71-12,967

OLSON, Gary Monroe, 1944-
ON THE COGNITIVE STRUCTURE OF NOUN
PHRASES: MEMORY FOR PRENOMINAL
ADJECTIVES IN ORDINARY ENGLISH SENTENCES.

Stanford University, Ph.D., 1970
Psychology, experimental

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THIS DISSERATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

ON THE COGNITIVE STRUCTURE OF NOUN PHRASES: MEMORY FOR PRENOMINAL ADJECTIVES IN ORDINARY ENGLISH SENTENCES

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF PSYCHOLOGY

AND THE COMMITTEE ON THE GRADUATE DIVISION

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

By

Gary Monroe Olson

August 1970

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Gordon Bower
(Principal Adviser)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Herbert H Clark

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

George A. Kaplan

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Roger N Shepard

Approved for the University Committee
on the Graduate Division:

Lincoln E Moses
Dean of the Graduate Division

Acknowledgements

This dissertation was completed while I held a Danforth Graduate Fellowship, and I am grateful to the Danforth Foundation for the relative freedom from financial anxiety provided by their support.

My deepest appreciation goes to my adviser, Gordon H. Bower, for his interest, encouragement, support, and advice during the carrying out of this work. We have learned much together about memory for complex verbal materials. I also thank Herbert H. Clark for his expert advice, which was often an antidote for my naivete. George A. Kaplan and Roger N. Shepard provided thoughtful advice and encouragement. Special thanks to my friend and associate, Michael R. Fehling, for his constant stimulation and his continuing excitement with mental phenomena.

Lana Boutacoff typed the final copy of the manuscript.

My wife, Joan, knows only too well the time and energy I withheld from her to complete this work. Her patience and devotion were a source of strength when the work became wearisome.

TABLE OF CONTENTS

		Page
Acknowledgements		iii
Table of Contents		iv
List of Tables		vi
List of Figures		vii
Chapter		
1	Introduction	1
	Demonstrations of Memory for Meaning in Connected Discourse	4
	What is Remembered: Historical Note	13
	What is Remembered: Linguistic Deep Structures	15
	What is Remembered: Mental Images	26
	What is Remembered: Semantic and Pragmatic Factors	33
	What is Remembered: Reprise	38
	On the Cognitive Structure of Adjectives	39
2	Theory	47
	Background	50
	The Structural Model	55
	Additional Theoretical Issues	69
	Memory Assumptions	71
3	Method	75
	Generation of Stimulus Materials	75

Page

Chapter

3 (continued)

Construction of the Memory Tests 80

Four-choice recognition test 80

Eight-choice recognition test 83

Recall test 84

Control Subjects 84

Design and Procedure of the Experiments 86

Study booklets 86

Test booklets 88

Procedure: Study session 89

Procedure: Test session 90

Design summary 91

Subjects 93

4 Results and Discussion 94

Procedural Note: Performance as a Function of Presentation Rate in Experiment 1 95

First-Order Effects: Performance on Four-Choice Recognition Tests 100

First-Order and Second-Order Effects: Performance on Eight-Choice Recognition Tests 108

Supplementary Data: Recall Evidence from Experiment 3 113

Analysis of Adjectives by Classes. 122

5 Summary and Conclusions. 129

References 136

LIST OF TABLES

Table		Page
2.1	Explication of the Subclasses of Prenominal Adjectives as Given by the Terminal Nodes in Figure 2.1	59
3.1	Examples of the Stimulus Materials	76
3.2	Examples of Recognition Test Items	81
3.3	Summary of Design and Experimental Parameters	92
4.1	Results for Four-Choice Recognition Items	97
4.2	Conditional Errors by Subjects for the Four-Choice Recognition Items	101
4.3	Polar versus Nonpolar Adjectives with Four-Choice Items	106
4.4	Polar versus Nonpolar Adjectives in Reduced Analysis of Eight-Choice Items	110
4.5	Recall Data for Polar and Nonpolar Adjectives	115
4.6	Overall Error Rate by Adjective Classes	123
4.7	Conditional Error Rate by Adjective Classes	124
4.8	Average Rank Order of Error Rate for Nonpolar Adjective Subclasses (From lowest to highest)	126

LIST OF FIGURES

Figures		Page
1.1	Correct recognition of identical or altered sentences for each test sentence type as a function of delay (after Sachs, 1967)	6
1.2	Correct forced-choice recognition of original sentence as a function of delay, type of change in the distractor, and instructions (after Wanner, 1968)	8
1.3	Relative prompt success as a function of the number of propositions per prompt word (after Wanner, 1968) .	25
1.4	Correct recognition of semantic and lexical changes as a function of the abstractness-concreteness of sentences (after Begg and Paivio, 1969)	28
2.1	Outline of the relationships among the subclasses of prenominal adjectives to be considered in the present research	58

Chapter 1

INTRODUCTION

I enjoy reading the newspaper every morning. If you were to ask me what I read in this morning's paper, I could tell you about the latest news from around the world, about yesterday's baseball results and the current status of the pennant races, about an editorial on Southeast Asia, about a series of letters to the editor from people disturbed about recent disruptions on campuses, about a column on the background of the latest developments in the Senate, and about a feature article on drug rehabilitation. I could fill you in on the market news, the local gossip, a review of a new movie, the comics, and even tell you about some of the ads. But if you were to ask me to repeat verbatim all that I had read, or even a substantial portion of it, I could not do it. You probably would not be surprised by this failure. It is perfectly ordinary that I could talk at length about what I had read but not be able to recite to you very much of it verbatim. In fact, were I to repeat the text of this morning's paper at length you would rightly suspect that there is something bizarre about me. Only a very small number of mnemonists can recall extensive amounts of material verbatim after a single reading, and for those of us with less auspicious mnemonic powers it is odd to take the incredible amount of time needed to memorize text from the San Francisco Chronicle.

But what is it I am remembering from this morning's paper? My ordinary experience with verbal materials suggests to me that what I

most typically remember of previous spoken, written, or mental verbal experiences is the "gist" of the material, very rarely the exact words. A strict verbatim recall criterion would be an absurd test of how well I have remembered what I read in the Chronicle. I cannot remember long sections of text as they originally appeared, but I can remember something, something that enables me to give you the "gist" of the text, often many details, "in my own words." My remembering something from previous experiences with complex verbal materials represents a special kind of ability, and the central problem of the study of memory for ordinary discourse is to explain what kinds of mental states and operations underlie this ability.

What is in my mind at the time of remembering that enables me to perform as I do? This is the central problem of this dissertation. It is a complicated question that cannot be answered with a small set of "definitive" experiments. In fact, it is the kind of question for which a satisfactory answer must necessarily be the result of a vast interrelated network of experimental and theoretical investigations. How shall we go about trying to answer this question? First, it will be necessary to examine some of the work of others to see what kinds of relevant facts and theories have been proposed. This will be the main function of Chapter 1. Second, it will be worthwhile to explore a smaller, manageable problem in greater detail, both theoretically and experimentally, to see if additional principles can be uncovered. This will be the task of Chapters 2, 3, and 4. Finally, taking into account both the work reviewed in Chapter 1 and the new work reported in subsequent chapters, can we formulate some proposals as an initial outline

of an answer to our question? This will be the role of Chapter 5.

In this chapter we want to examine some representative generalizations about what is remembered. First, we will review some demonstrations of the fact we obtained informally at the beginning of this chapter, that the main thing we remember from connected discourse is the meaning or gist. This will support and clarify our intuitions. The main task of this chapter is to review some suggestions for what we mean when we say we remember the gist of something. What properties does the mnemonic representation of gist have? What structural relationships can we uncover? There have been a number of proposals. A large body of work has focused on linguistic deep structures from transformational grammars (e.g., Chomsky, 1965) as models of gist memory, and we will examine some of these proposals to find out what they tell us that is relevant to our question. Then there is a growing body of work examining the role of mental imagery in memory for meaningful material. Although no one has a well-formulated language for talking about mental imagery, research has clearly shown that it is a cognitive phenomenon of great importance in memory, that it will have to be an important aspect of any model of memory for meaning. Finally, much research on memory for sentences and prose has explored the role of factors like permanent memory, the communicative setting, and affective responses. Although these factors have usually not been related to specific models for what is remembered, they are important for a general answer to our question.

The work to be presented in Chapters 2, 3, and 4 focuses on memory for prenominal adjectives in ordinary English sentences. The

reasons for looking at this are given at the beginning of Chapter 2. In order to give the work reported here some specific context, the present chapter will conclude with a brief review of empirical work on adjectives.

Demonstrations of Memory for Meaning in Connected Discourse

The intuitions about memory for meaningful verbal material highlighted in the discussion of the newspaper example have been supported by laboratory research. In an important study Sachs (1967) found that subjects were very good at recognizing changes which altered the meaning of prose but were very poor at recognizing changes which left the meaning intact but altered structural or lexical characteristics. Subjects listened to recorded prose passages, and at some point the passage was interrupted and a test sentence presented. These test sentences were either identical to a sentence presented earlier in the passage, were altered so as to change the form of an earlier sentence but leave the meaning unchanged, or were altered so as to change the meaning. If [1.1] represents a sentence originally heard in the text, [1.2] is a sentence with a changed meaning while [1.3] and [1.4] are sentences with altered structure but the same meaning (the labels in brackets refer to the respective curves in Figure 1.1). Subjects were

[1.1]	The tall boy hit the girl.	[IDENTICAL]
[1.2]	The girl hit the tall boy.	[SEMANTIC]
[1.3]	The girl was hit by the tall boy.	[ACTIVE/PASSIVE]
[1.4]	The boy who is tall hit the girl.	[FORMAL]

very good at correctly identifying the original sentence or any of the changed sentences on an immediate test, but when the test sentence was given after some intervening prose had been listened to subjects could only recognize changes of meaning with any accuracy. The curves for the four types of test sentence are shown in Figure 1.1.

Wanner (1968) used a clever technique to insure that his subjects would not be processing the target material in any special way. The subjects received tape-recorded instructions for a task very similar to Sachs'. But at the end of the instructions, instead of beginning the task on which they had just been instructed, Wanner tested subjects on their memory for the instructions themselves. Subjects were given a pair of sentences, one of which had been in the original instructions and another which differed from it only in the order of two of the words. For half the subjects the altered word order of the distractor did not affect the meaning of the sentence, while for the other half of the subjects the altered word order changed the meaning. When the change did affect meaning subjects made no errors in picking out the original sentence. But then the change did not affect meaning subjects were very poor at picking out the original, and in fact at a delay of 51 intervening syllables performance was at chance. These results show conclusively that when subjects do not expect a specially contrived memory test, but are only trying to understand the material, they remember the gist without remembering semantically irrelevant structural features.

