depth, or the interaction between the sentence length and mean depth, as can be seen in Figures 6a and 6b.

Third, it was found that the interaction of mean depth with the recall of N_1, V, and N_2 is significant, $\underline{F}$ (2, 240) = 42.503, $\underline{p} < .0001$; and that the interaction of length and mean depth of the sentences with the recall of N_1, V, N_2 is also significant, $\underline{F}$ (2, 240) = 68.832, $\underline{p} < .0001$. The meaning of these interactions can be further probed by examining the entries in Table 7 (p. 91) and what is presented in Figures 6a and 6b, that is to say, a pattern of the number of lexical items derived from Table 7 in a 2 (depth) x 3 (N_1, V, N_2) would predict very closely what is shown in Figure 6a;

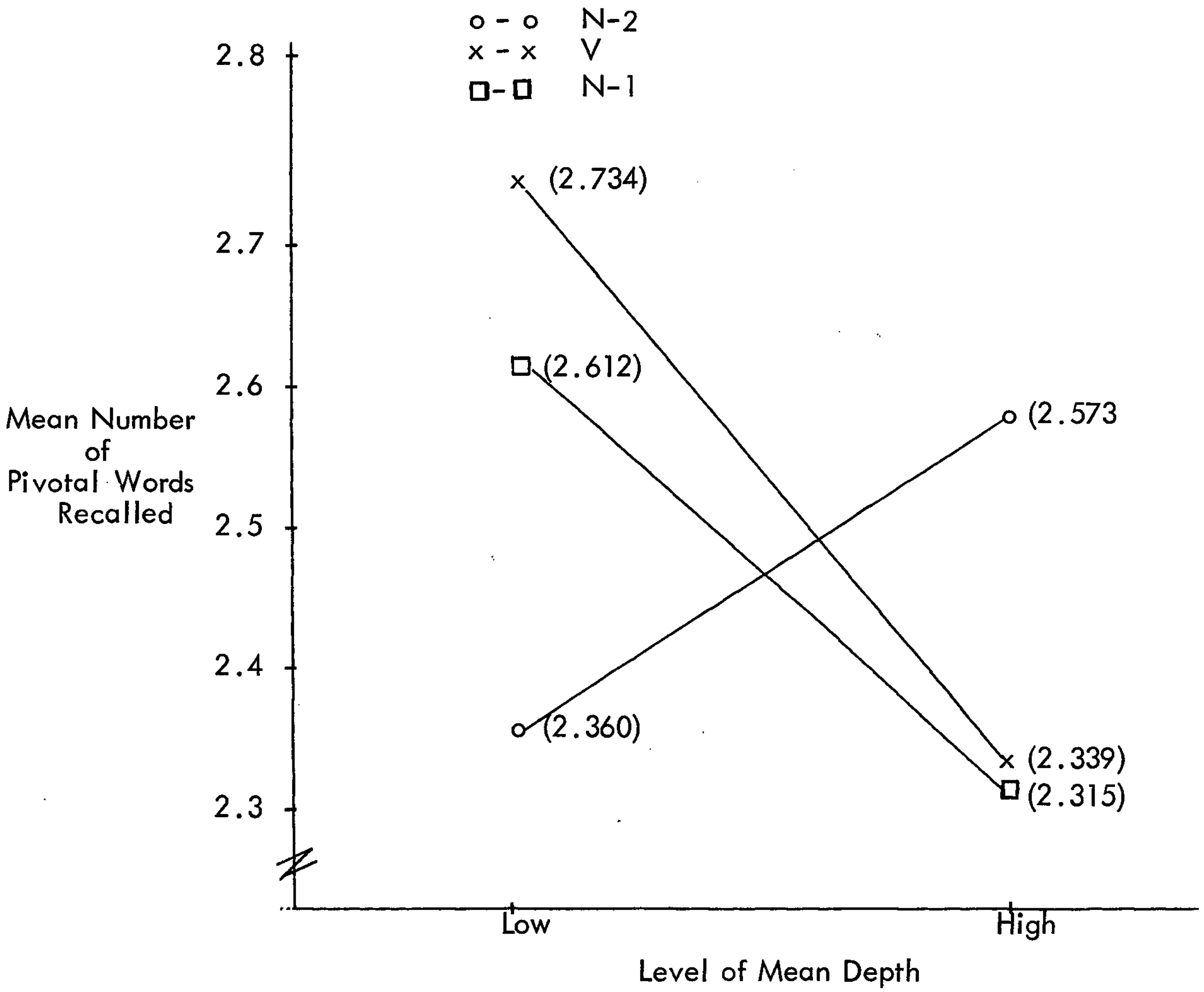


Figure 6a. Mean number of pivotal words recalled as a function of mean depth of sentence.

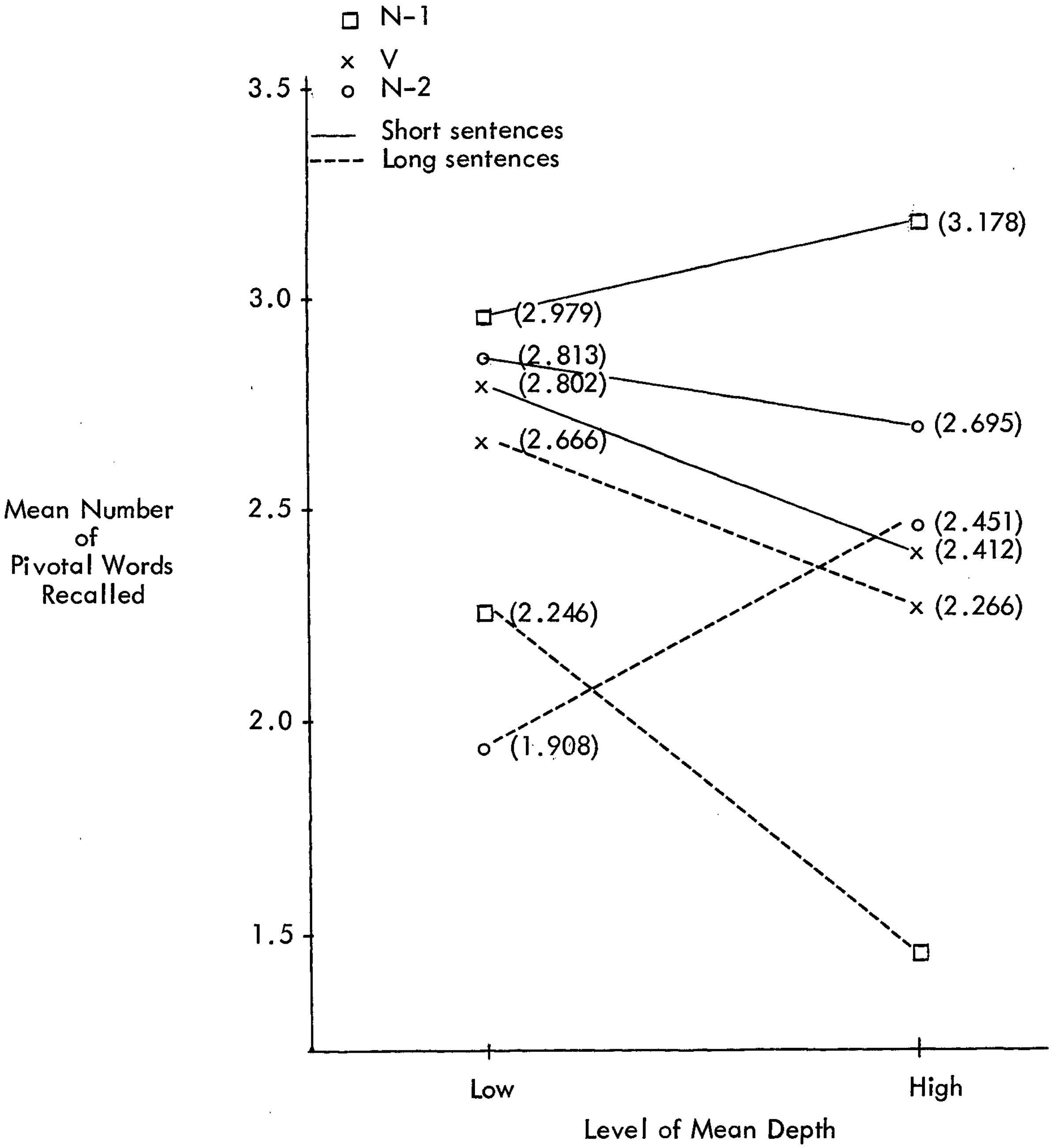


Figure 6b. Mean number of pivotal words recalled as a function of mean depth and sentence length.

and another pattern in a 2 (length) x 2 (depth) x 3 (N_1, V, N_2) would predict the outcomes plotted in Figure 6b. Thus, it can be said that the "gist" hypothesis is supported by the evidence presented on the interacting effect of the number of lexical items or density and the mean depth of sentences, as defined in the present study. In this respect, it is highly instructive to suggest that Ss, when asked to recall what they heard, might try to pick up first pivotal words in trying to get the "gist" of sentences, and then add further necessary qualifiers and determiners to complete syntactic forms of phrases and eventually sentences. This possibility is suggested by the differential interactions of these constituents with blocks and sentence length, as can be seen in Figure 7.

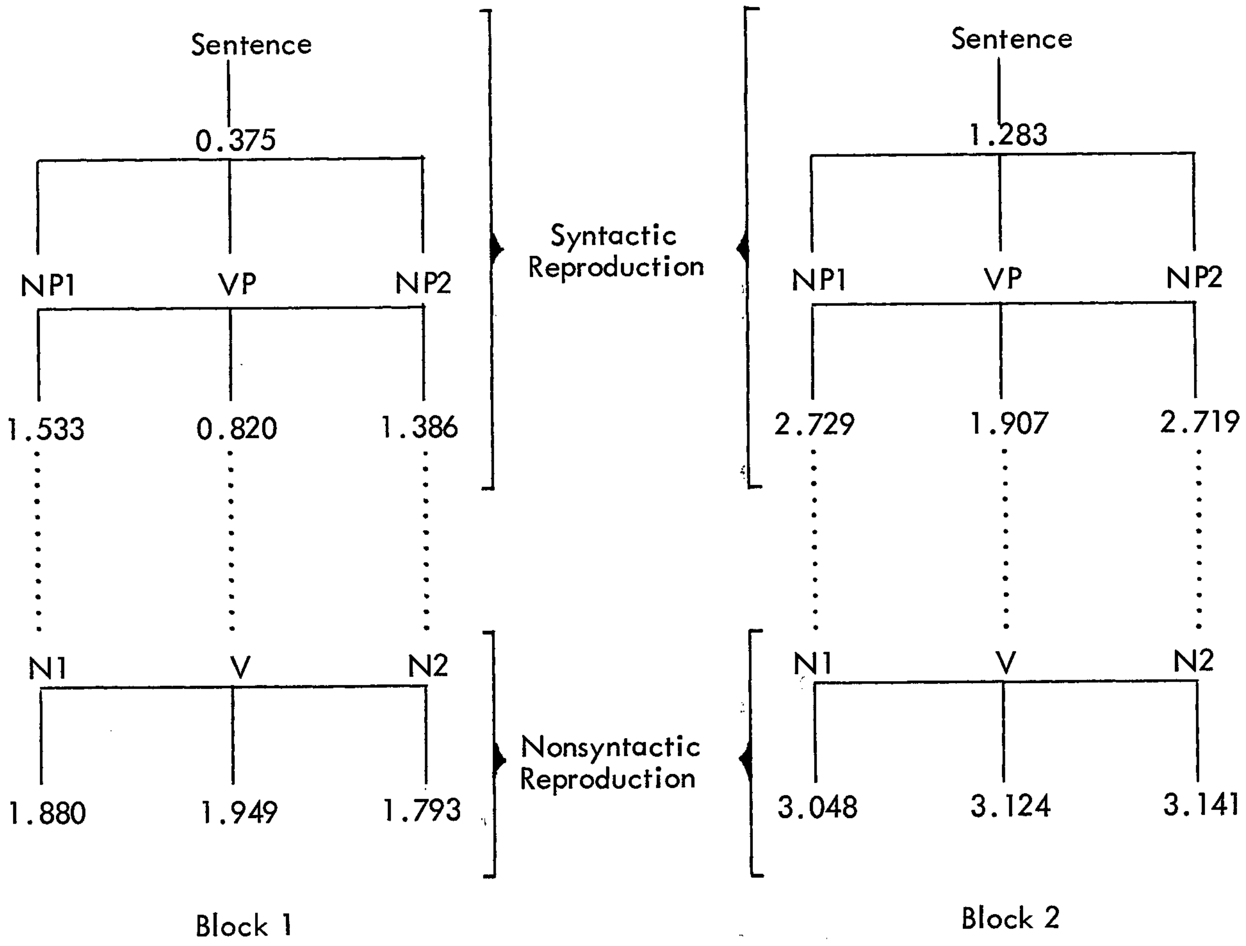


Figure 7. The amount of syntactic and nonsyntactic reproduction of sentence constituents as a function of trial blocks.

CHAPTER VI

SUMMARY

The purpose of the present study was to determine how selected stimulus and response variables affect the recall of sentences, phrases, and pivotal words. Of particular interest were the effects on recall of the mean depth (Martin and Roberts, 1966), type (Miller, 1962) and length (Perfetti, 1969a, b), of the to-be-remembered sentences, the mode of response and amount of practice at the task. There is some negative evidence regarding the effects on sentence recall of mean depth (Perfetti, 1969a and b; Perfetti, and Goodman, 1971) and the transformational type of the stimulus sentence (Martin and Roberts, 1966; Roberts, 1968; Mathews, 1968; Perfetti, 1969a; Wearing, 1969; Howe, 1970; André and Kulhavy, 1971; Bacharach and Kellas, 1971; James, Thompson, and Baldwin, 1973). There is very little evidence with regard to the effects on sentence recall of differential sentence length (Perfetti, 1969a, b). There are apparently no studies which have been concerned with differential response modes or repeated practice in the sentence recall context. It was hoped, therefore, that by factorially combining these variables in this study a clearer picture of their various roles in the sentence reproduction process would emerge.

An attempt was made to examine the assumptions about sentence reproduction process implied in Martin and Roberts' (1966) reformulation of Yngve's (1960) model of sentence production. An attempt was also made to examine processes implicit in Miller's (1962) and Chomsky's (1957) notions about how sentences are recalled. Verbatim phrase and pivotal word recall, the dependencies among phrases, and transformational errors were the variables examined in the hope they would be revealing of mechanisms involved in the sentence recall process. A general overview of the results of the present study would seem to indicate that the following mechanisms are involved in the sentence recall process.

First, the S initially recalls the "gist" of what he has heard with "gist" being represented at least by his early recall of nouns, and in many cases verbs. Pinpointing the nouns and the verb as serving the critical meaning-carrying function is in line with Reid's recent (1974) conceptualization of the critical function these elements play in capturing the meaning of a sentence. The findings in the present study also support Bobrow's (1970) contention that people remember the meaning of a sentence ven when they do not exactly remember the sentence itself.

Interestingly the two nouns and the verb are recalled better than the phrases of which they are a part with this being most noticeable early in the task (see Figure 7, p. 98). In his recent review of experimental psycholinguistics Johnson-Laird (1974) concludes that evidence from a number of studies points to the clause rather than the word as being the most meaningful unit for cueing memory. Certainly the two larger grammatical units in the sentences of the present study (i.e., the sentence and the phrases) do not serve this function, at least in early learning. It seems possible that in a sentence recall task, as opposed to sentence verification recognition, or a number of other memory tasks, when the S hears what he is to recall, rather than sees it, he is constrained by limited memory capacity and is able to focus only on the minimal units encapsulating the sentence's meaning . That appears to be what is happening in this study. It is also clear that there is no evidence in this study to support the notion that concrete subject nouns and the more abstract object nouns which often accompany transitive verbs (Clark, 1965) are differentially recalled. The three pivotal words, in combination, appear to serve the initial memory-cueing function for Ss in this study.

Once the S has recalled the "gist" of what he has heard, he seems to be able to add more and more lexical items, following the rules of the grammar, as is indicated by the increasing structure of his responses. As the S's recall of phrases increases his responses seem to indicate that the image or "gist" which cues his recall is related largely to the noun phrases. When the phrase recall is poorest, this seems to be a function of lexical density and surface structure complexities of the phrases, and to the interaction of these variables with sentence length. The difficulties the S has with verb phrase recall were largely attributable to the fact that it was in this phrase that surface structure was most complex, due to the discontinuous constituents. These findings were predicted from the mean depth hypothesis.

Phrase recall appeared to reflect the influence of left-to-right binary processing during encoding and decoding. The greater dependency of correct recall of verb phrases than object noun phrases on correct recall of the subject noun phrase, particularly in sentences of low mean depth, supports this assumption of the direction of processing which is central to Martin and Roberts' hypothesis.

The assumptions underlying the coding hypothesis appear to be partially confirmed in terms of phrase recall. The S seems to encode and decode verb phrases from active and affirmative sentences with more facility than those from passive or negative sentences, as the coding hypothesis might predict. Verb phrases, of course, are the most changed during the transformational process, so that the basis for this prediction seems to be entirely logical. Where verb phrases are changed the least is in changing an interrogative to a declarative sentence, or vice versa. As might be expected logically, though it does not follow from the coding hypothesis, verb phrases from these two sentence types were equally well recalled.

Superior recall of predicate verb and object noun phrases from negative versus active sentences is also left unexplained in terms of the transformational tag notion of memory processing. A possible explanation of the result is that negation is a strong signal of meaning change which facilitates recall of the predicate verb and object noun phrases. The memory-facilitating effect of negation is proposed as an explanation for the superior recall of negative sentences as well.

When the S begins to recall sentences, his errors are clearly predicted from the notion of transformational tags, which is supportive of this assumption underlying the coding hypothesis. When the S errs in recall of the given sentence type, the errors are generally simplifications toward the K sentence type. The apparent memory-depressing effect increased density has on recall of sentence constituents seems to be partially accounted for in Martin and Roberts' reformulation of Yngve's speech production model in that lexically dense constituents in this study are frequently those which entail the greatest number of listener expectations and speaker commitments relative to the other constituents in the same sentence.

The exact roles played by lexical density and mean depth remain unclear. While it is unfortunate, it also seems to be true that these two variables cannot be completely orthogonally varied while also systematically varying sentence length. Further study of the relationships among these variables seem meritted. It also seems worthwhile to investigate further the effects on sentence, phrase, and pivotal item recall of differential response modes. Recall generally was facilitated in the written response mode and it is conjectured that this is due to the increased opportunities available in this mode for the S to concretely organize his responses, as well as to the greater number of cues available. There would seem to be considerable practical value which might derive from examining the effects on memory of combining these response modes since little is known in this area. Finally, in response to Slobin's (1968) plea, it would seem worthwhile to examine the effects on recall of also systematically varying the stimulus presentation mode. The S generally has much fewer problems with the various aspects of sentences when he writes his responses, rather than says them. Again, the role of this variable in the sentence memory process is undefined and open to conjecture.

It seems fair to say that many critical aspects of Martin and Roberts' (1966) psychology of listening to and reproducing ordinary English sentences have been substantiated by the results of the present study. These results also substantiate Fodor, Bever, and Garrett's (1974) claim that the notion of a sentence family (Miller, 1962) is psychologically real, as well as their conclusion that it is unlikely that the integration of sentences is governed by the computational processes specified by transformational grammars (p. 273). It is felt that the present study is a step in the direction toward providing evidence which eventually may help to account for the role played by structural features in sentence processing, although much more study is warranted.

Bibliography

Anderson, B. The short-term retention of active and passive sentences. Unpublished Ph.D. Dissertation, 1963, The Johns Hopkins University.

André T., and R.W. Kulhavy. Category clustering in the free recall of sentences. Psychological Reports, 1971, 29, 631-634.

Bacharach, V.R. and G. Kellas. Phrase versus base structure effects on short-term retention. Journal of Verbal Learning and Verbal Behaviour, 1971, 10, 171-175.

Baker, W.J., G.D. Prideaux, and B.L. Derwing. Grammatical properties of sentences as a basis for concept formation. Journal of Psycholinguistic Research, 1973, 2, 201-220.

Beatty, J., and J. Borree. Effects of word class on the recall of sentences. Journal of Verbal Learning and Verbal Behaviour, 1971, 10, 604-607.

Bever, T.G. A survey of some recent work in psycholinguistics. In W.J. Plath (ed.), Specification and Utilization of a Transformational Grammar : Scientific Report Number Three. Yorktown Heights, N.Y.: Thomas J. Watson Research Center, International Business Machines Corporation, 1968.

Bobrow, S.A. Memory for words in sentences. Journal of Verbal Learning and Verbal Behaviour, 1970, 9, 363-372.

Cermak, L.S. Human Memory. New York: Ronald Press, 1972.

Chomsky, N. Syntactic Structures. The Hague : Mouton, 1957.

__________. Aspects of the Theory of Syntax. Cambridge, Mass.: M.I.T. Press, 1965.

Clark, H.H. The prediction of recall patterns in simple active sentences. Journal of Verbal Learning and Verbal Behaviour, 1966, 5, 99-106.

__________. Linguistic processes in deductive reasoning. Psychological Review, 1969, 76, 387-404.

Clifton, C., and P. Odom. Similarity relations among certain English sentence constructions. Psychological Monographs, 1966, 80 (Whole No. 613).

Coleman, E.B. Learning of prose written in four grammatical transformations. Journal of Applied Psychology, 1965, 49, 332-341.

Dever, R.B. Before transformational grammers: The case for data gathering. Journal of Special Education, 1971, 5, 119-126.

Duker, S. Listening : Reading. New York, Scarecrow Press, 1966.

Fillenbaum, S. On the use of memorial techniques to assess syntactic structures. Psychological Bulletin, 1970, 73, 231-237.

Fodor, J.A., T.G. Bever, and M.F. Garrett. The Psychology of Language. New York: McGraw-Hill, 1974.

Herriot, P. The comprehension of sentences as a function of grammatical depth and order . Journal of Verbal Learning and Verbal Behaviour, 1968, 7, 938-941.

Horowitz, L., and L. Prytulak. Redintegrative memory. Psychological Review, 1969, 76, 519-531.

Howe, E.S. Effects of grammatical qualifications on judgments of the depth and of the anxiety arousel potential of interpretive statements. Journal of Consulting and Clinical Psychology, 1970a, 34, 159-163.

__________. Transformation, associative uncertainty, and free recall of sentences. Journal of Verbal Learning and Verbal Behavior, 1970b, 9, 425-431.

Huxley, R., and E. Ingram. Language Acquisition : Models and Methods. New York, Academic Press, 1971.

James C.T. Theme and imagery in sentence recall. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 205-211.

__________. J.G. Thompson and J.M. Baldwin. The reconstructive process in sentence memory. Journal of Verbal Learning and Verbal Behavior, 1972, 12, 51-63.

Jarvella, R.J. Syntactic processing of connected speech. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 409-416.

Johnson-Laird, P.N. Experiemental psycholinguistics. Journal of Psycholinguistic Research, 1974, , 135-159.

Katz, J.J., and P.M. Postal. An Integrated Theory of Linguistic Descriptions. Cambridge, Mass. : M.I.T. Press, 1964.

Levelt, M.J.M. The scaling of syntactic relatedness : A new method in psycholinguistic research. Psychonomic Science, 1969, 17, 351-352.

Mandler, G., and J. Mandler. Serial position effects in sentences. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 195-202.

Martin, E., and K.H. Roberts. Grammatical Factors in sentence retention. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 211-218.

__________, K.H. Roberts and A.M. Collins. Short-term memory for sentences. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 560-566.

__________, and D.A. Walker. Subject uncertainty and word-class effects in short-term memory for sentences. Journal of Experimental Psychology, 1969, 80, 47-51.

Martin, J.G. Hesitations in the speaker's production and listener's reproduction of utterances. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 903-909.

Matthews, W.A. Transformational complexity and short term recall. Language and Speech, 1968, 11, 120-128.

Mehler, J. Some effects of grammatical transformations on the recall of English sentences. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 346-351.

Melton, A.W., and E. Martin (eds.). Coding Processes in Human Memory. New York: Wiley and Sons, 1972.

Miller, G.A. The magical number seven, plus or minus two. Psychological Review, 1956, 63, 81-97.

__________. Some psychological studies of grammar. American Psychologist, 1962, 17, 748-762.

__________, and N. Chomsky. Finitary models of language users. In R.D. Luce, R.R. Bush, and E. Galanter (eds.), Handbook of Mathematical Psychology, Vol. II, New York: Wiley, 1963.

__________, E. Galanter and K.H. Pribram. Plans and the Structure of Behavior. New York: Holt, Reinhart and Winston, 1960.

Mitchell, D.L. The effects on sentence recall of varying age, mean depth, and sentence type. Unpublished M.A. thesis, The University of British Columbia, 1968.

Paivio, A. Paired-associate learning and free recall of nouns as a function of concreteness, specificity, imagery, and meaningfulness. Psychological Reports, 1967, 20, 239-245.

__________. Mental imagery in associative learning and memory. Psychological Review, 1969, 76, 241-263.

__________, J.C. Yuille and S.A. Madigan. Concreteness, imagery, and meaningfulness values for 925 nouns. Journal of Experimental Psychology Monograph Supplement, 1968, 76, (1, Pt. 2).