Related results have been reported by Begg and Paivio (1969),

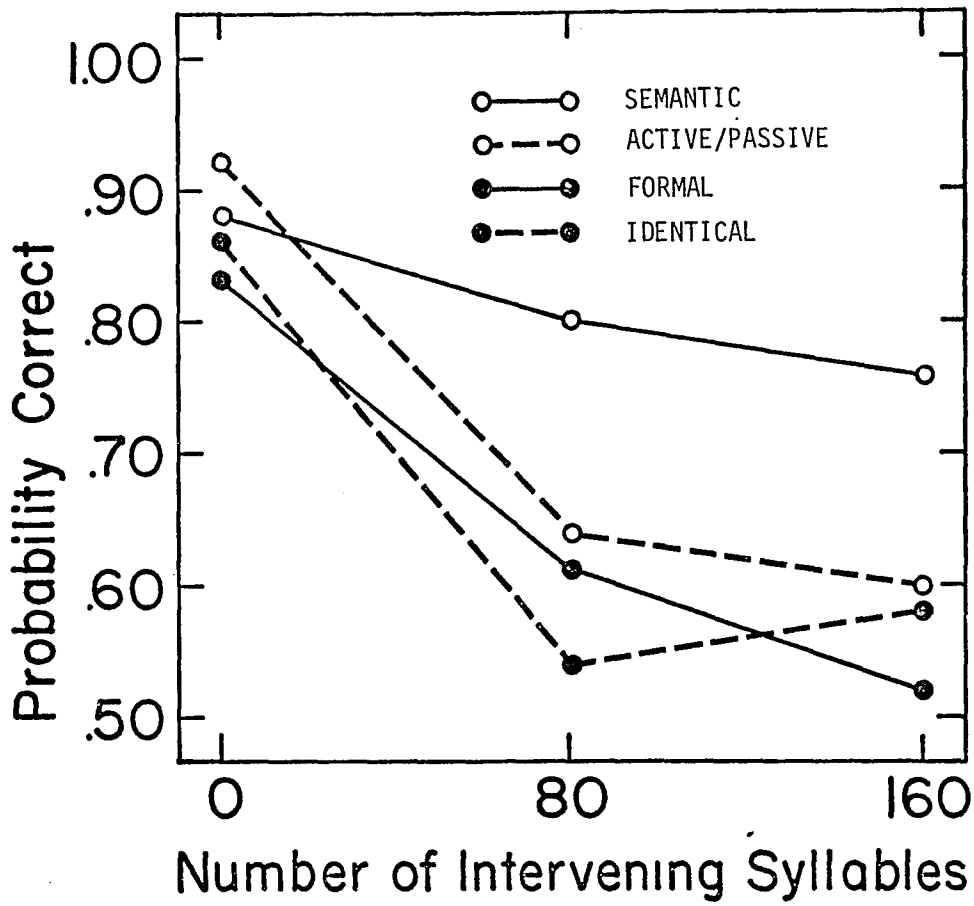

Figure 1.1. Correct recognition of identical or altered sentences for each test sentence type as a function of delay (after Sachs, 1967).

Bregman (1968), and Pompi and Lachman (1967). The contrast between memory for gist and memory for structural features is striking. The fact that in Sachs (1967) and Wanner (1968) memory for structural features was virtually gone after moderate retention intervals (a few minutes at the most) suggests that such information is not encoded in any permanent form. This generalization will recur throughout this chapter: the basic stored representation of sentences is semantic, while syntactic or structural information is essentially uncoded except for special circumstances.

Let us look at some of those special circumstances, because the phenomena observed when structural information is intentionally encoded are important. There was a second set of subjects in Wanner's (1968) experiment. These subjects participated in the same design described above except they were warned that they would be tested on their ability to remember the instructions. Not unexpectedly these subjects were better at noticing the semantically irrelevant changes. That is, they were able to retain significantly more information about structural or surface features of the material than the unwarned subjects, although they still remembered the semantic content better (i.e., they were better able to choose the original sentence when the distractor was altered in meaning). However, at longer delays they made many ereros in recognizing changes of word order that affected meaning, while the unwarned subjects made no errors at equivalent delays. See Figure 1.2, where Wanner's data have been reproduced. Intentionally trying to remember the materials verbatim impaired their ability to remember

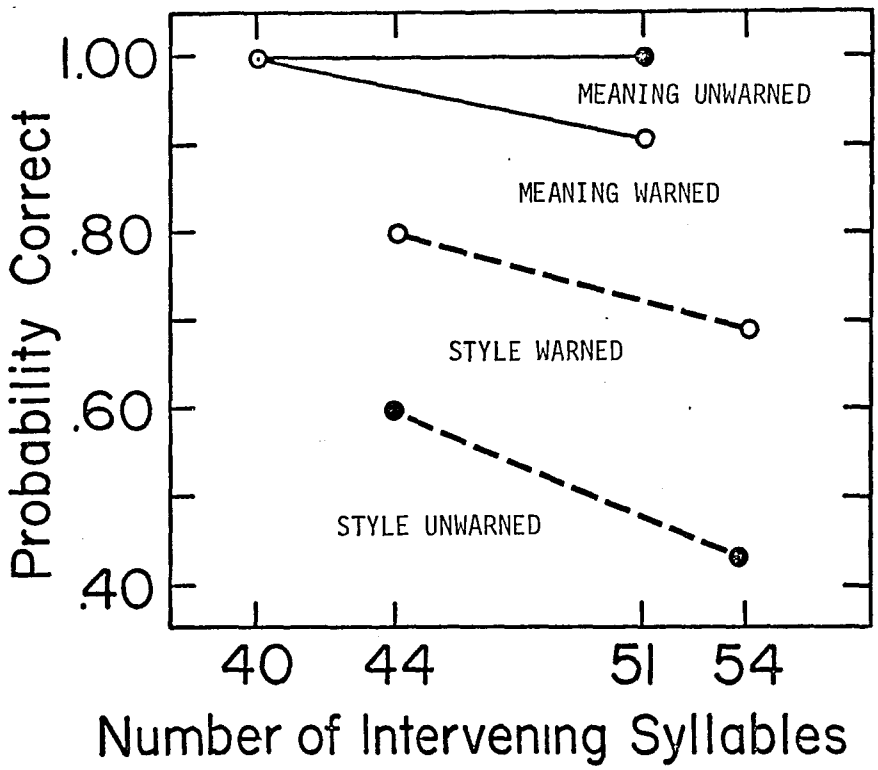

Figure 1.2. Correct forced-choice recognition of original sentence as a function of delay, type of change in the distractor, and instructions (after Wanner, 1968).

the semantic content of the instructions, as though trying to remember surface aspects of the material conflicted with trying to remember the meaning. This would seem to confirm the introspective conclusion we reached about the newspaper material: rote memorization is something quite different from ordinary comprehension, and in no sense do the two abilities seem to go together very well.

Related evidence was collected by Bregman and Strasberg (1968). They investigated the ability of subjects to remember the syntactic form of sentences. The important evidence for us are the extensive subject reports they collected after the experiment. These phenomenal reports revealed that memory for syntactic form is highly indirect, that subjects tended to reconstruct the original syntactic form of sentences by amalgamating a wide variety of information in memory, "including memories of word order, word endings, subject salience, affective response, truth value, interpretation and action imagery [p. 401]." These informal data suggest that memory for aspects of complex verbal material other than the meaning is secondary, that even when it does exist it is essentially reconstructed from other information in memory.

The development of transformational-generative grammars created a great deal of excitement among psychologists interested in complex verbal behavior, and a number of investigators attempted to explore the psychological "reality" of the early versions of these grammars (i.e., that of Chomsky, 1957). In particular, a memory model emerged as investigators tried to show that what was remembered from sentences was a kernel proposition capturing the basic meaning of the sentence

plus transformational markers that indicated what kinds of formal operations were required to get the kernel back to the original form. For instance, a passive sentence like [1.5] would be stored as a kernel proposition in [1.6] plus a transformational tag that indicated

[1.5] The boy was bitten by the dog.

[1.6] dog + bite + Past + boy

the passive form. Simple positive active declarative sentences were seen as the most primitive syntactic form, since their grammatical description entailed only obligatory transformations from the kernel to the surface form, and in terms of the memory model could be stored as an unmarked form. Examples of optional transformational markers that needed to be stored were passive, negative, question, wh- question, and emphatic, following the syntactic analysis of Chomsky (1957). It is obvious that in their model transformations like the negative change the meaning of the sentence (the issue as to whether or not transformations should change meaning has been actively disputed in the linguistic literature; see, e.g., Katz & Postal, 1964). Given that subjects typically remember the meaning of sentences quite well, this would seem to be an implausible memory model since forgetting would yield sentences whose meaning contradicted the original. But the formal simplicity of the model appealed to psychologists.

An experiment by Mehler (1963) attempted to demonstrate the correctness of this simple memory model. He wanted to show that the kinds of errors made in recalling sentences of various forms (active, passive, negative, question) provided evidence for the forgetting of transfor-

mational markers. Mehler's conclusion was that the pattern of errors indicated subjects remembered kernel propositions very well, but as transformational markers were lost with time the exact form of the original sentence became difficult to recall. However, a reanalysis of Mehler's data by Foa and Schlesinger (1965) indicated that, contrary to the early notions about kernel propositions, meaning, and transformational markers, almost all errors made by subjects in Mehler's experiment were meaning-preserving (e.g., active-passive confusions but rarely active-negative ones). That is, the errors were very systematic, and provided support for the general principle that some kind of abstract semantic representation is what is stored in memory at the time of comprehension, with little if any information about the transformational history of the sentence retained directly.

Another experiment analyzing recognition confusions yielded a similar conclusion. Fillenbaum (1966) presented subjects with a series of simple sentences and later gave them a specially-designed multiple-choice test. For instance, if subjects had studied the sentence [1.7], then [1.8] would be the test item. The basic finding was that errors

[1.7] The door is open.

[1.8] The door is (a) open (b) not open (c) closed (d) not closed.

which preserved meaning were more frequent than errors which changed the meaning of the original sentence but which were similar phonetically. In the example above, the most frequent error would be (d), not closed. The difference between meaning-preserving and meaning-violating errors was greater for contradictory pairs of lexical items,

where the negation of one member of the pair implies the other (e.g., open-closed, alive-dead), than for contraries, where the negation of one member does not necessarily entail the other (e.g., hot-cold, big-little). The finding of Fillenbaum's experiment, of course, is that subjects do not remember sentences verbatim or even phonetically but rather remember something which underlies the meaning that a pair like open-not closed has in common.

There have been other studies which have supported the general conclusion of Foa and Schlesinger's (1965) reanalysis of Mehler's (1963) data (Anderson, 1963; Coleman, 1965; Martin & Roberts, 1966; Clark & Clark, 1968; Bregman & Strasberg, 1968). Initial attempts to use the formal insights of transformational grammars were unsophisticated, but the research generated by the model did support the abstract semantic representation principle. Most of the research about to be reviewed also supports this general conclusion. We will take the general principle to be well enough established that we will not continue to belabor this point in reviewing further research. Rather, we shall turn to a consideration of what the semantically-based representation might look like, what its structure might be, since the research reviewed so far has given no indication of this.

Before going on, however, we should keep one thing in mind. We have emphasized so far that it is primarily semantic information that is remembered from connected discourse. The repeated use of the word "abstract," as in "abstract semantic representation," has been intentional and deserves elaboration. Semantic information is remembered better than structural information, in fact remembered much better as

we have seen. But the tasks we have looked at did not make very stringent demands on the meaning memory. Quite clearly we do not remember all the semantic information present in ordinary discourse. More likely we remember very little of it. Rather, we remember important themes, dominant actions, more salient characteristics of objects and events, without retaining all of the details. By "gist" we mean the essence of the material, the general idea, the central points. The concept of abstraction, for all its vagueness, is an important part of what we mean by memory for gist, and cannot be ignored.