Perfetti, C.A. Sentence retention and the depth hypothesis. Journal of Verbal Learning and Verbal Behavior, 1969a, 8, 101-104.

__________. Lexical density and phrase structure depth as variables in sentence retention. Journal of Verbal Learning and Verbal Behavior, 1969b, 8, 719-724.

__________, and D. Goodman. Memory for sentences and noun phrases of extreme depth. Quarterly Journal of Experimental Psychology, 1971, 23, 22-23.

Postman, L. A pragmatic view of organization theory. In E. Tulvig, W. Donaldson, Organization of Memory. New York, Academic Press, 1972.

Reber, A.S. Implicit learning of artificial grammars. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 855-863.

__________, and R.B. Millward. Event observation in probability learning. Journal of Experimental Psychology, 1968, 77, 317-327.

__________. On psycho-linguistic paradigms, Journal of Psycholinguistic Research, 1973, 2, 289-319.

Reid, L.S. Toward a grammar of the image. Psychological Bulletin, 1974, 81, 319-334.

Roberts, K.H. Grammatical and associative constraints in sentence retention. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 1072-1076.

Savin, H.B.,and E. Perchonock. Grammatical structure and the immediate recall of English sentences. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 348-353.

Schlesinger, I.M. Sentence Structure and the Reading Process. The Hague: Mouton, 1968.

Sequential Tests of Educational Process : Listening, Form 2B, Princeton, N.J., Educational Testing Service, 1957.

Slobin, D.I. Psycholinguistics. Glenview, Ill. : Scott, Foresman, 1971.

__________. Grammatical transformations and sentence comprehension in childhood and adulthood. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 219-227.

__________. Recall of full and truncated passive sentences in connected discourse. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 876-881.

Tulving, E., and W. Donaldson. Organization of Memory. New York, Academic Press, 1972.

__________, and D.M. Thompson. Retrieval processes in recognition memory: Effects of associative context. Journal of Experimental Psychology, 1971, 87, 116-124.

Underwood, B.J. Are we overloading memory? In A.W. Melton and E. Martin (eds.), Coding Processes in Human Memory, New York, Wiley and Sons, 1972.

Wang, M.D. The role of syntactic complexity as a determiner of comprehensivility. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 398-404.

Wanner, E. On remembering, forgetting, and understanding sentences: A study of the deep structure hypothesis. Unpublished doctoral dissertation, Harvard University, 1968.

Wearing, A.J. The storage of complex sentences. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 21-29.

__________. Word class and serial position in the immediate recall of sentences. Psychonomic Science, 1971, 25, 338-340.

__________. Remembering complex sentences. Quarterly Journal of Experimental Psychology, 1972, 24, 77-86.

Winer, B.J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1971.

Wright, P. Transformations and the understanding of sentences. Language and Speech, 1969, 12, 156-166.

Yngve, V.H. A model and a hypothesis for language structure. Proceedings of the American Philosophical Society, 1960, 104, 444-466.

Appendices

APPENDIX A

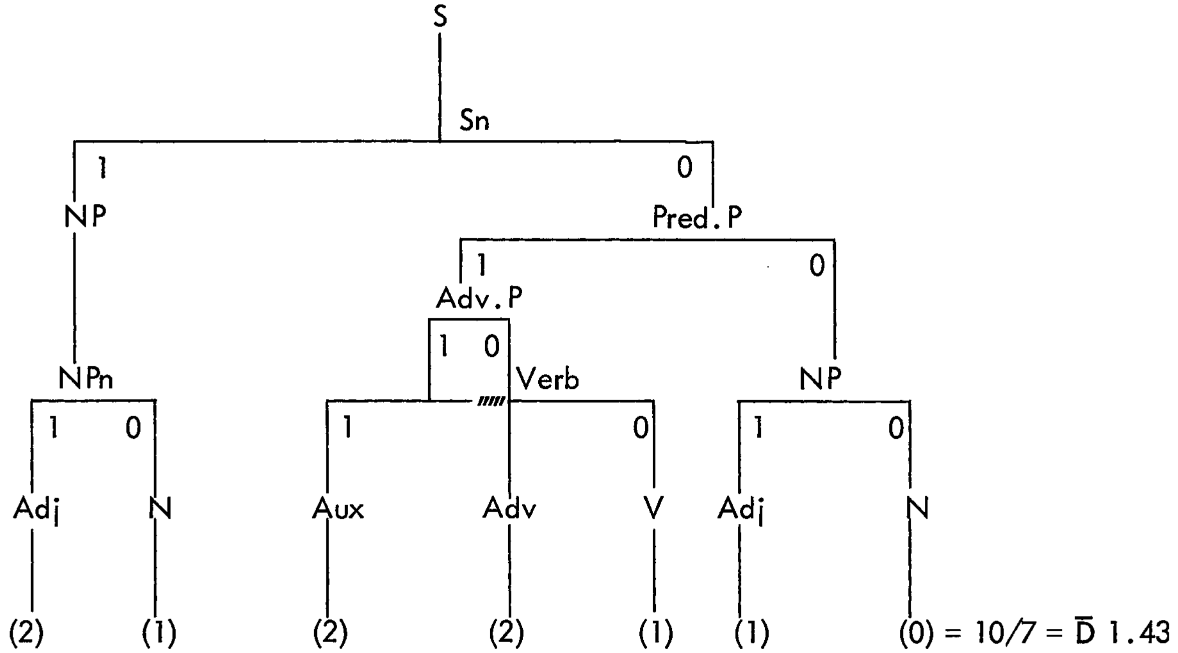
A Schematic Diagram of the Experimental Design

					SENTENCE TYPE									
					1	2	3	4	5	6	7	8		
					K	N	P	NP	KQ	NQ	PQ	NPQ		
				Trials	1-8	1-8	1-8	1-8	1-8	1-8	1-8	1-8	9	
Length	Depth	Response Mode	Group	Set										n = 8
1 (7 wks)	1 (low)	O	1	1									1	8
			2	2									2	8
		W	3	1									3	8
			4	2									4	8
	2 (high)	O	5	1									5	8
			6	2									6	8
		W	7	1									7	8
			8	2									8	8
2 (10 wks)	1 (low)	O	9	1									9	8
			10	2									10	8
		W	11	1									11	8
			12	2									12	8
	2 (high)	O	13	1									13	8
			14	2									14	8
		W	15	1									15	8
			16	2									16	8
													N	= 128

Appendix B

Binary Tree Diagrams of All Stimulus Sentences Used

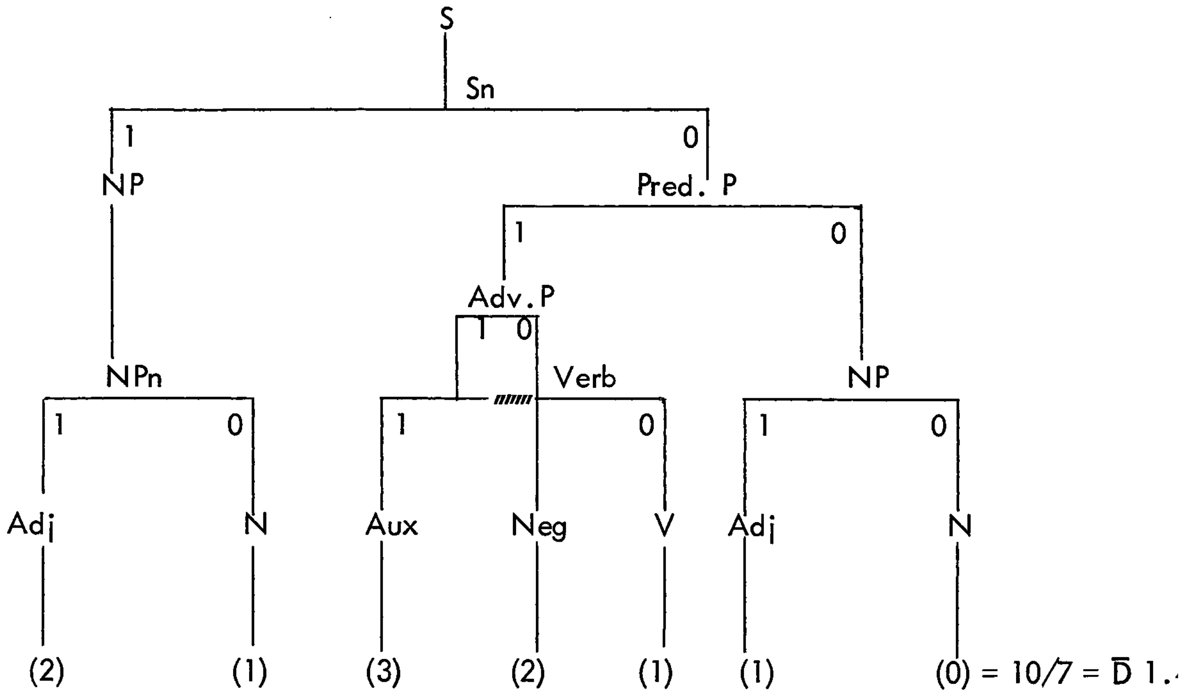
KI



Set one - Active children were already boarding large buses.

Set two - Patient workers were finally repairing broken dikes.

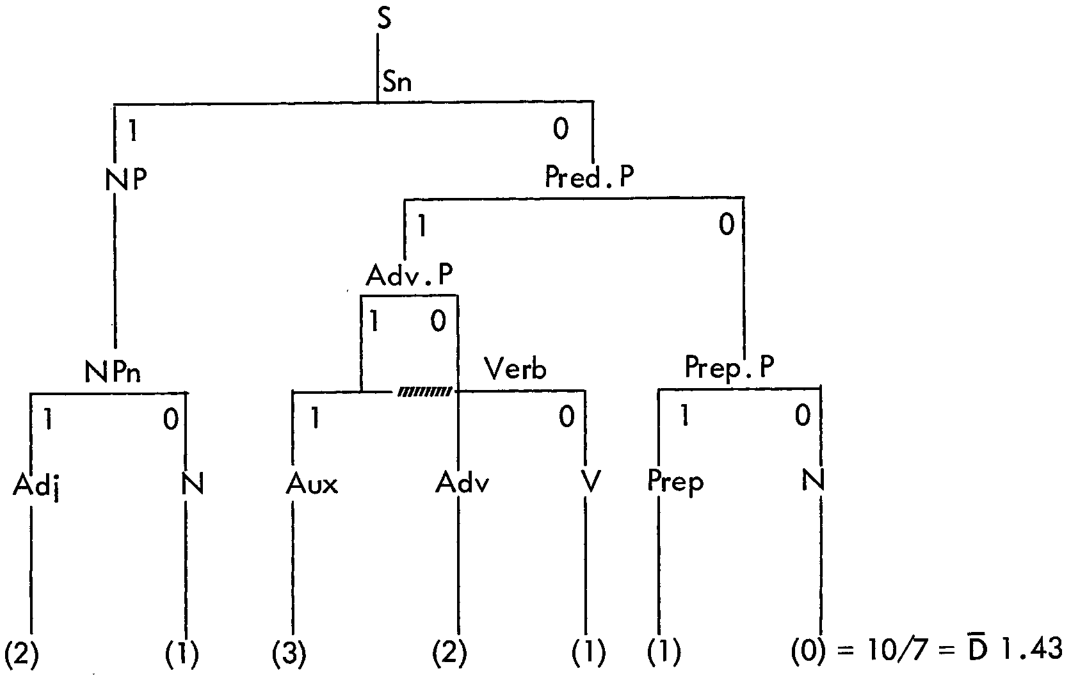
NI



Set one - Experienced farmers are not planting leafy vegetables.

Set two - Conservative residents are not joining costly clubs.

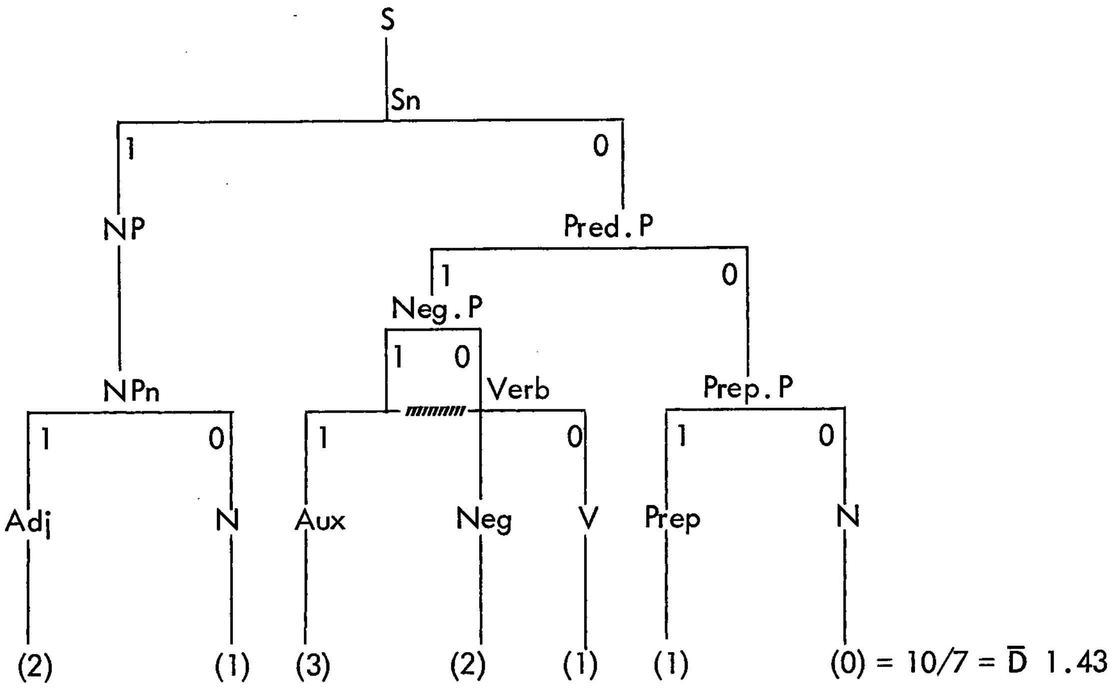
PI



Set one - Small houses were soon purchased by couples.

Set two - Rugged mountains were often eroded by rains.

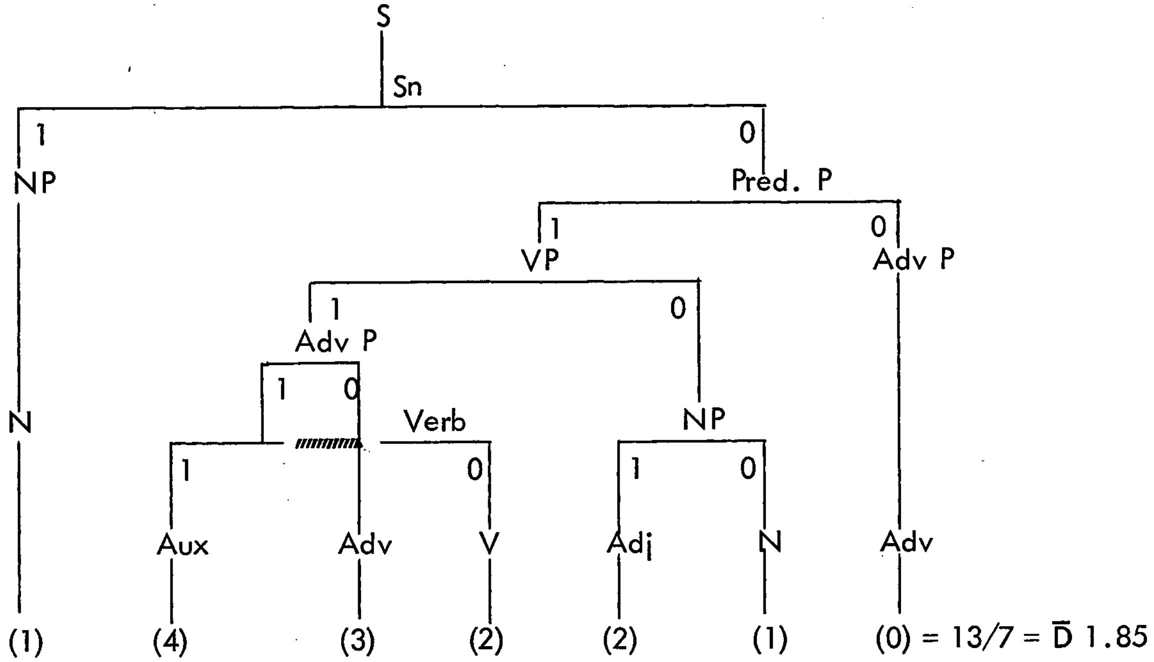
NPI



Set one - Bulky machines were not fixed by mechanics.

Set two - Imported cheeses were not stocked by merchants.

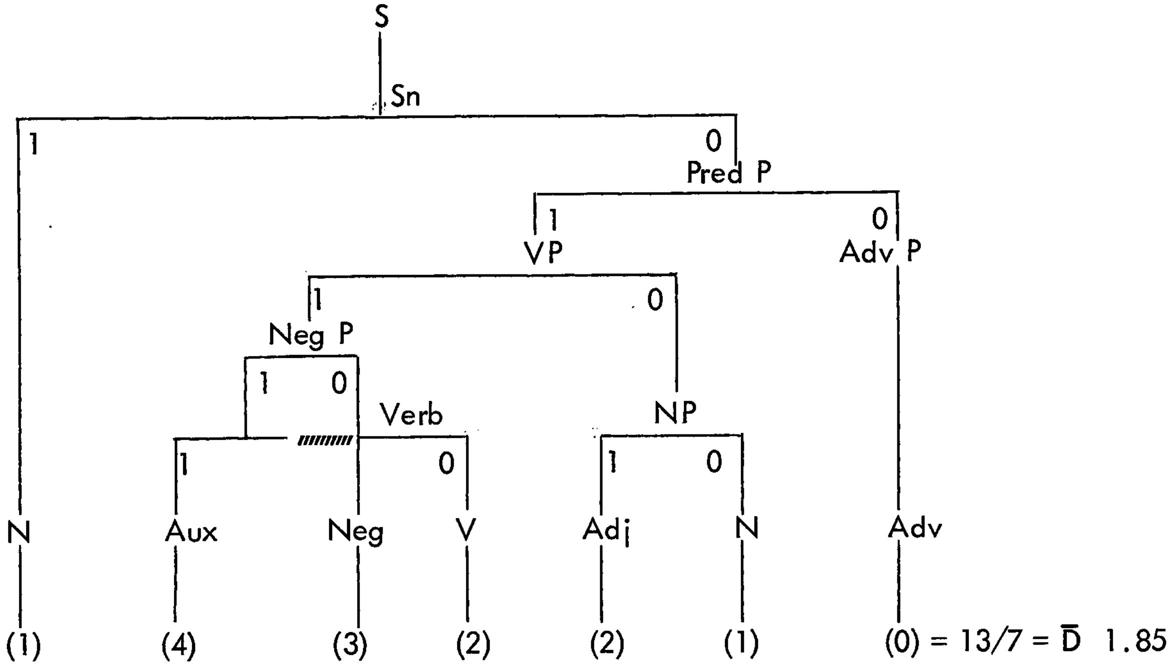
KII



Set one - Children were already boarding large buses eagerly.

Set two - Workers were finally repairing broken dikes carefully.

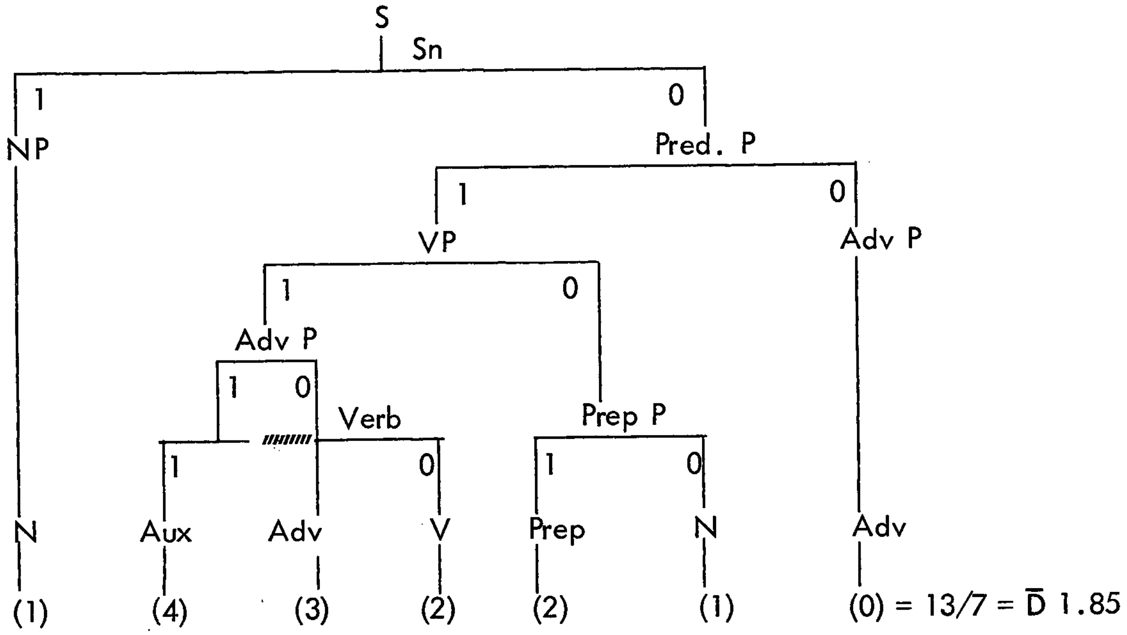
NII



Set one - Farmers are not planting leafy vegetables willingly.

Set two - Residents are not joining costly clubs readily.

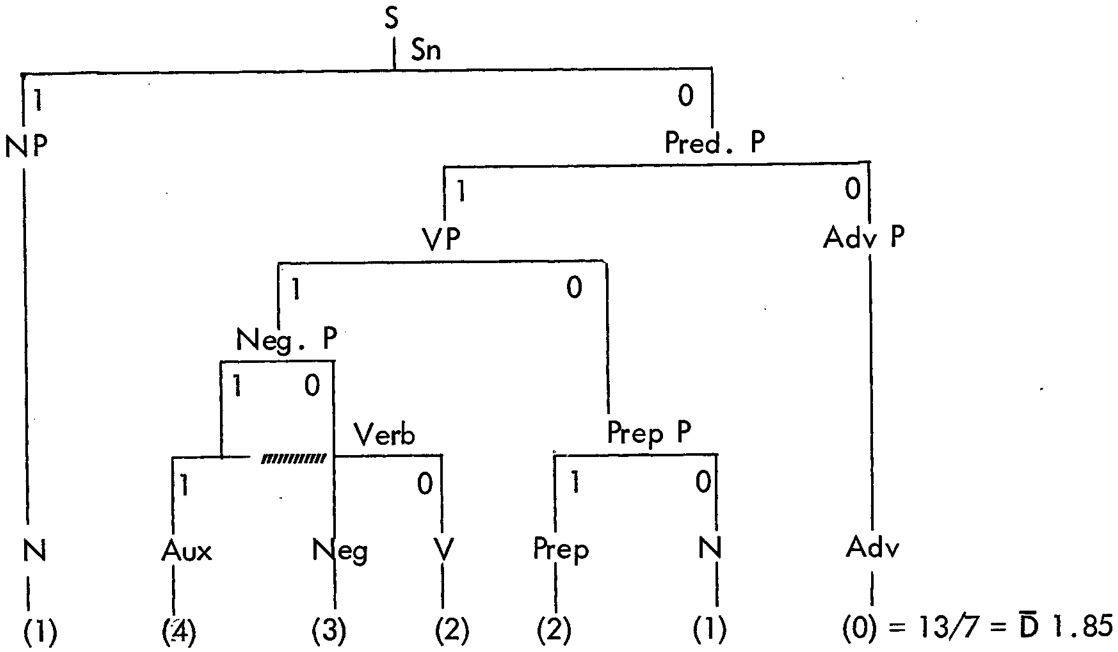
PII



Set one - Houses were soon purchased by couples enthusiastically.