What is Remembered: Historical Note

Remarkably little attention has been given to the problem of what is remembered from complex verbal material until recently. Welborn and English (1937), in an extensive review of research on the learning of meaningful verbal material, cited few studies that considered this particular question. The most notable exception to this was the work of Bartlett (1932). In a series of interesting and important experiments Bartlett attempted to uncover factors in the recall of both complex verbal and pictorial materials. He studied the systematic distortions in reproduction over time, and concluded that socially or culturally determined norms or schemata were of primary importance in the remembering of complex materials. His claim was that whatever is remembered is strongly influenced by these schemata, and with time the material tends to incorporate more and more of these cultural norms and lose its distinctive ties with the original material. These pro-

cesses of simplification and assimilation to verbal and cultural norms suggest that the representation given to learned material depends heavily upon relatively permanent semantic and pragmatic (in the sense of Morris, 1938, having to do with extralinguistic social and cultural factors involved in language as communication) structures in memory. The advantage of using such structures is that the encoding of new material is facilitated. That is, by using well-learned and easily accessible structures in memory one can organize and store new material more easily, (see Bower, 1970, for a more recent discussion of such encoding advantages). The great interest of Bartlett's work is that it revealed the dynamic properties of a system which must rely upon such long-term structures, showing the systematic distortion of the original material that can take place as the original material and the well-learned long-term material become merged and interwoven.

The research tradition initiated by Bartlett has continued to be active, and has produced a large number of studies (representative ones are Oldfield, 1954; Gomulicki, 1956; Paul, 1959; King, 1968; Johnson, 1970). In general the work has concentrated on developing the notion of a "schemata." For example, there have been attempts to give the notion of "schemata" a more rigorous flavor by relating it to notions of a code as developed in information theory (Oldfield, 1954). Gomulicki (1956) has examined the role of "abstraction" with verbal materials of varying complexity. Despite continuing research in this area, little insight has been gained into the specific details of the kind of schematic representation referred to by these theorists. The language of

"schemata," despite the attempts of people like Oldfield (1954), has remained vague and suggestive, and there have been no ties between this work and the more linguistically motivated work or the recent work on mental imagery. Nevertheless, the evidence gathered by Bartlett and those that have followed him is among the richest in complex verbal memory, and certainly needs more systematic and analytic attention than it has been given.

What is Remembered: Linguistic Deep Structures

We have already mentioned the effect of an earlier form of transformational grammar (Chomsky, 1957) on memory research, and the generally negative findings with respect to the transformational marker theory of memory. Recent advances in linguistic theory, including revisions and refinements of transformational grammars (e.g., Katz and Postal, 1964; Chomsky, 1965) have provided a source of formalizations for a different kind of theory of verbal memory. The details of the deep structure itself, rather than the transformational relationships between deep and surface structure, have been the basis of new memory models. The research to be examined in this section has attempted to isolate specific characteristics or parameters of linguistically determined deep structures which are important predictors of memory performance. To the extent these predictions work, there is evidence of at least something isomorphic to linguistic deep structures in memory.

There is a general theoretical issue which needs to be raised with respect to the strong form of the hypothesis that deep structure re-

lationships parallel memory relationships. This strong form says that linguistic deep structures are in fact the exact representation given to sentences in long-term memory. Although this has been a seductive hypothesis for formalism-starved psychologists, it is clearly false. Deep structures represent grammatical abstractions, formal idealizations of a narrowly defined set of relationships that are important for well-formedness distinctions among strings in natural languages. Too many other factors not included in linguistic deep structures have been shown to be important in memory. What we can hope for, however, is that linguistic deep structures will be able to capture some formal relationships that are important to mnemonic representations. What the research about to be reviewed may tell us is that interrelations among elements of a sentence as specified by linguistic deep structures are important in the mnemonic representation as well, not that linguistic deep structures are the actual representation.

One of the most obvious formal parameters of deep structure for memory is some kind of complexity measure on the phrase structure trees specified by linguistic theory. The general hypothesis is that sentences with greater complexity will be more difficult to remember or will make extra demands on the memory system. A study by Savin and Perchonock (1965) investigated a particular model of sentence complexity. They defined complexity largely in terms of transformational history as specified by the earlier version of Chomsky's (1957) transformational grammar. They had subjects learn a sentence and a list of eight words, and then asked for immediate recall of both the sentence and the words. They reasoned that sentences with greater complexity

should take more space in memory, and thus subjects should recall fewer of the accompanying words. This result was obtained. For instance, subjects recalled more words when they correctly recalled an active positive declarative sentence than when they correctly recalled a negative. There are several problems with these results. First, since their analysis was based on a memory model (transformational marker hypothesis) that has been rejected for other reasons (see our review of Mehler's research in an earlier section), the basis for their predictions is weakened. They could of course reformulate their model so as to give essentially equivalent predictions. Second, a number of more recent investigations have had difficulty replicating their results, or have replicated them while at the same time uncovering serious confoundings in the original experiment that leave the complexity hypothesis untested (e.g., Matthews, 1968; Epstein, 1969; Glucksberg and Danks, 1969). The Savin and Perchonock study was an ingenious if unsuccessful attempt to propose and test a particular complexity measure.

Martin and Roberts (1966) have examined another complexity measure, the Yngve (1960) number computed for the surface phrase marker of relatively simple sentences. They found that sentences with a larger Yngve number, hence with greater structural complexity, were more difficult to recall than sentences of equal length with a smaller Yngve number. Furthermore, Martin and Roberts reanalyzed Mehler's (1963) data, and found that Yngve complexity (which is of course confounded with transformational markers as specified by the memory hypothesis

tested by Mehler) predicted the probability of recall quite well. Although Martin and Roberts found a main effect for Yngve complexity in their experiment, they also found a main effect for sentence type with complexity held constant, and found interactions between sentence type and complexity. Thus at best Yngve complexity is a partial predictor of memory performance, and is a puzzling measure for them to have chosen since it is not tied to the kinds of semantic relationships which have been found to be so important in sentence memory.

A more sophisticated use of the complexity of deep structure for predicting memory performance was the work of Rohrman (1968, 1969; Rohrman & Polzella, 1968; Polzella & Rohrman, 1969). His basic finding has been an apparent inverse relationship between the ease of recalling English nominalizations and the complexity of their deep structure computed by counting the number of nodes in the appropriate phrase-structure tree. Most of the work has centered upon the differences between subject and object nominalizations. The noun phrase growling lions is an example of a subject nominalization, and its deep structure is shown schematically by the labeled bracketing in [1.9]. The

[1.9] $[\ [\text{lions}]_{NP}\ [\text{growl}]_{VP}\ ]_{S}$

phrase digging holes is an object nominalization, and its deep structure is shown in [1.10]. The PRO ("someone") is a dummy element for

[1.10] $[\ [\text{PRO}]_{NP}\ [\ [\text{dig}]_{V}\ [\text{holes}]_{NP}\ ]_{VP}\ ]_{S}$

the implied subject of the verb dig. The presence of this PRO form adds an extra node to the deep structure of the object nominalization, and under the hypothesis that greater complexity leads to poorer recall,

object nominalizations should be more difficult to recall than subject nominalizations. Rohrman has discovered this in a number of experiments. Polzella and Rohrman (1969) extended this reasoning to the recall of transitive and intransitive verbs. They found that on the average transitive verbs (like dig) were more difficult to recall than intransitive verbs (like growl). As with the difference in recallability between the two kinds of nominalizations, they reasoned that transitives were more difficult to recall because there is more implicit deep structure associated with them than with intransitives. Transitives normally have both a subject and an object in sentential contexts while intransitives have only a subject. If one assumes that the lexical entry for a verb contains information about the syntactic frames within which the verb can appear, then transitive verbs will have more information, i.e., an additional implicit node, than intransitives, and thus will be more difficult to remember. Or, equivalently, if subjects attached an object noun to each transitive verb they study, they would increase the average memory load for these verbs. At a first approximation, then, the data for both the nominalizations and the transitive-intransitive verbs indicated that ease of recall could be related to a relatively straightforward complexity measure on the deep structure of the materials. However, these data were not conclusive, and in fact Paivio (1970) has come up with an interpretation based on mental imagery which is quite different than Rohrman's. Because of the importance of Paivio's study we defer consideration of his data until a later section on mental imagery, but his data cast serious doubt on Rohrman's complexity interpretation.

Weisberg (1969) has explored a different deep structure parameter, the relative proximity of lexical elements in the underlying propositional structure. He found that changes in word associations due to experience with sentences were accounted for on the basis of deep structure proximity, not surface structure proximity. He demonstrated this in two ways. First, sentence pairs with similar surface structure but different deep structures were presented to different groups of subjects. The sentences [1.11] and [1.12] were such a pair in Weisberg's

[1.11] Places with new planes flying can be dangerous.

[1.12] Even with new planes, flying can be dangerous.

experiment. The two sentences can be broken down into the sets of elementary propositions shown in [1.13] and [1.14]. The basic prediction

[1.13] Places are dangerous. Planes are new. Planes are flying.

[1.14] Planes are new. Flying is dangerous.

was that word associations elicited after subjects had studied these sentences would reflect proximity measures within these deep structure elementary propositions, not surface structure proximity. With these examples, if flying were given as the stimulus, one would expect planes to be the most frequent intrasentence associate in [1.11] while dangerous would be the most frequent in [1.12]. This was what Weisberg found. Most of the sentences in his experiment differed slightly in surface structure as well, and the results were not conclusive in eliminating surface structure as a variable. So Weisberg did a second experiment in which a single deep structure was given four different surface realizations. If the elements in [1.15] represent the deep

structure relations to be studied, then [1.16a] through [1.16d] are all different surface realizations of these same deep structure propositions.

[1.15] Children eat bread. Children are slow. Bread is cold.

[1.16a] Slow children eat cold bread.

[1.16b] Slow children eat, as you know, cold bread.

[1.16c] Cold bread is eaten by slow children.

[1.16d] Children who are slow eat bread that is cold.

The prediction that intrasentence word associations to stimuli taken from these sentences should be constant regardless of the particular surface structure a subject studied was confirmed. This finding depended upon supporting the null hypothesis, but was consistent with the interpretation of the first experiment. To the extent that the organizational processes involved in understanding a sentence yield an internal representation that influences later word association behavior, it is deep structure relationships rather than surface structure ones that determine the nature of this influence.

Davidson (1968) reported a related finding. He had subjects learn simple sentences over five learning-test trials, and analyzed the protocols for transition error probabilities (TEPs) between successive elements of the recalled sentences (subjects were encouraged to guess when they could not recall exactly, so most of the protocols were complete sentences). The deep structure of the sentences ordered the observed TEPs better than the surface structure, indicating the subjects' behavior during the recall portion of each trial was influenced more by the relationships among constituents of the deep structure than by the surface structure. Both the Weisberg (1969) and the Davidson (1968)

studies supported the principle that deep structure proximity is more important than surface structure proximity in predicting the behavior of sentence constituents. This supports the general metatheoretic point we developed earlier, but provides little insight into the details of mnemonic representation.

Blumenthal (1967; Blumenthal & Boakes, 1967) proposed that the scope or range of a word's function in deep structure should predict the effectiveness of that word as a prompt for recall of the entire sentence. Words with greater deep structure scope, that is words inserted into the phrase marker closer to the initial S symbol, should be more effective prompts. For instance, the pair of sentences in [1.17a] and [1.18a] have the same surface structure but differ in deep structure (shown in [1.17b] and [1.18b]).