Set two - Mountains were often eroded by rains gradually

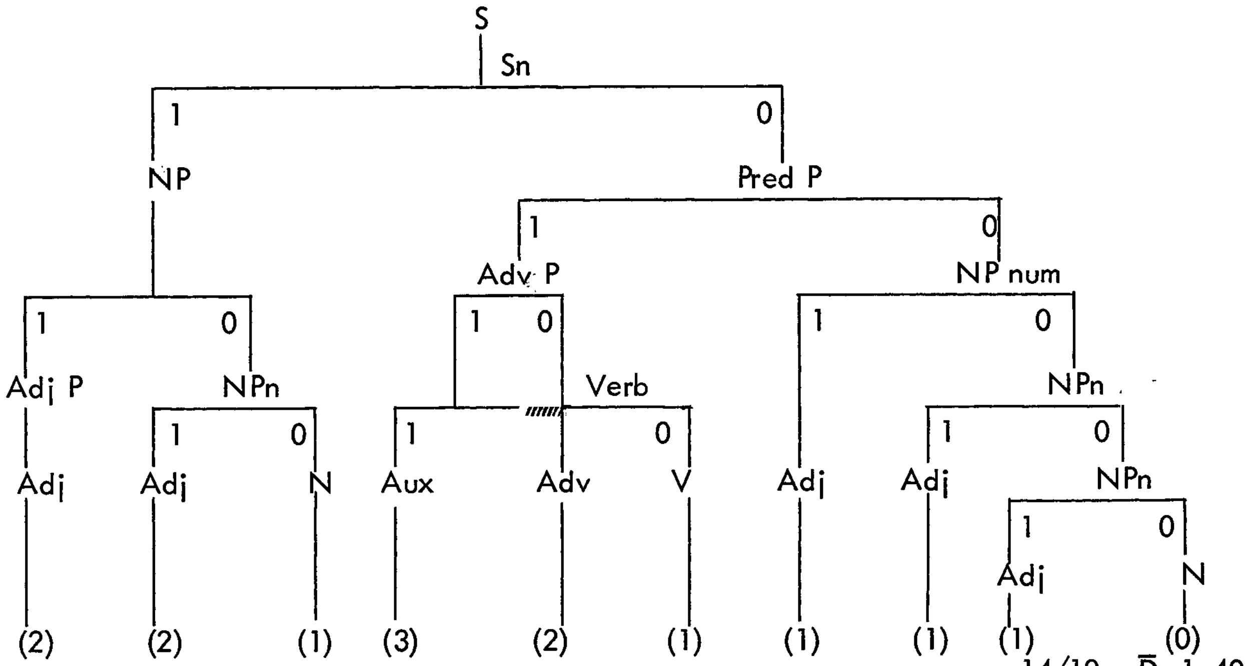
NPII



Set one - Machines were not fixed by mechanics quickly.

Set two - Cheeses were not stocked by merchants immediately.

KIII

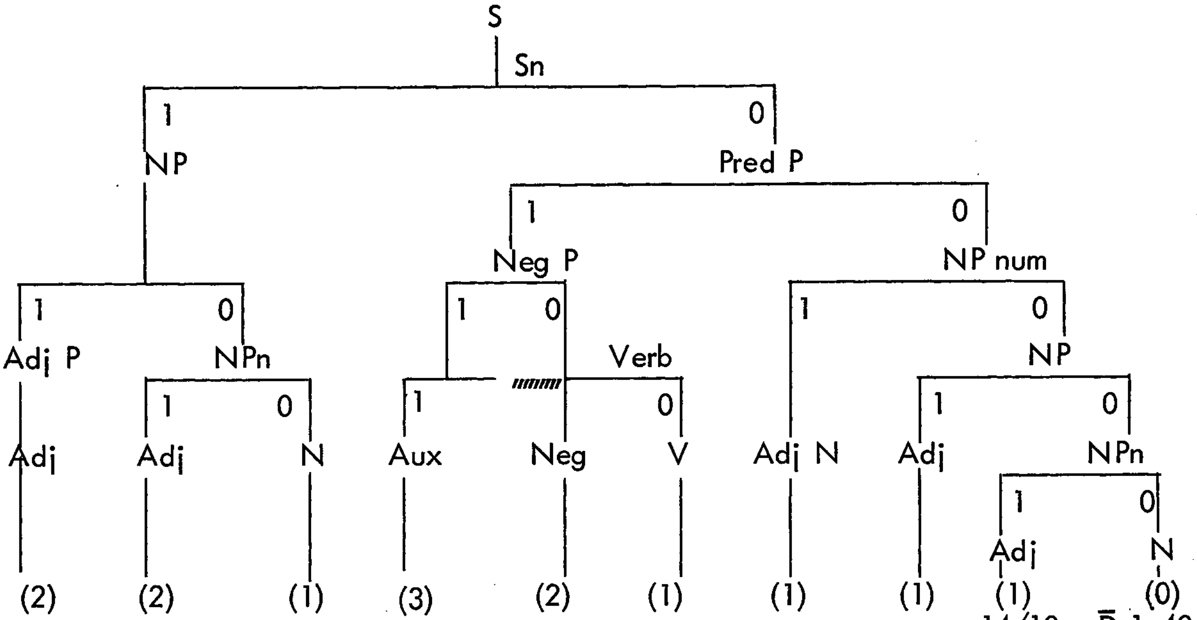


= 14/10 = $\bar{D}$ 1.40

Set one - Sturdy active children were already boarding several large yellow buses.

Set two - Patient tired workers were finally repairing most broken flooded dikes.

NIII

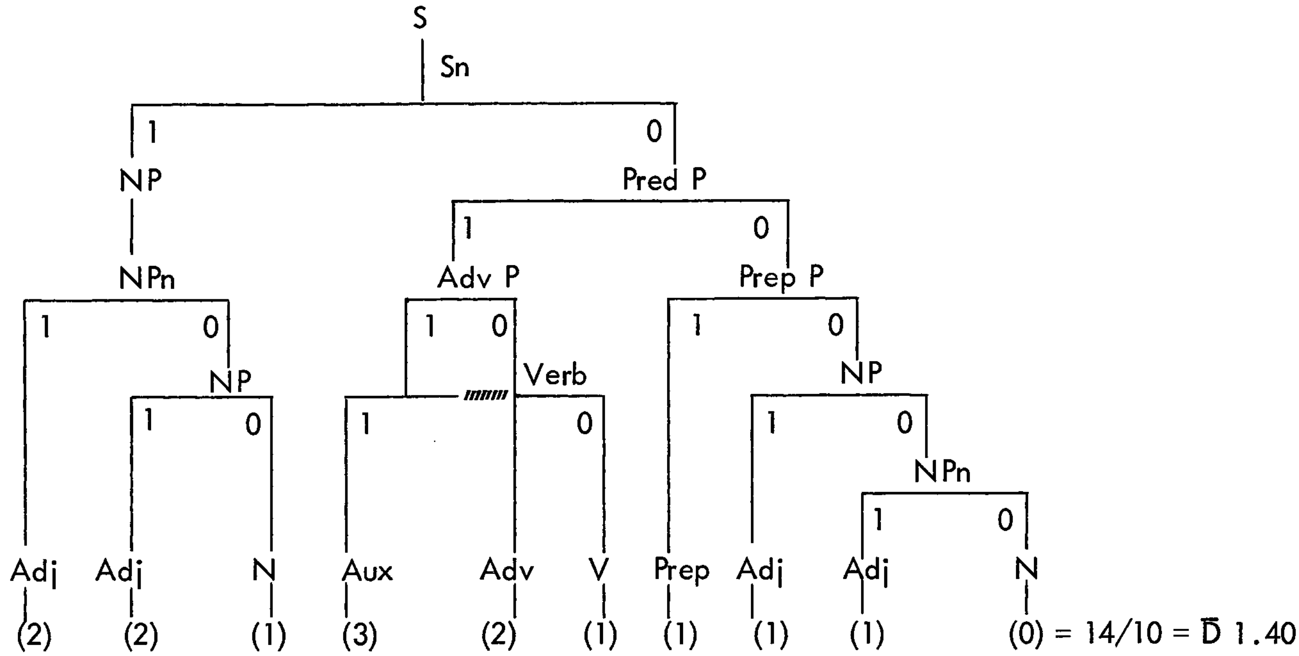


= 14/10 = $\bar{D}$ 1.40

Set one - Experienced local farmers are not planting many leafy perishable vegetables

Set two - Retired conservative residents are not joining any exclusive new clubs

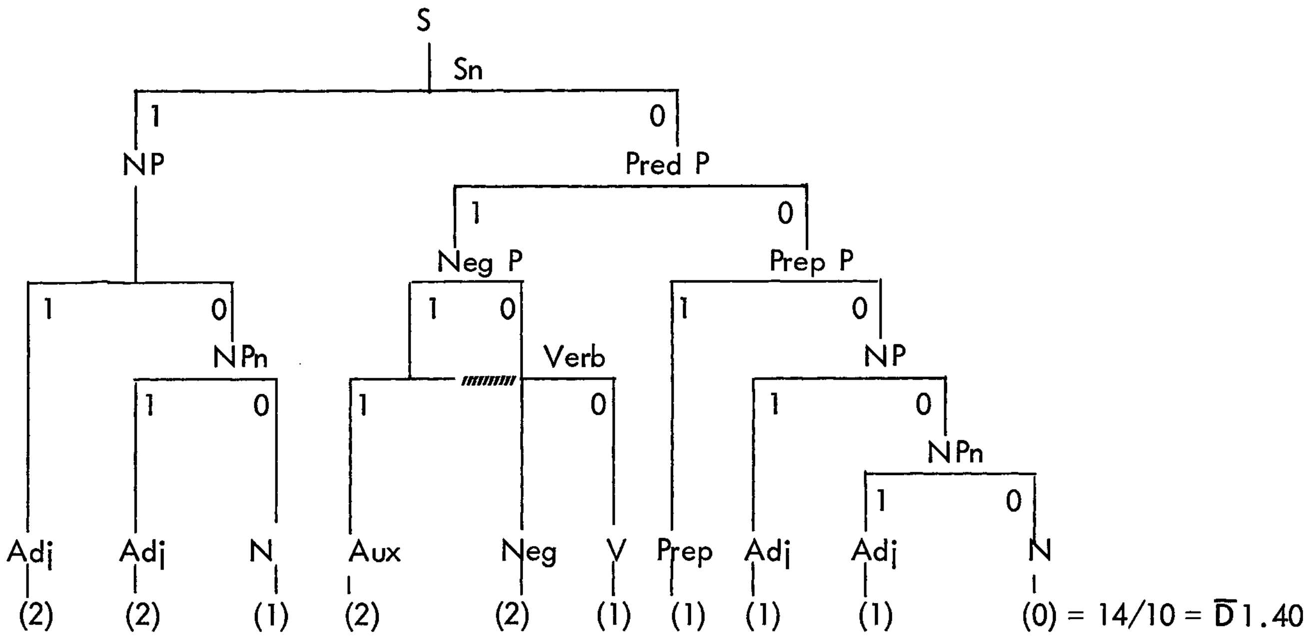
PIII



Set one - Small compact houses were soon purchased by young settled couples.

Set two - Steep rugged mountains were often eroded by pounding heavy rains.

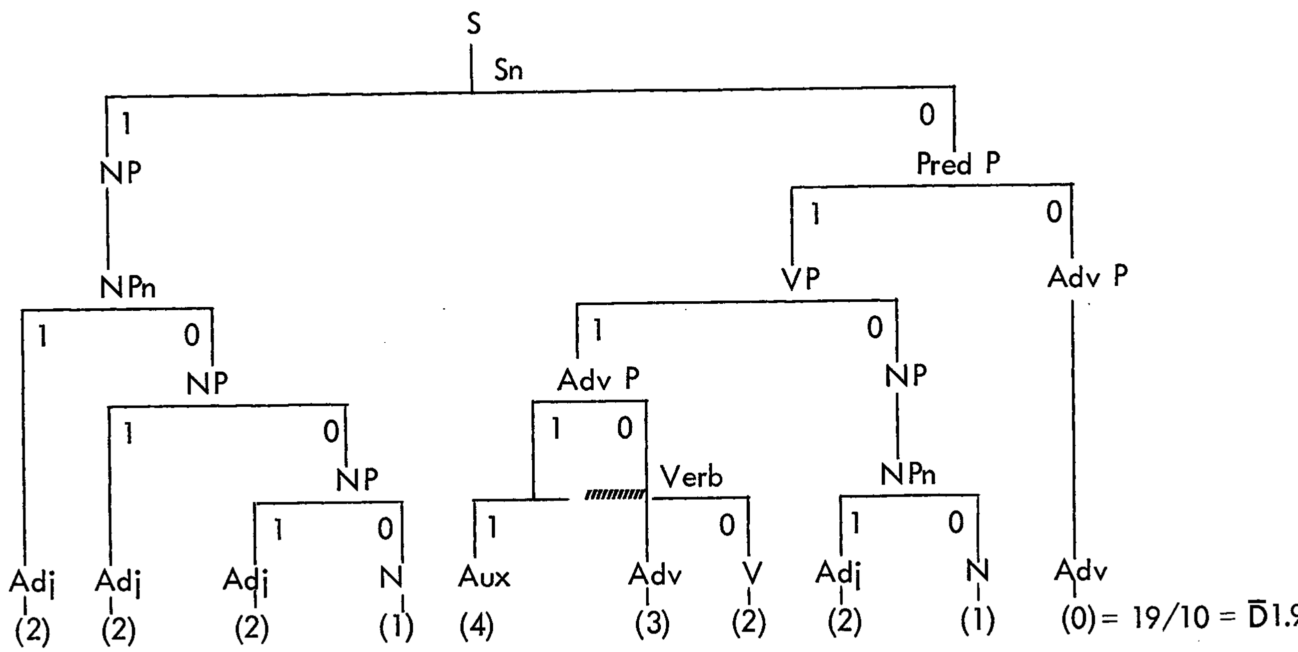
NPIII Binary



Set one - Huge bulky machines were not fixed by irritable rushed mechanics.

Set two - Imported waxy cheeses were not stocked by cautious older merchants.

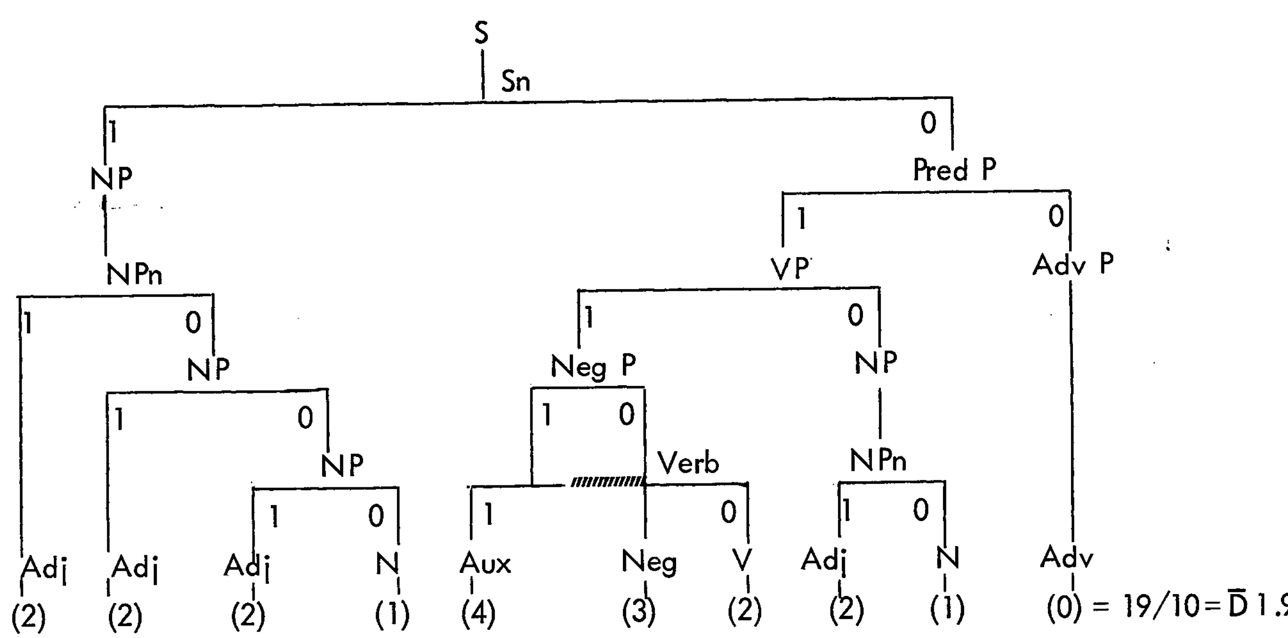
KIV



Set one - Some sturdy active children were already boarding large buses eagerly.

Set two - Many tired patient workers were finally repairing broken dikes carefully.

NIV



Set one - Several experienced local farmers are not planting leafy vegetables willingly.

Set two - Most conservative male residents are not joining costly clubs readily.

Appendix C

Written and Oral Instructions Given to Subjects

"Are you listening? You are asked to participate with many others from this school in this study. It will take about half an hour and is in no way connected with your schoolwork. We hope you will be serious in your efforts, however, so that what you do will be a real help in this study.

You are going to hear several ordinary English sentences. Listen to them all. Then, when you hear 'start,' repeat as many of the sentences as you can remember in any order, into the microphone. Speak clearly so I can understand what you say. Sometimes you will only remember a word or two, sometimes more. Say anything you can remember.

Let's review that. First you'll hear several sentences. Next you'll hear, 'Start." Finally repeat as many sentences, or parts of sentences, as you can remember, in any order. Any questions?

O.K. let's try it. Ready? Here are the sentences."

Following this the S heard the recorded sentences and "Start," after which he said what he could remember. Subsequent trials were preceded by the following remarks and instructions:

Trial Two: "Trial Two: I'm going to repeat the same sentences again. Listen to them all. When you hear 'Start,' repeat as many as you can remember, or as many fragments as you can recall. Ready? Listen to the sentences again."

Trial Three: "Trial Three: Listen to them again. Wait 'til you hear 'Start' before speaking. Ready?"

Trial Four: "Trial Four: You're remembering more and more. I'll say them again. We'll do this several more times and each time it will get easier. Ready?"

Trial Five: "Trial Five: O.K., you're doing fine. Please listen to the sentences again. Remember, don't speak 'til I say, 'Start.' Ready?"

Trial Six: "Trial Six: This is the sixth of eight trials. You're nearly finished. Keep listening closely. Ready?"

Trial Seven: "Trial Seven: O.K. I'll say the sentences again. Listen carefully. Ready?"

Trial Eight: "Trial Eight: O.K. This is the last trial. Listen carefully. Ready?"

_Ss_ in the Written Response groups heard similar tape-recorded instructions, amended to suit the response mode.

Appendix D

Scoring Criteria and Exemplars of Sentence Recall

GRAMMATICAL SUBJECT AND OBJECT NOUN PHRASE

Marking Criteria	(C) Example of Correct Sentence (I) Example of Incorrect Sentence
0. omitted	
1. correct	
2. added article or optional determiners (i.e. of the)	(C) Were active children already boarding large buses? (I) Active children were already boarding the large buses.
3. singular noun	(C) Rugged mountains were often eroded by rains. (I) Rugged mountain were often eroded by rains.
4. alternate noun or noun phrase synonym	(C) Were bulky machines not fixed by mechanics? (I) Were bulky machinery not fixed by mechanics?
5. alternate noun or noun phrase other	(C) Were cheeses not stocked by merchants immediately? (I) Were cheeses not stocked by workers immediately?
5a. omitted noun	(C) Steep rugged mountains were often eroded by constant heavy rains. (I) Rugged, steep mountains eroded by constant heavy rains.
6. alternate adjective synonym	(C) Were children already boarding large buses eagerly? (I) Children were boarding big buses eagerly.
7. alternate adjective other	(C) Farmers are not planting leafy vegetables willingly. (I) Farmers were willingly planting green vegetables.
8. extra adjective or adjective clause	(C) Small houses were soon purchased by couples. (I) Small houses were soon purchased by young couples.
9. omit adjective	(C) Were rugged mountains often eroded by rains? (I) Were mountains eroded by rains?

9a.	extra noun	(C) Were mountains often eroded by rains gradually? (I) Were mountains eroded by <u>wind</u> and rain?
10.	unmarkably incomplete	(C) Were rugged mountains often eroded by rains? (I) Rocky hills with rain.
11.	intra-sentence intrusion single word	(C) Rugged <u>mountains</u> were often eroded by rains. (I) Rugged mountains are often eroded by mountains.
12.	inter-sentence intrusion single word	(C) Are conservative residents not joining costly clubs? (C) Small houses were soon purchased by <u>couples</u>. (I) Conservative <u>couples</u> are not joining costly clubs.
15.	preposition omitted or alternate	(C) Mountains were often eroded <u>by</u> rains gradually. (I) Mountains were gradually eroding <u>from</u> rain.

PREDICATE VERB PHRASE

Marking Criteria	(C) Example of Correct Sentence (I) Example of Incorrect Sentence
0. omitted	
1. correct	
2. optional tense change	(C) Rugged mountains were often eroded by rains. (I) Rugged mountains are often eroded by rains.
3. auxiliary omitted	(C) Are residents not joining costly clubs readily? (I) Residents not joining costly clubs willingly.
3a. question transform (do and why)	(C) Are conservative residents not joining costly clubs? (I) Why were conservative residents not joining expensive clubs?
3b. adverb transform or adjective	(C) Are farmers not planting leafy vegetables willingly? (I) Farmers were not willing to plant leafy vegetables.
3c. aspect change (i.e. were not importing wished to repair)	(C) Were bulky machines not fixed by mechanics? (I) Were bulky machines not being fixed by mechanics?
3f. verb omitted	(C) Were small compact houses soon purchased by young settled couples? (I) Were large houses by unsettled couples?
4. adverb alternate synonym	(C) Children were already boarding large buses eagerly. (I) Children were already boarding large buses enthusiastically.
5. adverb alternate other	(C) Rugged mountains were often eroded by rains. (I) Rugged mountains were usually eroded by rain.
6. verb alternate synonym	(C) Small houses were soon purchased by couples. (I) Small houses were soon bought by couples.

7. verb alternate other	(C) Conservative residents are not <u>joining</u> costly clubs. (I) Conservative residents are not <u>buying</u> costly clubs.
8. adverb omitted	(C) Small houses were soon purchased by couples. (I) Were small houses purchased by couples?
9. extra adverb or adverb phrase added	(C) Are conservative residents not joining costly clubs? (I) Are conservative residents <u>already</u> joining costly clubs?
10. unmarkably incomplete	(C) Machines were not fixed by mechanics quickly. (I) Machines are not
11. intra-sentence intrusion single word	(C) Were <u>imported</u> cheeses not stocked by merchants? (I) Were imported cheeses not <u>imported</u> by merchants?
12. inter-sentence intrusion single word	(C) Patient workers were finally repairing broken dykes. (C) Were bulky machines not <u>fixed</u> by mechanics? (I) Patient workers are finally <u>fixing</u> the broken dykes.

Appendix E

Confusion Error Matrices of Eight Sentence Types Presented and Recalled by Experimental Treatments

LENGTH = SHORT
DEPTH = LOW

Type of Stimulus Sentence

RESPONSE	Block		Recalled as	K	N	P	NP	Q	NQ	PQ	NPQ
ORAL	Block = 1		Omission	20	17	17	18	21	20	27	23
			K	34	6	4	0	12	0	2	0
			N	1	30	1	5	0	13	1	1
			P	0	0	35	4	0	0	8	3
		Recalled as:	NP	0	1	0	27	0	1	0	3
			Q	3	0	0	0	22	7	1	0
			NQ	1	5	2	1	4	18	0	0
			PQ	0	0	1	0	1	0	22	5
			NPQ	0	0	0	1	0	0	1	18
			Unclassified	5	5	4	8	4	5	2	11
	Block = 2		Omission	8	6	9	4	14	12	12	8
			K	50	1	0	0	15	3	3	3
			N	3	52	0	0	0	13	2	1
			P	0	0	50	2	0	0	9	4
		Recalled as:	NP	0	1	1	54	0	0	1	9
			Q	1	0	0	0	35	3	4	0
			NQ	0	4	0	0	0	33	0	3
			PQ	0	0	2	0	0	0	32	11
			NPQ	0	0	0	3	0	0	0	25
			Unclassified	2	0	2	1	0	0	1	0
WRITTEN	Block = 1		Omission	1	0	0	0	1	0	1	0
			K	60	0	0	0	14	1	0	0
			N	0	61	0	0	0	2	0	0
			P	0	0	62	0	0	0	11	1
		Recalled as:	NP	0	0	0	63	0	0	0	5
			Q	2	0	0	0	47	1	0	0
			NQ	1	3	0	0	2	60	0	0
			PQ	0	0	2	0	0	0	47	4
			NPQ	0	0	0	1	0	0	5	54
			Unclassified	0	0	0	0	0	0	0	0
	Block = 2		Omission	7	8	14	9	12	15	14	14
			K	54	3	3	2	15	0	1	0
			N	0	48	0	1	0	7	0	2
			P	0	0	42	1	0	0	9	2
		Recalled as:	NP	0	0	0	42	0	0	0	3
			Q	1	1	1	0	30	5	2	0
			NQ	0	2	0	0	2	35	1	1
			PQ	0	0	0	2	0	0	28	9
			NPQ	0	0	0	2	0	0	7	29
			Unclassified	2	2	4	5	5	2	2	4

LENGTH = SHORT
DEPTH = HIGH

Type of Stimulus Sentence

RESPONSE = ORAL

Block		Recalled as	K	N	P	NP	Q	NQ	PQ	NPQ
Block = 1		Omission	14	22	15	19	24	25	27	24
		K	40	6	6	1	10	5	1	3
		N	0	25	0	3	1	12	0	3
		P	3	0	38	6	1	0	13	4
	Recalled as:	NP	0	0	1	24	2	0	0	4
		Q	3	5	0	1	17	8	2	3
		NQ	1	1	0	0	4	10	0	3
		PQ	1	0	1	2	0	0	16	4
		NPQ	0	0	0	3	0	0	0	8
		Unclassified	2	5	3	5	5	4	5	8
Block = 2		Omission	11	9	8	9	8	4	16	6
		K	34	6	2	0	10	3	2	2
		N	3	33	0	6	9	14	1	5
		P	1	0	44	5	0	0	11	4
	Recalled as:	NP	0	0	1	36	0	0	1	14
		Q	6	8	1	2	28	7	7	4
		NQ	4	6	1	2	8	33	6	10
		PQ	1	0	4	0	0	0	17	6
		NPQ	0	0	3	4	0	1	2	12
		Unclassified	4	2	0	0	1	2	1	1