[1.17a] Gloves were made by tailors.

[1.17b] $[[\text{tailors}]_{NP}[[\text{made}]_{V}[\text{gloves}]_{NP}]_{VP}]_{S}$

[1.18a] Gloves were made by hand.

[1.18b] $[[\text{someone}]_{NP}[[\text{made}]_{V}[\text{gloves}]_{NP}[\text{by hand}]_{MANNER}]_{VP}]_{S}$

Blumenthal reasoned that tailors should be a more effective recall cue than hand since by hand is only a part of the verb phrase whereas by tailors is a passified logical subject. As the sentence's logical subject, tailors has greater scope or range of function in [1.17a] than hand has in [1.18a]. Similarly, for the pair of sentences in [1.19a] and [1.20a], eager is a nominal modifier while easy is a sentence modifier (see [1.19b] and [1.20b]), so that easy should be a

[1.19a] John is eager to please.

[1.19b]

$$[[\text{John}]_{NP}[[\text{be eager}]_{V}[[\text{John}]_{NP}[[\text{please}]_{V}[\text{someone}]_{NP}]_{VP}]_{S}]_{VP}]_{S}$$

[1.20a] John is easy to please.

[1.20b]

$$[[[\text{it}][[\text{someone}]_{NP}[[\text{please}]_{V}[\text{John}]_{NP}]_{VP}]_{S}]_{NP}[\text{be easy}]_{VP}]_{S}$$

more effective prompt for recall of the entire sentence than should eager. These predictions were supported by Blumenthal's data.

A different deep structure parameter was offered by Wanner (1968) as the basis for predicting the relative effectiveness of words as prompts for sentence recall. He reasoned that it is the number of times a lexical item is entered in deep structure, not its scope or range, that determines its effectiveness as a prompt. The examples discussed earlier in [1.13], [1.14], and [1.15] show how different lexical items can appear in more than one deep structure proposition. Wanner used sentences like those in [1.21] and [1.22], pitting his proposal against Blumenthal's. Governor appears equally often in the

[1.21] The governor asked the detective to cease drinking.

[1.22] The governor asked the detective to prevent drinking.

deep structure of these sentences (once), while detective appears twice in the deep structure of [1.22] and three times in the deep structure of [1.21]. Neither detective nor governor differ in their scope in the two sentences. Subjects first free-recalled sentences like these, and then were given the equivalents of detective or governor as prompts. Wanner calculated the relative improvement in sentence recall with the

prompts compared to the free recall case, and plotted this as a function of the number of occurrences of the prompt in the deep structure of the sentence. The results are shown in Figure 1.3, and support Wanner's hypothesis. Whatever is stored in memory at the time of studying sentences like these seems to preserve the relationship among the lexical constituents of the sentence that is captured by counting the number of elementary propositions in which a particular word is implicated in the linguistic deep structure. Since Wanner did not independently vary the scope or range of items while keeping the number of deep structure propositions containing the items constant, his experiment does not rule out the possibility that Blumenthal's measure on deep structures may have some validity as well.

The research we have reviewed in this section represents attempts to predict memory performance on the basis of relationships among constituents of linguistic deep structures. Unfortunately most of the research was carried out in order to demonstrate that semantic relationships rather than surface structure characteristics are more important in remembering linguistic strings. Hence, there has been little if any attention given to developing a systematic approach to the question of what is remembered. Although important relationships hold among the constitutents of deep structures, relationships related to the meaning of the sentence, it is not certain this is the way we should be talking about memory. Certainly the research we have reviewed has not convinced us that it is. But no well-formulated alternatives exist.

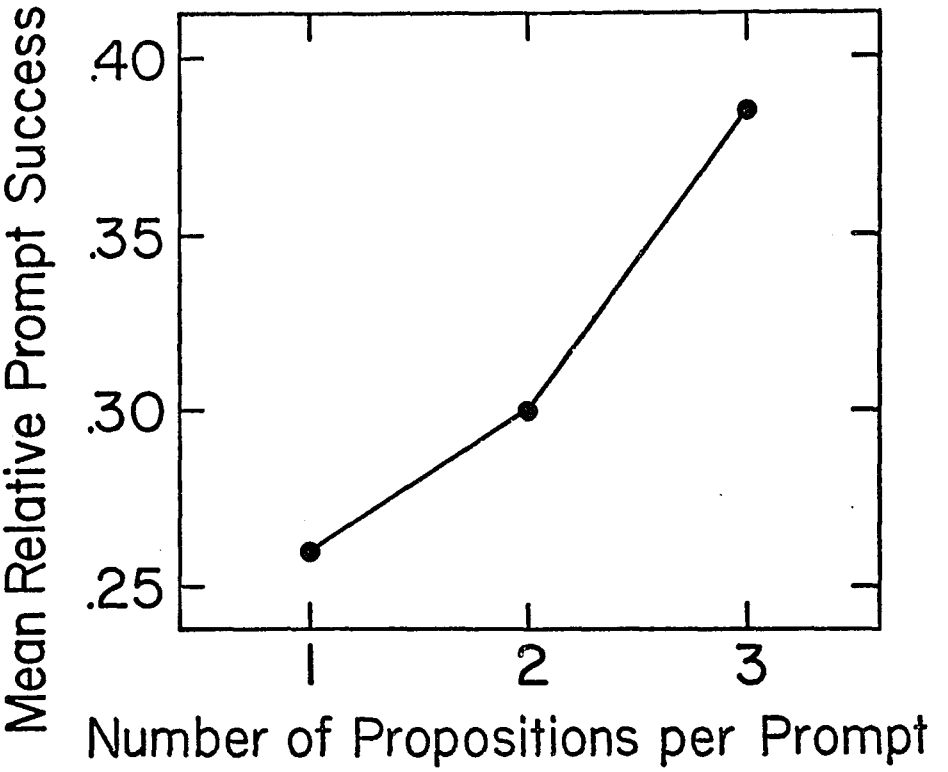

Figure 1.3. Relative prompt success as a function of the number of propositions per prompt word (after Wanner, 1968).

What is Remembered: Mental Images

Another important candidate for what is remembered from complex verbal material is mental imagery. There has been a rapid proliferation of work on the nature and use of mental imagery in cognitive functioning (e.g., Hebb, 1968; Paivio, 1969; Richardson, 1969; Bower, in press), and it is only natural that this renewed interest in imagery should lead to work on the role of imagery in the comprehension and storage of natural language. Recent work by Paivio and his associates (Paivio, 1970; Begg and Paivio, 1969; Yuille and Paivio, 1969) has revealed facts that have been largely overlooked by the majority of investigators with a more linguistic flavor to their work, yet these results are of major importance for our general question of what is remembered. Some research on the role of concreteness-abstractness in the representation of adjectives will be reviewed in the last section of this chapter.

Begg and Paivio (1969) questioned the generality of Sachs' (1967) results, and manipulated the concreteness-abstractness of the materials they used in her paradigm. Using independent groups of subjects as judges, Begg and Paivio obtained two sets of well-formed sentences separated on the basis of abstractness-concreteness. Experimental subjects heard both abstract (e.g., [1.23]) and concrete (e.g., [1.24]) sentences, and were later tested with sentences which were either

[1.23] The arbitrary regulation provoked a civil complaint.

[1.24] The spirited leader slapped a mournful hostage.

identical to one studied, changed so as to alter the meaning, or changed

so as to alter the form but not the meaning. For concrete sentences Begg and Paivio found the same results as Sachs: subjects were very good at recognizing changes of meaning but were only slightly better than chance in recognizing changes of syntactic form (i.e., changes that did not alter meaning). However, with abstract sentences exactly the reverse was obtained. Subjects were very good at picking out changes that altered structure but preserved meaning while they were extremely poor at picking out changes which altered meaning (in Begg and Paivio's study, all meaning-changing alternations were reversals of subject and object). The results are shown in Figure 1.4. This interaction was particularly striking because there was no main effect for abstractness-concreteness, that is, the overall level of performance did not differ for the two kinds of sentences. A re-examination of the sentences used in Sachs' study revealed that most of her sentences were concrete, so her results were consistent with the Begg and Paivio findings.

Why there should be such a striking interaction as a function of properties of the stimulus materials is not at all clear. The effect for concrete sentences seems straightforward. Subjects apparently encode some representative of the meaning of the sentences which is independent of the structural form of the original. Begg and Paivio argued that the subjects encode concrete sentences in terms of some kind of action imagery, their argument relying on informal subject reports elicited throughout Paivio's research on concreteness-abstractness (see Paivio, 1969). It is not clear how one would go about distinguishing between storing sentences as linguistic deep structures

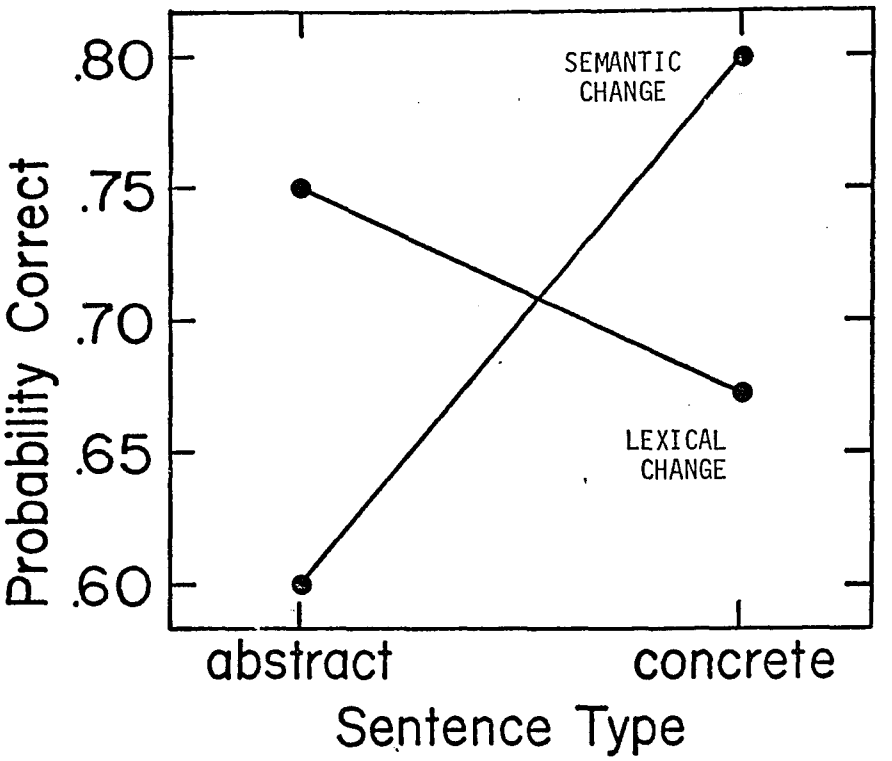

Figure 1.4. Correct recognition of semantic and lexical changes as a function of the abstractness-concreteness of sentences (after Begg and Paivio, 1969).

and as action imagery. It is likely, in fact, that both are used, or perhaps equivalently, that these are two different modes of access to some common underlying representation. However, the results for abstract sentences are puzzling. Begg and Paivio discussed the results in terms of subjects storing a direct representative of the linguistic string itself rather than some abstracted representation of the meaning. This would explain the ability of the subjects to recognize lexical changes that do not affect meaning, but leaves unexplained their inability to recognize permutations of subject and object. Begg and Paivio argue that the information is represented in terms of words and phrases, but without higher order grammatical information about the entire string itself. This view needs the support of additional data. The fact that the abstract sentences Begg and Paivio used were quite bizarre, which they were, limits the generality of their conclusions regarding this material but in no way helps us to understand the result they did obtain. The important point to be had from this experiment, however, is that the original Sachs result is subject to constraints having to do with the relative concreteness of the material. What the exact nature of these constraints might be must await further research.