RESPONSE = WRITTEN

Block		Recalled as	K	N	P	NP	Q	NQ	PQ	NPQ
Block = 1		Omission	11	10	8	15	10	12	12	10
		K	43	5	5	1	8	3	11	2
		N	2	40	0	5	3	10	2	9
		P	0	0	44	4	1	0	5	5
	Recalled as:	NP	0	0	1	32	0	0	0	1
		Q	2	0	0	0	33	6	4	5
		NQ	2	4	0	1	5	31	1	1
		PQ	0	0	3	0	0	0	23	3
		NPQ	0	0	1	2	1	0	6	25
		Unclassified	4	5	2	4	3	2	0	3
Block = 2		Omission	1	1	2	0	1	0	1	2
		K	53	6	2	0	12	1	2	0
		N	0	52	0	2	1	19	0	1
		P	0	0	60	9	0	0	8	1
	Recalled as:	NP	0	0	0	44	0	0	2	8
		Q	9	0	0	0	45	0	2	2
		NQ	0	5	0	0	4	44	3	1
		PQ	0	0	0	0	1	0	45	3
		NPQ	1	0	0	7	0	0	1	46
		Unclassified	0	0	0	2	0	0	0	0

LENGTH = LONG
DEPTH = LOW

Type of Stimulus Sentence

RESPONSE = ORAL

			K	N	P	NP	Q	NQ	PQ	NPQ
Block = 1		Omission	22	20	15	18	20	15	15	16
	Recalled as:	K	35	4	6	1	18	1	3	0
		N	0	28	0	3	0	16	0	1
		P	0	0	33	0	1	0	13	5
		NP	0	3	0	31	0	0	1	14
		Q	2	0	0	0	20	7	2	0
		NQ	0	3	0	1	3	20	0	0
		PQ	0	0	3	4	1	0	16	5
		NPQ	0	0	0	2	0	0	5	18
		Unclassified	5	6	7	4	1	5	9	5
Block = 2		Omission	3	4	3	6	4	7	4	5
	Recalled as:	K	58	0	0	0	15	1	1	0
		N	0	52	0	0	0	15	0	0
		P	0	0	56	4	1	0	17	1
		NP	0	0	0	45	0	0	1	16
		Q	3	1	0	1	41	5	1	0
		NQ	0	7	0	0	2	36	0	2
		PQ	0	0	4	2	1	0	36	5
		NPQ	0	0	0	6	0	0	3	34
		Unclassified	0	0	1	0	0	0	1	1

RESPONSE = WRITTEN

			K	N	P	NP	Q	NQ	PQ	NPQ
Block = 1		Omission	15	19	14	6	14	13	12	17
	Recalled as:	K	36	1	0	4	16	4	1	0
		N	0	37	0	1	0	11	0	0
		P	0	0	44	3	0	0	10	6
		NP	0	0	0	31	0	0	0	6
		Q	4	0	0	1	26	6	1	0
		NQ	0	5	0	0	1	26	1	1
		PQ	1	0	2	2	0	0	24	4
		NPQ	0	0	0	2	0	0	5	22
		Unclassified	8	2	4	14	7	4	10	8
Block = 2		Omission	1	3	3	1	0	2	2	1
	Recalled as:	K	59	0	0	0	6	0	0	0
		N	0	55	0	0	0	7	0	0
		P	0	0	60	0	0	0	2	2
		NP	0	0	0	61	0	0	0	5
		Q	4	0	0	0	53	1	0	0
		NP	0	5	0	0	0	51	0	0
		NQ	0	0	1	0	0	0	59	0
		NPQ	0	0	0	2	0	0	1	54
		Unclassified	0	1	0	0	5	3	0	2

LENGTH = LONG
DEPTH = HIGH

Type of Stimulus Sentence

RESPONSE = ORAL

			K	N	P	NP	Q	NQ	PQ	NPQ
		Omission	22	19	21	17	23	28	12	16
		K	33	1	6	0	21	3	1	2
		N	1	37	0	1	0	13	0	1
		P	1	0	28	2	0	0	22	4
Block = 1	Recalled as:	NP	0	0	1	27	0	0	0	14
		Q	0	0	0	0	13	1	2	0
		NQ	0	2	0	0	2	14	2	0
		PQ	2	0	2	1	0	0	12	6
		NPQ	0	0	0	2	0	0	1	12
		Unclassified	5	5	6	14	5	5	12	9
		Omission	12	9	6	11	8	13	2	4
		K	44	2	2	0	20	3	1	0
		N	0	46	0	0	0	16	0	0
		P	1	0	52	1	2	0	19	1
Block = 2	Recalled as:	NP	0	0	0	43	0	0	0	18
		Q	6	0	0	1	30	1	1	0
		NQ	0	5	0	0	2	28	0	1
		PQ	0	0	1	2	0	0	36	6
		NPQ	0	0	0	5	0	0	2	31
		Unclassified	1	2	3	1	2	3	3	3

RESPONSE = WRITTEN

			K	N	P	NP	Q	NQ	PQ	NPQ
		Omission	13	11	17	15	13	12	7	6
		K	37	2	2	1	13	5	2	0
		N	4	40	0	0	0	6	0	0
		P	0	0	34	1	0	0	5	3
Block = 1	Recalled as:	NP	0	0	0	31	0	0	1	9
		Q	5	0	0	0	25	2	1	0
		NQ	0	6	0	1	10	35	1	0
		PQ	0	0	6	0	0	0	29	4
		NPQ	1	0	0	3	1	0	13	33
		Unclassified	4	5	5	12	2	4	5	9
		Omission	1	1	0	2	1	1	2	1
		K	53	0	0	0	13	0	0	0
		N	0	60	0	2	0	3	0	0
		P	0	0	61	1	0	0	2	0
Block = 2	Recalled as:	NP	0	0	0	54	0	0	1	2
		Q	9	0	0	0	44	1	0	0
		NQ	1	3	0	0	6	59	0	0
		PQ	0	0	2	1	0	0	41	6
		NPQ	0	0	1	2	0	0	18	55
		Unclassified	0	0	0	2	0	0	0	0

Appendix F

Analysis of Variance Summary Table for Verbatim

Recall of Sentences

Appendix F. Analysis of Variance Summary Table for Verbatim Recall of Sentences

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	PROB. F EXCEEDED
MEAN	1407.73975	1	1407.73975	301.94116	.000
L	91.96118	1	91.96118	19.72443	.000
D	45.71753	1	45.71753	9.80579	.002
R	399.03149	1	399.03149	85.58685	.000
LD	.43945	1	.43945	.09426	.759
LR	7.99902	1	7.99902	1.71568	.193
DR	9.03101	1	9.03101	1.93703	.167
LDR	.00757	1	.00757	.00162	.968
ERROR	559.47583	120	4.66230		
B	422.31006	1	422.31006	259.38818	.000
BL	1.03328	1	1.03328	.63465	.427
BD	14.78268	1	14.78268	9.07572	.003
BR	127.99335	1	127.99335	78.61826	.000
BLD	2.39262	1	2.39262	1.46958	.228
BLR	2.00011	1	2.00011	1.22849	.270
BDR	1.32019	1	1.32019	.81088	.370
BLDR	.03125	1	.03125	.01919	.890
ERROR	195.37201	120	1.62810		
T	26.19971	7	3.74282	4.49092	.000
TL	10.38770	7	1.48396	1.78056	.088
TD	9.04443	7	1.29206	1.55031	.147
TR	16.35815	7	2.33688	2.80397	.007
TLD	12.35669	7	1.76524	2.11807	.039
TLR	5.85815	7	.83688	1.00415	.427
TDR	1.98340	7	.28334	.33998	.936
TLDR	8.22559	7	1.17508	1.40996	.198
ERROR	700.07178	840	.83342		
BT	7.20068	7	1.02867	1.91994	.064
BTL	11.54443	7	1.64920	3.07813	.003
BTD	8.54443	7	1.22063	2.27823	.027
BTR	5.45239	7	.77891	1.45379	.180
BTLD	4.99756	7	.71394	1.33252	.232
BTLR	2.70288	7	.38613	.72068	.655
BTDR	.56982	7	.08140	.15193	.994
BTLDR	1.67139	7	.23877	.44565	.873
ERROR	450.05566	840	.53578		

L = sentence length, D = sentence mean depth R = response mode B = trial block
T = sentence type

Appendix G

Appendix G-1. Analysis of Variance Summary Table for Verbatim Recall of Subject-Non Phrases

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	PROB. F EXCEEDED
MEAN	9298.65234	1	9298.65234	1471.20068	.000
L	895.07422	1	895.07422	141.61556	.000
D	22.77515	1	22.77515	3.60341	.060
R	407.91211	1	407.91211	64.53844	.000
LD	83.69312	1	83.69312	13.24164	.000
LR	.38257	1	.38257	.06053	.806
DR	9.29810	1	9.29810	1.47111	.228
LDR	3.12427	1	3.12427	.49431	.483
ERROR	758.45410	120	6.32045		
T	31.60767	7	4.51538	3.68519	.001
TL	12.56714	7	1.79531	1.46522	.176
TD	27.64673	7	3.94953	3.22337	.002
TR	7.42676	7	1.06096	.86590	.533
TLD	11.17700	7	1.59671	1.30314	.246
TLR	8.04590	7	1.14941	.93808	.476
TDR	5.53662	7	.79095	.64552	.718
TLDR	.86646	7	.12378	.10102	.998
ERROR	1029.23438	840	1.22528		
B	731.51221	1	731.51221	482.17261	.000
BL	27.65872	1	27.65872	18.23111	.000
BD	4.13243	1	4.13243	2.72387	.101
BR	37.73624	1	37.73624	24.87364	.000
BLD	1.42378	1	1.42378	.93848	.335
BLR	10.69531	1	10.69531	7.04976	.009
BDR	.86134	1	.86134	.56775	.453
BLDR	3.78130	1	3.78130	2.49242	.117
ERROR	182.05406	120	1.51712		
TB	5.27271	7	.75324	1.17743	.313
TBL	15.02026	7	2.14575	3.35413	.002
TBD	7.63989	7	1.09141	1.70604	.104
TBR	4.09888	7	.58555	.91531	.494
TBLD	11.59863	7	1.65695	2.59006	.012
TBLR	6.67114	7	.95302	1.48971	.167
TBDR	6.87964	7	.98281	1.53627	.151
TBLDR	11.55396	7	1.65056	2.58008	.012
ERROR	537.37622	840	.63973		

L = sentence length; D = sentence depth; R = response made; T = sentence type; B = trial block

Appendix G-2. Analysis of Variance Summary Table for Verbatim Recall of Predicate Verb Phrases

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	PROB. F EXCEEDED
MEAN	3808.78906	1	3808.78906	756.92247	.000
L	9.15869	1	9.15869	1.82013	.180
D	142.90088	1	142.90088	28.39899	.000
R	439.75098	1	439.75098	87.39261	.000
LD	.46899	1	.46899	.09320	.761
LR	9.16382	1	9.16382	1.82114	.180
DR	.03931	1	.03931	.00781	.930
LDR	4.59375	1	4.59375	.91293	.341
ERROR	603.82813	120	5.03190		
T	76.90698	7	10.98671	8.11820	.000
TL	14.50244	7	2.07178	1.53086	.153
TD	17.09985	7	2.44284	1.80504	.083
TR	28.01343	7	4.00192	2.95706	.005
TLD	8.97900	7	1.28271	.94781	.469
TLR	7.81445	7	1.11635	.82488	.567
TDR	12.00244	7	1.71463	1.26696	.264
TLDR	8.41650	7	1.20236	.88843	.515
ERROR	1136.80957	840	1.35334		
B	604.85352	1	604.85352	381.58447	.000
BL	17.81334	1	17.81334	11.23792	.001
BD	6.68367	1	6.68367	4.21653	.042
BR	84.90565	1	84.90565	53.56451	.000
BLD	5.18011	1	5.18011	3.26798	.073
BLR	3.36388	1	3.36388	2.12217	.148
BDR	.59816	1	.59816	.37736	.540
BLDR	.05907	1	.05907	.03726	.847
ERROR	190.21324	120	1.58511		
TB	3.01831	7	.43119	.58888	.765
TBL	16.41650	7	2.34521	3.20291	.003
TBD	7.32715	7	1.04673	1.42955	.190
TBR	3.51050	7	.50150	.68491	.685
TBLD	2.54980	7	.36426	.49748	.837
TBLR	6.61621	7	.94517	1.29084	.252
TBDR	1.13159	7	.16166	.22078	.981
TBLDR	2.17090	7	.31013	.42355	.888
ERROR	615.05908	840	.73221		

L = sentence length; D = sentence depth; R = response made; T = sentence type; B = trial block.

Appendix G-3. Analysis of Variance Summary Table for Verbatim Recall of Object-noun Phrase

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	PROB. F EXCEEDED
MEAN	8625.18750	1	8625.18750	1665.19019	.000
L	212.01221	1	212.01221	40.93135	.000
D	18.19287	1	18.19287	3.51234	.063
R	441.60364	1	441.60364	85.25755	.000
LD	52.20801	1	52.20801	10.07935	.002
LR	3.36353	1	3.36353	.64937	.422
DR	7.87451	1	7.87451	1.52026	.220
LDR	4.98120	1	4.98120	.96168	.329
ERROR	621.56421	120	5.17970		
T	242.24463	7	34.60637	22.69585	.000
TL	28.44409	7	4.06344	2.66492	.010
TD	35.51001	7	5.07286	3.32693	.002
TR	11.36157	7	1.62308	1.06446	.385
TLD	49.11060	7	7.01580	4.60116	.000
TLR	13.03687	7	1.86241	1.22142	.288
TDR	6.41650	7	.91664	.60116	.755
TLDR	18.09155	7	2.58451	1.69499	.107
ERROR	1280.82202	840	1.52479		
B	909.73462	1	909.73462	750.07495	.000
BL	9.98567	1	9.98567	8.23317	.005
BD	6.23405	1	6.23405	5.13997	.025
BR	54.14484	1	54.14484	44.64233	.000
BLD	2.19167	1	2.19167	1.80702	.181
BLR	.59824	1	.59824	.49324	.484
BDR	1.37146	1	1.37146	1.13077	.290
BLDR	5.58995	1	5.58995	4.60891	.034
ERROR	145.54303	120	1.21286		
TB	9.93604	7	1.41943	1.83334	.078
TBL	13.60425	7	1.94346	2.51018	.015
TBD	4.37012	7	.62430	.80635	.582
TBR	11.78735	7	1.68391	2.17493	.034
TBLD	4.69434	7	.67062	.86617	.533
TBLR	5.92871	7	.84696	1.09393	.365
TBDR	7.73315	7	1.10474	1.42688	.191
TBLDR	8.63940	7	1.23420	1.59409	.134
ERROR	650.35620	840	.77423		

L = sentence length; D = sentence mean depth; R = response mode; T = sentence type; B = trial block.

Appendix H

Appendix H-1. Analysis of Variance Summary Table for Correct Recall of Pivotal Word (Subject Noun)

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	PROB. F EXCEEDED
MEAN	12432.16410	1	12432.16410	2578.60107	.000
L	772.64868	1	772.64868	160.25790	.000
D	45.11548	1	45.11548	9.35757	.003
R	237.89868	1	237.89868	49.34344	.000
LD	126.01343	1	126.01343	26.13690	.000
LR	.09570	1	.09570	.01985	.888
DR	.78125	1	.78125	.16204	.688
LDR	13.13184	1	13.13184	2.72372	.101
ERROR	578.55396	120	4.82128		
T	102.04248	7	14.57750	13.51928	.000
TL	28.52856	7	4.07551	3.77966	.000
TD	18.82666	7	2.68952	2.49428	.015
TR	17.40308	7	2.48615	2.30568	.025
TLD	17.74121	7	2.53446	2.35048	.022
TLR	7.15356	7	1.02194	.94775	.469
TDR	3.96777	7	.56682	.52568	.816
TLDR	2.56934	7	.36705	.34040	.935
ERROR	905.75122	840	1.07827		
B	698.42676	1	698.42676	585.29468	.000
BL	32.00040	1	32.00040	26.81693	.000
BD	7.26703	1	7.26703	6.08991	.015
BR	14.44533	1	14.44533	12.10546	.001
BLD	.86131	1	.86131	.72180	.397
BLR	7.99994	1	7.99994	6.70410	.011
BDR	4.68942	1	4.68942	3.92982	.050
BLDR	6.34573	1	6.34573	5.31784	.023
ERROR	143.19496	120	1.19329		
TB	5.60083	7	.80012	1.19647	.302
TBL	18.96729	7	2.70961	4.05186	.000
TBD	7.48169	7	1.06881	1.59826	.132
TBR	3.05420	7	.43631	.65245	.712
TBLD	5.80957	7	.82994	1.24106	.277
TBLR	4.51514	7	.64502	.96454	.456
TBDR	5.23169	7	.74738	1.11761	.350
TBLDR	9.34131	7	1.33447	1.99552	.053
ERROR	561.73560	840	.66873		

L = sentence length; D = sentence mean depth; R = response mode; T = sentence type; B = trial block

Appendix H-2. Analysis of Variance Summary Table for Correct Recall of Pivotal Word (Predicate Verb)

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	PROB. F EXCEEDED
MEAN	13177.21880	1	13177.21880	2183.90161	.000
L	10.25684	1	10.25684	1.69990	.195
D	80.07666	1	80.07666	13.27136	.000
R	255.24487	1	255.24487	42.30252	.000
LD	.01221	1	.01221	.00202	.964
LR	19.72681	1	19.72681	3.26939	.073
DR	8.37964	1	8.37964	1.38878	.241
LDR	4.78491	1	4.78491	.79302	.375
ERROR	724.05566	120	6.03380		
T	46.16846	7	6.59549	6.04525	.000
TL	8.36963	7	1.19566	1.09591	.364
TD	17.52930	7	2.50418	2.29527	.025
TR	20.00610	7	2.85801	2.61958	.011
TLD	7.88940	7	1.12706	1.03303	.406
TLR	3.45581	7	.49369	.45250	.869
TDR	12.06885	7	1.72412	1.58028	.138
TLDR	8.22534	7	1.17505	1.07702	.376
ERROR	916.45801	840	1.09102		
B	706.61816	1	706.61816	528.50903	.000
BL	8.37981	1	8.37981	6.26761	.014
BD	1.81662	1	1.81662	1.35873	.246
BR	5.59044	1	5.59044	4.18132	.043
BLD	.53172	1	.53172	.39770	.529
BLR	.59816	1	.59816	.44739	.505
BDR	.00439	1	.00439	.00329	.954
BLDR	.30513	1	.30513	.22822	.634
ERROR	160.44041	120	1.33700		
TB	11.72534	7	1.67505	2.33502	.023
TBL	9.52222	7	1.36032	1.89628	.067
TBD	10.41284	7	1.48755	2.07364	.044
TBR	11.88916	7	1.69845	2.36764	.021
TBLD	2.04150	7	.29164	.40655	.899
TBLR	4.00635	7	.57234	.79784	.589
TBDR	4.02246	7	.57464	.80104	.587
TBLDR	1.97144	7	.28163	.39260	.907
ERROR	602.58276	840	.71736		

L = sentence length; D = sentence mean depth; R = response mode; T = sentence type; B = trial block

Appendix H-3. Analysis of Variance Summary Table for Correct Recall of Pivotal Word (Object Noun).

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	PROB. F EXCEEDED
MEAN	12461.74610	1	12461.74610	2547.58423	.000
L	168.78320	1	168.78320	34.50475	.000
D	23.21143	1	23.21143	4.74517	.031
R	291.01440	1	291.01440	59.49277	.000
LD	55.77856	1	55.77856	11.40294	.001
LR	3.78101	1	3.78101	.77296	.381
DR	.00171	1	.00171	.00035	.985
LDR	.01758	1	.01758	.00359	.952
ERROR	586.99121	120	4.89159		
T	55.04517	7	7.86360	5.99577	.000
TL	6.10864	7	.87266	.66538	.702
TD	30.11401	7	4.30200	3.28015	.002
TR	12.98315	7	1.85474	1.41418	.196
TLD	29.03589	7	4.14798	3.16272	.003
TLR	8.99097	7	1.28442	.97934	.445
TDR	2.09863	7	.29980	.22859	.979
TLDR	20.05127	7	2.86447	2.18407	.034
ERROR	1101.68091	840	1.31152		
B	929.83374	1	929.83374	979.85181	.000
BL	18.75911	1	18.75911	19.76820	.000
BD	2.12642	1	2.12642	2.24080	.137
BR	15.82059	1	15.82059	16.67162	.000
BLD	.86118	1	.86118	.90750	.343
BLR	2.53133	1	2.53133	2.66749	.105
BDR	1.87679	1	1.87679	1.97774	.162
BLDR	1.64238	1	1.64238	1.73073	.191
ERROR	113.87439	120	.94895		
TB	9.07715	7	1.29673	1.68373	.109
TBL	7.70215	7	1.10031	1.42868	.190
TBD	3.84937	7	.54991	.71402	.660
TBR	8.73340	7	1.24763	1.61997	.126
TBLD	7.80225	7	1.11461	1.44725	.183
TBLR	6.55396	7	.93628	1.21570	.291
TBDR	6.66138	7	.95163	1.23563	.280
TBLDR	5.30200	7	.75743	.98348	.442
ERROR	646.93018	840	.77015		

L = sentence length; D = sentence mean depth; R = response mode; T = sentence type; B = trial block

Appendix I

Analysis of Variance Summary Table for Conditional Error Probabilities of Phrases

Appendix I. Analysis of Variance Summary Table for Conditional Error Probabilities of Phrases

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F	PROB. F EXCEEDED
MEAN	1467.27075	1	1467.27075	1597.11890	.000
L	27.33015	1	27.33015	29.74876	.000
D	.00761	1	.00761	.00829	.928
R	36.64076	1	36.64076	39.88333	.000
LD	2.20334	1	2.20334	2.39833	.124
LR	.46495	1	.46495	.50610	.478
DR	3.02342	1	3.02342	3.29098	.072
LDR	.03250	1	.03250	.03538	.851
ERROR	110.24382	120	.91870		
B	67.65102	1	67.65102	159.27794	.000
BL	4.55705	1	4.55705	10.72915	.001
BD	.25783	1	.25783	.60703	.437
BR	2.65289	1	2.65289	6.24599	.014
BLD	.04581	1	.04581	.10785	.743
BLR	.74773	1	.74773	1.76045	.187
BDR	.28868	1	.28868	.67967	.411
BLDR	.07600	1	.07600	.17894	.673
ERROR	50.96828	120	.42474		
C	52.91757	2	26.45879	88.95976	.000
CL	13.62378	2	6.81189	22.90294	.000
CD	22.80730	2	11.40365	38.34135	.000
CR	.99590	2	.49795	1.67420	.190
CLD	3.01140	2	1.50570	5.06246	.007
CLR	1.00031	2	.50015	1.68161	.188
CDR	3.19247	2	1.59624	5.36687	.005
CLDR	.65135	2	.32568	1.09499	.336
ERROR	71.38182	240	.29742		
BC	.38489	2	.19244	.62504	.536
BCL	.14114	2	.07057	.22921	.795
BCD	1.69229	2	.84615	2.74820	.066
BCR	.38814	2	.19407	.63032	.533
BCLD	.19380	2	.09690	.31473	.730
BCLR	.77356	2	.38678	1.25623	.287
BCDR	.04519	2	.02259	.07337	.929
BCLDR	.66602	2	.33301	1.08156	.341
ERROR	73.89363	240	.30789		

Appendix J

Appendix J-1: Analysis of Variance Summary Table for Counter Instances of Simplification Transformational Errors as assessed by Off-diagonal Pairwise Comparisons

Source	Sum of Squares	Degrees of Freedom	Mean Square	*F*	*p*
Mean	28.9928	1	28.9928	1846.678	<.001
Length	0.0037	1	0.0037	<1	
Depth	0.0443	1	0.0443	2.822	
Response mode	0.0037	1	0.0037	<1	
Error	0.0628	4	0.0157		
Block	0.0003	1	0.0003	<1	
Error	0.0905	7	0.0129		

Appendix J-2: Analysis of Variance Summary Table for Proportions of 1-tag, 2-tag, and 3-tag Sentences Recalled as K-type Sentences

Source	Sum of Squares	Degrees of Freedom	Mean Square	F	p
Mean	7.4955	1	7.4955		
Length	0.0768	1	0.0768	8.2581	<.050
Depth	0.1374	1	0.1374	14.7731	<.025
Response mode	0.2602	1	0.2602	27.9744	<.010
Error	0.0374	4	0.0093		
Blocks	0.4174	1	0.4174	15.6914	<.010
Error	0.1859	7	0.0266		
Proportion (by number of tags)	2.7399	2	1.3700	82.0329	<.0001
Error	0.2337	14	0.0167		