Another important study was reported recently by Paivio (1970). This research stemmed from Rohrman's work in the recall of English nominalizations, reviewed in the preceding section. Rohrman attributed the difference in recallability of subject and object nominalizations to the extra PRO element in the deep structure of the object nominali-

zations, which created a greater load on the memory for these object nominalizations. Paivio reran one of Rohrman's experiments (Rohrman, 1968), but controlled for the relative abstractness-concreteness of the materials. Paivio assembled a pool of subject and object nominalizations which had been rated for imagery value (see Paivio, Yuille, and Madigan, 1968, for the technique). He selected stimulus materials from this pool in such a way that he could factorially manipulate both the relative concreteness of the materials and type of nominalization (subject or object) in the same experiment. When level of concreteness was held constant, there was no difference in the recallability of the two kinds of nominalizations. But there was a large difference in recallability as a function of level of concreteness independent of type of nominalization. Paivio had access to Rohrman's original stimulus materials and obtained imagery ratings for them along with the larger pool of materials mentioned above. He found that Rohrman's subject nominalizations had significantly higher average imagery ratings than the object nominalizations. Thus, it appears Rohrman's original results were due to the abstractness-concreteness of his materials, not to the difference in deep structure he originally postulated. These results of Paivio's are important, because abstractness-concreteness has not been controlled for in the studies we reviewed in support of linguistic deep structure as a predictor of memory performance. This suggests that considerable care must be taken in generalizing from the kinds of results reviewed in the last section. Further, it seems imperative now that extensive research be done on the

relationship between parameters of linguistic deep structures and the structural features of action imagery found in memory performance.

A final experiment of Paivio's group that is of direct relevance to the problem we are considering is one reported by Yuille and Paivio (1969) on abstractness-concreteness in memory for connected discourse. Three levels of abstractness (concrete, medium, abstract) were combined with two levels of organization (grammatical, random) to create six different 79-word paragraphs. Subjects studied one of these paragraphs for two trials, free recalling as many words from the paragraph as they could after each trial. Subjects learning the concrete material in syntactic order performed much better than any of the other groups. In fact, for subjects learning the medium and abstract material, performance was slightly better for the random presentation than for the syntactically ordered presentation. Of greater interest, however, were the details of the subjects' protocols. For the concrete materials, subjects with the grammatically ordered paragraphs recalled a much larger proportion of nouns than those with randomly ordered paragraphs. There was no difference in the proportion of nouns recalled as a function of organization for either the abstract or medium groups. In all levels of abstractness, more function words were recalled with random presentation than with syntactically ordered presentation. Finally, intrusion errors were anlyzed into theme-related and theme-unrelated intrusions. For the concrete material there was no difference in the proportion of theme related errors for grammatically or randomly ordered materials. However, for the medium and the abstract material, a larger proportion of the intrusions were theme-related with the

syntactically ordered presentation than with the random presentation (cf. Pompi and Lachman, 1967).

These results also demonstrated the importance of the abstractness-concreteness of the materials to be remembered. Not only was the overall level of performance different as a function of concreteness, but there were interactions in the internal analyses of the pattern of recall which suggested that very different things were going on in the concrete case. The pattern of responses suggested that imagery may have a very important role in the organization of thematic material, and may be a central form of storage in memory for verbal contexts larger than individual sentences.

The general lesson of these few results on the role of imagery in remembering complex verbal material is that any discussion of what is remembered from the comprehension of sentences or longer connected discourse must include imaginal encoding as a central component of whatever is going on. It is clear from the kinds of reanalysis of linguistically oriented research that Begg and Paivio (1969) and Paivio (1970) have done that most of the psycholinguistic research on memory for meaningful material has had an unreasonably narrow view about the possible forms of representation meaning might take. This is in part natural, since linguistic theory had already provided a well-formalized candidate for the canonical form of representation for semantic memory, i.e., linguistic deep structure. No similarly formalized structural model is available from the recent research on visual imagery. In fact, the most important problem in this area is the development of some kind of well-formalized structural model of imagery that will enable researchers

in memory to make more detailed predictions about the nature of remembering. Simply pointing out the gross differences as a function of stimulus characteristics, instructions, or individual differences will not be satisfactory.

What is Remembered: Semantic and Pragmatic Factors

In the metaphor of the transformational grammarian, we have been trying to develop the idea that some kind of deep structure, semantically based and more abstract than a direct copying of the linguistic string, is what is remembered from sentences. We have reviewed some evidence related to two candidates for such a model, actual linguistic deep structures and mental images. Neither of these has been developed in detail as a memory model, and the evidence we reviewed was in general sketchy and programmatic. There are many other factors that have been shown to be important in remembering sentential material, factors not necessarily related to a general hypothesis about what is remembered but important nonetheless. We cannot systematically review here these many factors and the research relevant to them, but rather will briefly indicate the role some of them have in the remembering of complex verbal materials.

These factors can be subsumed under several broad categories. First, a large number of important factors are associated with the use of our general store of information about the world. Each of us has in memory an extensive dictionary of information about words, knowledge about permissible combinations of experience, normative information on

the relative frequency of experiences and combinations of experiences, general presuppositional information, and specific kinds of factual information (as might be used to answer questions). We normally rely extensively on such information to help us understand sentences and to interpret and encode in memory the meaning of verbal material. This reliance on our general store of information can take many forms, and we will only look at a few examples.

Clark (1969) emphasized the cognitive importance of a particular lexical phenomenon, the distinction between marked and unmarked members of pairs of polar adjectives. For example, in the pair long-short, long is the unmarked member while short is the marked one. There are various kinds of linguistic evidence for this distinction. The unmarked member of the pair, long is the basis for the name of the dimension, length. Similarly, long has a neutral sense in certain contexts, like questions. How long is it? does not imply that the object is necessarily long, but the question How short is it? does imply shortness. This marked-unmarked asymmetry in polar adjectives is a very general linguistic phenomenon (see Greenberg, 1966), and Clark (1969) has shown that marked members of such pairs seem to have a more complex cognitive representation than unmarked members. In the deductive reasoning problems Clark investigated, the data made sense if one assumed that the unmarked member of the pair was the canonical form of representation for the elements of the pair in memory. That is, short is in fact represented in memory as something like not long, or to be more precise, the sentence X is short is represented as false (X is long). Clark and Card (1969) found that the marking of the marked member could be

lost, resulting in memory confusions with the unmarked member. Further, the memory confusions between members of such polar pairs were asymmetric toward the unmarked member.

Associative information as reflected in free association data is another example of permanent encodings of experiential regularities which are important for verbal memory. The relationships among words revealed by such data certainly have a role in the remembering of sentences and connected discourse, although such information by itself is insufficient as a general model of memory for complicated verbal materials. The modeling of associative relationships (e.g., Quillian, 1968) is an important step in specifying the kinds of dictionary information we have in memory, but work by psychologists (e.g., Fillenbaum, 1969) and by linguistis (e.g., Katz and Fodor, 1963; Fillmore, 1970) has shown the need for more complicated kinds of dictionary entries than mere nodes in lexical graphs.

So-called associative integration is an example of the role of general stored experience in the learning and remembering of sentences. In a sense this work can be seen as a logical outgrowth of Miller and Selfridge's (1950) work on recall of successive approximations to English. Miller and Selfridge's data is often interpreted as showing that as closer and closer approximations to ordinary English are obtained, individual words are recoded into well-integrated higher order units from permanent memory. Research on associative integration uses materials which are always grammatical but whose constituents have different degrees of associative relationships. Rosenberg (1966, 1968,

1969) has examined this phenomenon extensively. The sentence in [1.25] is a highly integrated sentence, whose words have high between-word associative strengths, while [1.26] is a poorly integrated sentence.

[1.25] The old king ruled wisely.

[1.26] The poor king dined gravely.

Stated in terms of the recoding principle, in remembering [1.25] one can rely upon previous experience, recoding the sentence into a unitized, effectively retrievable representation (like a visual image, perhaps), while [1.26], being a novel proposition, cannot be easily represented in well-integrated units on the basis of experience. The finding, of course, is that [1.25] is easier to recall than [1.26] (Rosenberg, 1968). Rosenberg (1969), in a different experiment, presented a more compelling demonstration of the importance of associative integration. Subjects studied a constant criterion material, a fixed set of words, within a sentential context which was either semantically well-integrated or poorly-integrated. This criterion material was more easily recalled when accompanied by the semantically well-integrated context (see Dolinsky and Michael, 1969, for a similar result). These results are consonant with the idea that the criterion material in the well-integrated context is remembered better because it is represented in terms of previously organized information in permanent memory. Such general recoding economies are extremely important in remembering sentential material.

A second broad category of factors are pragmatic factors associated with the communicative situation. These are elements of both the linguistic and non-linguistic contexts, and include knowledge of what is

being said or talked about, knowledge of the characteristics of the speakers involved, the interpretation of emphasis, tone of voice, and other factors associated with the physical transmission of the signal itself. Such factors are clearly important in the comprehension of the message, but they are also components of the message itself.

Turner and Rommetveit (1968) investigated the role of focus of attention in the recall of active and passive sentences. Their goal was to investigate experimentally the pragmatic difference between actives and passives. Suppose subjects had studied either [1.27] or [1.28]. They found that showing a picture of a rabbit or of a rabbit

[1.27] The rabbit was eating the carrot.

[1.28] The carrot was being eaten by the rabbit.

eating a carrot at the time of test facilitated the recall of [1.27] with respect to [1.28], while showing a picture of a carrot facilitated the recall of [1.28]. Showing a picture of the syntactic subject at the time of test facilitated the recall of passives and actives, indicating that the element emphasized in the sentence (logical subject in the active, logical object in the passive) is in some sense a better index of complete sentence content. They also found that when the picture of the syntactic object was shown, subjects tended to recall the sentences in the opposite voice of that studied. The net result of their findings, then, is that the pragmatic difference between active and passive sentences, which by most syntactic theories have equivalent deep structures, does make a difference in the effectiveness of sentence constitutents as retrieval cues.

A third kind of factor, which we will not consider in detail here, are affective factors. These factors have been largely ignored by recent psycholinguistic investigators, but have been extensively investigated by Osgood and his associates (e.g., Osgood, Suci, and Tannenbaum, 1957). The problem of whether we remember material which has positive affect better than material with negative affect has had a long and controversial history, and we shall not go into it here. But we should be aware of such factors, since they will be important in the development of a complete model for the representation of sentences in memory.

What is Remembered: Reprise

We are going to assume as metatheory for the ensuing chapters that when we understand a sentence and store it in memory we represent the sentence in some abstract, semantically based fashion, with little information about the specific syntactic or lexical form of the original sentence. Further, we are going to assume that deep structures posited by linguistic theory offer insight into certain features of this abstract representation, and that one task for research will be to determine which aspects of the linguistics of deep structures are relevant to mnemonic representation. We will also assume that mental images play a dominant role in most peoples memory for connected discourse, and that factors like the abstractness of the materials to be remembered will be an important parameter to consider in discussing the representation of linguistic materials in memory. As the brief survey of additional factors sketched in the previous section should have suggested, the hypo-

theses outlined above represent a great oversimplification of what is actually going on. But it is an oversimplification which we feel can lead to insight at the initial stages of development of a complete model of memory for sentences.

On the Cognitive Structure of Adjectives

In the next two chapters we will be presenting and testing some ideas about the mnemonic representation of adjective-noun relationships in sentential contexts. Although adjectives have often been used as materials in traditional verbal memory tasks, there has been little systematic study of them as a functional word class in the context of ordinary English sentences. Unfortunately, what little study there has been of adjectives as a functional word class in sentences has not included work on how they are remembered. Most investigations of connected discourse have concentrated on more wholistic characteristics of the material, and work seems to be just beginning which examines more detailed characteristics of such material. We shall review here some relevant research on adjectives as they are found in phrases and sentences as a background to the work to be presented in the next two chapters.

Adjectives seem to be highly susceptible to error when recalled in the context of a sentence. Gomulicki (1956) found that with relatively short, simple verbal materials the most frequently misremembered parts were adjectives. In fact, for the simplest, 13-word sentences he studied, adjectives were the only part misremembered with any frequency. With longer texts, more complex modifying phrases and clauses were the most

susceptible to misremembering. The general trend Gomulicki discovered was that the abstractive processes which simplify and distort connected discourse over time affected modificatory relationships the most. Martin and Roberts (1966) reported a similar finding. They found that the most frequent errors in free recalling simple sentences, apart from omissions, were substitutions of adjectives or adverbs in the original sentences. Once again, adjectives seemed to be the most poorly represented sentence constituent in memory. (This is relative to other constituents, since in absolute terms performance with respect to adjectives is quite good.) None of these investigators carried the work a step further and checked the nature of the intrusion errors obtained for adjectives, to see if the errors reflected any partial information the subjects might have retained. So these results do not tell us much about the manner in which modificatory relationships are represented in memory, only that however they are represented, they are more susceptible to distortion (or loss through abstraction over time) than other components of sentences or phrases.

Coleman (1965) compared the learning of sentence pairs which differed by a single transformation. He found that active-verb transformations were easier to learn in a multitrial immediate recall task than their corresponding nominalizations, active sentences were easier to learn than their passives, and nonembedded sentences were easier to learn than the corresponding embedded sentences. He also applied this technique to adjectivalizations and their corresponding detransformations. Sentences [1.29] and [1.30] give an example of an adjectivalization and

its detransformed adjective version. Subjects in this experiment were given two learning trials, and were scored on how many correct content

[1.29] The urgency of immediate demands is allowed to usurp attention.

[1.30] Immediate demands are so urgent that they are allowed to usurp attention.

words they recalled, scoring for correct singletons, correct pairs, correct triples, and correct intact sentences. On all of these measures the two types of sentence did not differ. Coleman alludes to differences among different classes of adjectivalizations but provides no data to support this claim. Thus, Coleman's results do not tell us very much about the nature of the cognitive representation for modificatory components of ordinary sentences. A further complicating factor in this research was that most of his materials were highly abstract (as in [1.29] and [1.30]), and as we have seen from the work of Begg and Paivio (1970) and Yuille and Paivio (1969) marked qualitative differences in remembering are associated with differences in abstractness-concreteness in sentential material.

Concreteness has been specifically implicated in the cognitive representation of adjectives. Lockhard (1969a, 1969b) found that if subjects who had studied simple adjective-noun pairs were later cued with one or the other word for recall of the entire pair, the adjective was less effective than the noun as a recall cue. The important factor in this asymmetry seems to be the relatively greater concreteness of nouns as a class than adjectives as a class (see, e.g., Lambert and Paivio, 1956; Paivio, 1963; Yuille, Paivio, and Lambert, 1969). Paivio (1969)

has reviewed many experiments that support the general principle that stimulus concreteness is more critical than response concreteness in recall. Thus, the superiority of nouns as recall cues in Lockhart's experiments seems to be consistent with an interpretation based on relative concreteness. Lockhart and Martin (1969) have explored this kind of effect further. They had subjects study adjective-adjective-noun triples and later used one of the three words from the triple as a recall cue for the entire triple. Not surprisingly, the noun was the most effective cue word. But Lockhart and Martin found that of the two adjectives, that adjective which was preferred by most speakers of the language to be closer to the noun was the more effective recall cue. For example, most speakers of English prefer the phrase the big red table to the red big table, and in fact red was a more effective recall cue than big for this pair. Lockhart and Martin attempted to relate preferred ordering to the dimension of abstractness-concreteness (see Martin, 1969a, 1969b), with adjectives of greater relative concreteness being preferred closer to the noun in English.

The linguistic fact about adjectives which seems to have interested psychologists most has been their relative ordering in the language when multiple adjectives are used in a noun phrase. These ordering relationships are extensive, and in fact Vendler (1968), in work which will be central to the theoretical developments in the next chapter, has systematized these ordering relationships for all classes of adjectives in English. Thus, the ostensible preferred orderings shown

in [1.31] and [1.32] can be described by the ordering of adjective classes in his theory of adjectives.

[1.31] huge scintillating luminous blue star

[1.32] horrible grinning toothless simian mouth

Psychologists have tried to assess the cognitive "reality" responsible for these preferred orderings. We have already alluded to Martin's explanation in terms of relative concreteness (Martin, 1969a, 1969b). Martin (1969b) had subjects rate adjectives on a variety of measures, and summarized this series of observations, "the most important correlate of adjective order considered . . . is definiteness of denotation [p. 701]." A rough correlate of definiteness of denotation discussed at length by Martin (1969b) is the extent to which choice of an adjective depends upon knowing the noun. In the phrase the big red table red is the more definite adjective since it is essentially invariant with respect to the noun it modifies. A red table is also a red object in the universe. Big depends upon the choice of the noun, since a big table need not be a big object in the universe. He relates this ordering principle to the accessibility of adjectives, and in a number of experiments with both English- and Indonesian-speaking subjects (Martin, 1969a) finds support for this reasoning. The basic idea is that the more concrete or definite an adjective is, the more accessible it is, and greater accessibility means it should take less time to produce a given adjective. Subjects are considerably faster, for example, at producing the value for a color than at producing a size when shown a picture of an object and given the dimension.

Bever (1970) has used the basic ordering arguments outlined by Martin but has elaborated a processing model which incorporates these ideas. He wanted to explain real-time processing of noun phrases, and proposed that the ordering of prenominal adjectives is one basis for deciding when a noun phrase has ended. The most definite or noun-like adjectives are closest to the noun being modified. By a perceptual analogy, this defines a gradient of noun-like-ness that culminates in the modified noun itself, and thus yields the following strategy as a model for establishing the boundary of a noun phrase:

> [1.33] After "determiner . . .." the boundary of the head noun phrase is marked by (1) a set of morpheme classes that signal the end of a noun phrase (such as "s") or immediately subsequent morphemes that signify the beginning of a new noun phrase (such as "the", proper nouns) or a relative clause (such as "that") and (2) a subsequent lexical item that is less uniquely a noun. [p. 323]

For this strategy to work, [1.34] must be unacceptable while [1.35] is

[1.34] *The plastic red large box . . ..

[1.35] The large red plastic box . . ..

acceptable. By the strategy in [1.33], [1.34] would be segmented after plastic, after red, and after box, while [1.35] would be correctly segmented only after box. Bever's general claim is that "the restriction on prenominal adjective ordering is an example of the effect of perceptual strategies on 'grammatical' structure [p. 325]," the perceptual strategy in question being the segmentation of a series at a break in the gradient. However, Bever ignores other segmentation cues like stress, intonation, context, and semantic well-formedness, all of which would

seem to be more important than his gradient idea.

A series of investigations by Oller and Sales (1969) explored the role of pragmatic factors on prenominal adjective ordering. They were interested in the communicative uses of language, in language performance within the extra-linguistic context. They state a principle of adjective ordering which takes into account pragmatic factors:

> [1.36] Relative to a particular context of situation, modifiers in the English NP are ordered from the least limiting to the most limiting proceeding away from the head noun, such that each additional modifier denotes a sub-class included by the denotation of the unit it modifies; when no restricting context is present, the order of high frequency contexts will prevail. [p. 222]

Oller and Sales demonstrated this principle by giving subjects examples from a collection of pictures of objects and asking them to give a noun phrase that described the object. For instance, the experimenter would point to a picture of an object with a number alongside it, and the subject would say, "Number 3 is a . . ..", where something like "big red triangle" would be inserted in the blank. The context was manipulated by the items making up the collection of objects, and hopefully changes in this context would be reflected by changes in the produced ordering of prenominal adjectives. Thus, if a set of three red triangles differing in size were the collection, a subject would say "big red triangle." But if the collection were three big triangles differing in color, the subject would say "red big triangle." This was found for a variety of contexts with both real and nonsense adjectives, supporting the pragmatic principle given in [1.36].

This is by no means a complete survey of recent work on adjectives. However, it is representative, and shows the lack of experimentation on the way in which adjectives are remembered. The work to be presented in the next two chapters was designed to obtain evidence for particular kinds of abstract structures in memory associated with adjectival modification.

Chapter 2

THEORY

Adjectives[1] and verbs have much in common, in that both basically predicate various things of nouns or noun-equivalents. This idea was at the heart of the Aristotelian analysis of language, where no distinction was made between adjectives and verbs. This kind of intuition about language is being recaptured in contemporary discussions of the deep structure of adjectives and verbs (e.g., Fillmore, 1968; Anderson, 1969). The separation of adjectives and verbs into different grammatical classes is probably a surface feature of languages like English. In fact, there are a large number of languages which have no grammatical class corresponding to English adjectives, which subsume the functional aspects of our adjectives in the grammatical class verb. One such language is Chinese. As Hockett (1958) points out, although the Chinese equivalents of such English adjectives as red or small form a separate subclass of verbs, they clearly have the same syntax as the Chinese equivalents of come and eat. The point of this, of course, is that the distributional class called "adjective" in English is an accidental phenomenon, not an essential one, when the nature of the functions performed by elements of this class is considered. In the previous

---

[1] As in Vendler (1968), a functional definition of "adjective" is being used here. The words large, steel, and bouncing are all functioning as "adjectives" in the phrases the large rock, the steel fence, and the bouncing ball, although as lexical items in isolation we are more likely to think of steel as a "noun" and bouncing as a "verb." Functional notions are, of course, more important at the level of deep structure.

chapter the evidence in support of a particular metatheory was reviewed. This metatheory stated that the important parameters of the representation of complex verbal materials in memory are more abstract parameters analogous to measures derivable from linguistic deep structures. Thus, in considering the way in which people remember prenominal adjectives in ordinary English sentences we will be less interested in the accidental properties of these materials than in the abstract, functional, deep-structure properties. Some abstract properties of prenominal adjectives will be presented in this chapter, and predictions based on these properties will be examined in light of some new empirical memory evidence to be presented in Chapter 4.

Why should we be interested in prenominal adjectives in our investigation of what it is that people remember from ordinary discourse? There are three basic reasons. First, by the kind of broad characterization we are giving this class of linguistic phenomena, the elements that make up this class are a very heterogeneous collection, as will be evident from the description of subclasses of adjectives to be presented later in this chapter. Elements of the substantive grammatical classes in English, namely noun, verb, adjective (narrowly defined), and adverb, can all function as prenominal modifiers. More important, in many cases it is difficult to distinguish their "normal" functioning as elements of their primary distributional classes from their functioning as adjectives. It seems very arbitrary, for instance, to call <u>glass</u> as "adjective" in [2.1] and a "noun" in [2.2] when it seems obvious that it is fulfilling the same function in both cases. Thus a wide range of

[2.1] the glass door

[2.2] the door that is made of glass

predicative functions are subsumed under the broad class of prenominal adjectives, and an adequate characterization of the mnemonic representation of this class of linguistic entities will give us broad power for the characterization of the conceptual relations underlying linguistic functions in general. A second, related reason for studying this grammatical class is that it is an easily identifiable class, by ordinary syntactic criteria, and furthermore, the broad scope of the class is available with extremely "simple" linguistic materials. All of the many subclasses of adjectives we shall outline later in this chapter can be used with trivial subject-verb-object sentences (although we shall not opt for this great simplicity in the present research). A third, and final, reason for examining this grammatical class is that there is already an extensive body of linguistic theory for it. Both traditional linguists and contemporary generative-transformational linguists have given prenominal modifiers a great deal of attention, so that a large body of systematic information, both formal and informal, is available.

This chapter will be organized as follows. First, there will be a brief background discussion of the development of the deep structure representation of adjectival modification, in the framework of fairly classical Chomskyan generative-transformational grammar. This will provide the basis for the next section, which will present a detailed outline of a theory of different classes of prenominal adjectives. Third, brief mention will be made of recent developments in linguistics that

may be relevant to the description of adjectival modification. Finally, the linguistic discussion will be tied to some claims about cognitive representations, and some predictions will be worked out for memory experiments.

Background

We have mentioned the heterogeniety of functions present among prenominal adjectives but we have not yet motivated this point, either linguistically or psychologically. In this section we want to examine some elementary linguistic facts and develop some structures that can capture the important features of these facts. Before dealing with these directly, however, it will be instructive to consider briefly some historical antecedents of the representation to be discussed.

The formation of nominal compounds is a linguistic process that offers some interesting insights into the more general problem of predicative modification. The problem of compounding has interested linguists for a long time, since it seems to be a "creative" process, a means by which new words or expressions are added to the language. We are talking about nominal compounds like freight train, ice cream, blackbird, door knob, crybaby, and the like, where words that can function as separate words according to distributional criteria function in particular contexts as an inseparable unit (a nice example from the vocabulary of children acquiring the language is the compound allgone, which typically enters the child's production lexicon before either of the parts enters).

A large body of linguistic literature exists on the problem of classifying nominal compounds according to various shemata. Many handbooks of language (e.g., Curme's old handbooks of English and German, Panini's analysis of Sanskrit) contain descriptions of nominal compounds which attempt to classify them on the basis of the semantic relations which hold between the constituents of the compound. The difficulties inherent in such an analysis are great, since well-defined semantic criteria have always been elusive. Grammarians like Bloomfield (1933) and Jespersen (1942) attempted to establish classificatory shemata based on syntactic criteria, by deriving or explaining a given compound on the basis of a phrase or clause that more adequately reflected the relationship between the elements of the compound. For instance, door knob really means knob of a door, talking machine is (figuratively) a machine which talks, laughing gas is a gas which causes laughing, and so on. These analyses seem very much like syntactic analyses of nominals and adjectives that we will be discussing in just a moment, but there are some important differences. First, these classical grammarians extended this analysis only to nominal compounds, not to adjectival modification in general. Second, the essential flaw in this work was the literalness and superficialness of the analyses. Such representations were not developed to account for what phrases like [2.1] and [2.2] have in common, but rather to account for why phrases like door knob and laughing gas seem different. That is, to use the jargon of contemporary linguistics, such analyses were solely in terms of surface structure, and a number of relationships that hold among the elements of compounds

could not be captured because the form of representation was not abstract enough. Nonetheless, these early attempts at classification were similar to current deep structure analyses of nominalizations and adjectives, so that this work can be seen as a historical antecedent to much of what we will be discussing in this chapter (see Lees, 1960, for a more complete discussion of the relationship between these earlier efforts at classifying compound nominals and more recent transformational analyses of nominalizations).

With the development of generative transformational grammars a particular kind of representation for modificatory relationships became prevalent, and can be seen as a radical extension of the kind of work we have just described. The distinction between deep and surface structure was crucial to this analysis. Singular modificatory predicates were expanded in the deep structure into more complete expressions (or kernel propositions, if you will) which were then embedded into the main sentence according to particular rules. Let us take the example of a very simple English sentence, like [2.3]. The deep structure of such a

[2.3] The small boy kicked the brown dog.

sentence would contain a set of simple propositions, like [2.4], [2.5] and [2.6], with [2.5] and [2.6] embedded in the appropriate places in

[2.4] The boy kicked the dog.

[2.5] The boy is small.

[2.6] The dog is brown.

[2.4]. This general form of representation has become standard for the explication of a variety of predications, simple and complex, and it

will be basic to the present analysis. (Recent developments in case grammars and generative semantics have changed the picture with respect to the general canonicity of these forms. However, the general points to be made in this section and the next would be compatible with a change in the formalization used to capture the regularities to be discussed here. This issue will be raised again in a later section of this chapter.)

It is obvious that one can predicate many different kinds of things of a given noun, and there are good reasons for wanting to be able to capture various aspects of these differences in the deep structure of adjective-noun relationships. In Syntactic Structures Chomsky has a nice example of this point. Although the examples in [2.7a & b] and [2.8a & b] have many similarities in surface structure, [2.7c] and [2.8c]

[2.7a] the sleeping child

[2.7b] the child is sleeping

[2.7c] *the very sleeping child

[2.8a] the interesting book

[2.8b] the book is interesting

[2.8c] the very interesting book

reveal that there is something very different about these noun phrases, for most speakers of the language would accept [2.8c] as well-formed but would reject [2.7c]. We can take this example as showing why, linguistically, it is important to distinguish among various classes of adjectives and to characterize the differences in a formal manner in the deep structure of the language. Later we will hope to show that such

differences are important for cognition as well, and that it will be useful to have a model of these differences for making statements about psychological phenomena like errors in remembering.

Work on transformational grammars of English has yielded a fairly standard way to represent the distinctions among the different classes of adjectives in the language, namely, by rewriting and expanding the adjective-noun pair in terms of a restrictive relative clause (see, e.g., Smith, 1961, 1964; Lees, 1960; Vendler, 1963, 1968). For instance, the noun phrase the red book would be rewritten as the restrictive clause the book which is red. The claim is that the deep structure of these two linguistic expressions does not differ, and that the similarity of meaning of these two examples or [2.1] and [2.2] mentioned earlier arises because they are derived transformationally from a common source. The advantage of spelling out a noun phrase in terms of a relative clause is that distinctions which are not obvious in the surface structure of pre-nominal adjectives can be made explicit in the equivalent relative clause. In fact, the difference between [2.7a] and [2.8a] is easily seen in the rewritings given in [2.7d] and [2.8d]. By putting these kinds of re-writings into canonical forms, a considerable amount of insight about the

[2.7d] the child who is sleeping ——> the sleeping child

[2.8d] the book which is interesting to read [for someone] ——> the interesting book

structure of adjectives in English can be captured formally.

The Structural Model

The most extensive analysis of adjective types in English has been done by Zeno Vendler (1968), and the general form as well as many of the details of the model about to be elaborated are derived from his work. But prior to going into the details of the model to be proposed here, a number of points should be kept in mind. First, Vendler's superb analysis was carried out as a linguistic and philosophical effort, not as a psychological one, so in a number of respects the results of his work must be taken as a rough framework within which to construct a partial model for the cognitive structures implicit in prenominal adjectives. Since Vendler's analysis purports to be exhaustive, a number of his classes will be ignored here. Some classes of adjectives occur primarily in other than prenominal positions, and the present concern is with prenominals only. Further, some classes contain only a small number of exemplars, and these are not very useful for experimental purposes.

Second, since the motivation for the present analysis differs from that of Vendler (and indeed from that of others who have discussed adjectival constructions, such as Ross, 1966; Smith, 1961; Oller and Sales, 1969), we shall want to make slightly different kinds of distinctions among the subclasses of adjectives to be considered. This means that sometimes Vendler's classes will be split, sometimes combined, in order to capture certain semantic distinctions with respect to the kinds of claims that will be made about the cognitive structure underlying human memory for complex verbal materials. We should think of the model in

terms of the kinds of information contained in the various formal structures, and how subjects might use this information in performing in memory tasks. The general claim is that the structures outlined in the model represent the kind of information that is input into whatever retreival or recognition scheme the subject might employ given a particular memory demand. In general, however, we will not be concerned with modeling the exact form of the cognitive structures involved, but rather, to use the language of Shepard and Chipman (1970), we shall be seeking an isomorphism of relations between structures in the model and structures in the mind. The specific implications of such an isomorphism will be detailed in the final section of this chapter.

Third, it should be remembered that the model is being developed in the context of an elaborate linguistic theory of English adjectives, and it is not clear how many of the fine points of the model are language dependent. Vendler's (1968) analysis made no claims to linguistic universality, but one would hope that the psychological claims from research such as that presented here would not be too heavily tied to a particular linguistic community. This is a general problem in work on verbal behavior, and it is not at all clear at what level of theoretical discourse one can feel certain that one is modeling cognitive structures characteristic of broad populations of humans. In the context of the present problem, for instance, we have noted that there are languages which have no adjectives in the surface structure. Rather than a distinction between adjectives and verbs, Chinese has a distinction between stative verbs and verbs of action (see Lyons, 1968, pp. 323-325, for a more

complete discussion, including comments on the relevance of the action-stative distinction for the deep structure of English). To capture the related aspects of the cognitive structures for speakers of Chinese one might wish to develop a very different model than the kind proposed here. As linguists move toward a more abstract characterization of deep structures, in which the distinctions among superficial grammatical categories become minimal (see, for instance, the work of Anderson, 1969, in preparation, for an attempt at an extremely parsimonious set of categories in the deep structure), one can hope for more generalizeable theories of cognitive structures.

Finally, the structures about to be proposed should be taken as only one component of a very complicated process. The psycholinguistics of even a piece of English sentences like the adjective is many-faceted, and the contribution of the present model should be viewed in that context. No claims for completeness are made here. Later we will of course have to discuss some of the other factors involved in the cognitive representation of these materials.

Figure 2.1 and Table 2.1 present a rough outline of the classification of English prenominal adjectives to be discussed here. The structure present in Figure 2.1 is intended to suggest how the various subclasses are interrelated, but the exact form of the tree-structure presented there should not be taken too seriously. The subclasses represented by the terminal nodes in Figure 2.1 and more fully explicatd in Table 2.1 are more important for our present purposes. Most of the formal rewritings in Table 2.1 are taken from Vendler (1968), although

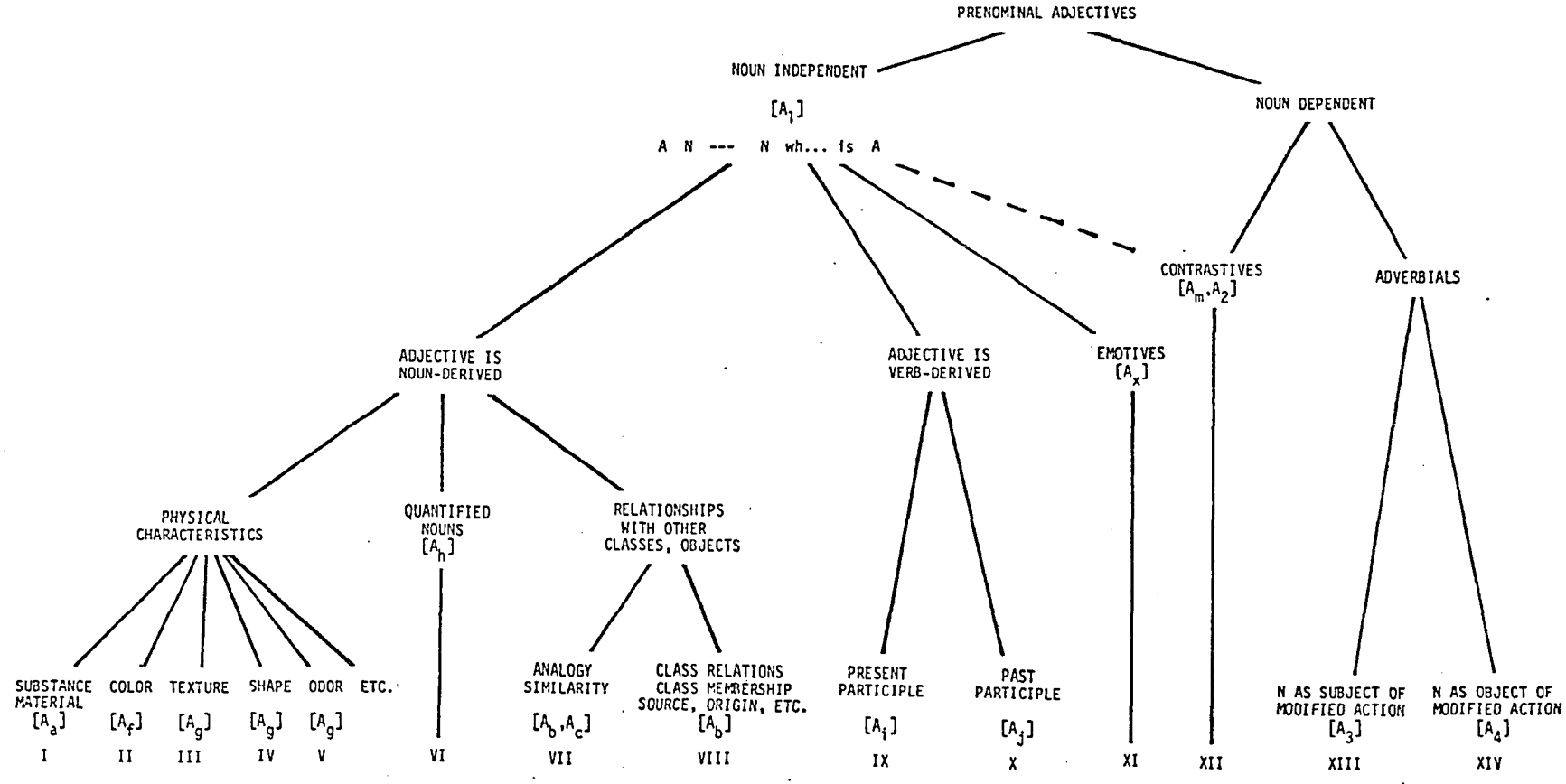

Figure 2.1. Outline of the relationships among the subclasses of prenominal adjectives to be considered in the present research.

Notes: a. The terminal nodes, given in this figure as Roman numerals, are more fully explicated in Table 2.1.

b. The expressions in brackets (e.g., $[A_3]$) refer to Vendler's (1968) descriptors for his classification scheme, and can be used to cross-reference the present classification with his.

Table 2.1

Explication of the Subclasses of Prenominal Adjectives as Given by the Terminal Nodes in Figure 2.1[a]

Subclass	Description	Formal Structure (A N ⟶)	Examples
I	substance, material of N	N wh... {consists / is made of / etc.} N*	iron bar ⟶ bar which consists of iron silk scarf ⟶ scarf which is made of silk
General Structure for Subclasses II - V, etc.		N wh... [ ] is A_{N*}	
II	color	N whose color is A_{N*}	red table ⟶ table whose color is red
III	texture	N whose texture is A_{N*}	granular stone ⟶ stone whose texture is granular
IV	shape	N whose shape is A_{N*}	square window ⟶ window whose shape is square
V	odor	N whose odor is A_{N*}	acrid fumes ⟶ fumes whose odor is acrid
VI	quantified nouns	N wh... {has / contains / etc.} QN*	verbose report ⟶ report that contains plenty of words shapeless rock ⟶ rock that has little [identifiable] shape colorful room ⟶ room which has lots of color
VII	analogy similarily	N wh... is like N* N wh... is like that of N*	childlike face ⟶ face which is like that of a child metallic surface ⟶ surface which is like metal

Table 2.1 (continued)

Subclass	Description	Formal Structure (A N ⟶)	Examples
VIII	class relations class membership source, origin etc.	N wh... {is / V} P N*	Bolivian ore ⟶ ore that is from Bolivia Wagnerian opera ⟶ opera that is [written] by Wagner canine teeth ⟶ teeth that are [found] in dogs children's shoes ⟶ shoes that are for children
IX	present state of activities of N, or properties of N with respect to current activities (typically of present participle form)	N wh... V [P N*]	floating ice ⟶ ice that is floating [on water] sleeping girl ⟶ girl who is sleeping dormant volcano ⟶ volcano which is in a state of inactivity ⟶ volcano which is "sleeping"
X	present conditions of N implicating or implying past activities, inferences from past activities (typically of past participle form)	N wh... is V [by N*]	broken pot ⟶ pot which was broken [by someone] aged wine ⟶ wine which has been aged
XI	emotives: attribute to N the capacity to evoke some kind of emotional reaction	N wh... evokes {causes} etc. reaction {emotion } etc. in N*	dreadful face ⟶ face which evokes dread [in someone] abominable monster ⟶ monster which evokes abomination (disgust, loathing) [in someone]

Table 2.1 (continued)

Subclass	Description	Formal Structure (A N —>)	Examples
XII	contrastives, measure adjectives, dimensionalized adjectives (often in polar pairs)	N wh... is A for [being] N	small elephant —> elephant which is small for being an elephant good thief —> thief who is good for being a thief weak king —> king who is weak for being a king —> king who is weaker than $king_{norm}$
XIII	adverbial, where N is the subject of the action being modified	N wh... V D_A	fast runner —> runner who runs fast —> runner whose running is fast [for being a runner] just king —> king who rules justly careful scientist —> scientist who experiments / theorizes / works / thinks, etc. carefully [for being a scientist]
XIV	adverbial, where N is the object of the action being modified	N wh... is D_A [P N*] to V	comfortable chair —> chair that is comfortable [for one] to sit on easy problem —> problem that is easy [for someone] to solve

[a]The following abbreviations are used in the formal notation: N = the modified noun, and in some cases the generic term for the class from which the noun is a member; A = adjective; A_N = noun-derived adjective; wh... = relative pronouns (who, whom, whose, which, what, that); P = prepositions; V = verb; D_A = adverbial form of adjective; Q = expressions for quantity like plenty of, lots of, little, no, etc; N* = noun different from the modified noun.

in a number of instances his somewhat cumbersome notation has been simplified. Also, Figure 2.1 contains bracketed references to the descriptors Vendler used to label his classes and subclasses (e.g., $[A_3]$, $[A_b]$).

The largest class of prenominal adjectives in English are those labeled A_1 in Vendler's notation (terminal nodes I to XI in Figure 2.1), and include those adjectives which have the property of being noun independent. As indicated in Figure 2.1, the general rewrite rule for this class of adjectives is given as in [2.9]. Noun independence refers to the transitivity which holds between the modifier and the noun, and the

[2.9] A N ⟶ N wh... is A

modifier and all (superordinate) classes of which the noun is a member. For instance, a red book is also a red object, a German girl is also a German female is also a German person, and so forth. This transitivity over class membership does not hold for the other major category of adjectives, the noun dependents. A large flea is not necessarily a large insect or a large animal. Similarly, a fast tortoise is not necessarily a fast animal. This general distinction between noun-dependence and noun-independence (not to be confused with noun-derived, to be discussed in a moment) is an important one, for in the case of noun-dependent adjectives there is generally more complexity of structure in the restrictive clause which is equivalent to the adjective-noun phrase. The precise nature of such dependencies will be explicated when we discuss those subclasses from the noun dependent portion of Figure 2.1 (XII to XIV).

First, however, let us consider the class of noun independent adjectives, which, as mentioned, includes the bulk of prenominal adjectives in English. Vendler distinguishes 13 subclasses of A_1 adjectives, but not all of these will be discussed here. In fact, only 9 (something will be said about a tenth later) of the 13 will be used in the present analysis, and as is evident from Figure 2.1 these subclasses will be broken down somewhat differently. The noun independent adjectives have been broken down into three broad categories in Figure 2.1: adjectives derived from nouns, adjectives derived from verbs, and emotives. As the discussion of the subclasses subsumed under these categories proceeds, the conceptual significance of the breakdown should become clearer.

Three general groupings of noun-derived adjectives are shown in Figure 2.1. The first includes those adjectives which specify the physical characteristics of the noun, and include the subclasses A_a, A_f, and A_g from Vendler's analysis. The reason these adjectives fall into different subclasses in Vendler's analysis has to do with the order in which speakers of the language prefer adjectives before a noun when there are more than one, but these distinctions will not be too important for the memory research to be outlined later. The examples given in Table 2.1 should make the distinctions among the subclasses clear. All types of subclasses from the A_g's have not been included here, simply because a few examples make the nature of the representation obvious. Certainly other perceptual-sensory dimensions like taste, brightness, and so on would fall into the general form shown in Table 2.1.

One important point needs to be stressed with respect to the subclasses of noun-derived adjectives having to do with physical characteristics. Many adjectives used to describe physical characteristics like size, texture, and brightness, for example, fall into the noun-dependent class of contrastive or measure adjectives (XII in Figure 2.1). There are good reasons, both linguistic and psychological, to believe that these belong in a different class because of differences in their deep structure. In other words, the separation is not arbitrary.

Adjectives in subclass VI are a peculiar lot, in that they actually cross many conceptual boundaries. They differ from other subclasses in that they specify quantitative characteristics of various attributes of nouns. In many cases adjectives in this class have superordinate relations with adjectives from other classes (e.g., <u>colorful</u> from VI, <u>red</u>, <u>blue</u>, <u>green</u> from II). These kinds of relationships will be important to keep in mind when the details of the research presented here are discussed later.

Finally, there are those adjectives which relate the modified noun to other nouns by means of analogy or similarity, class membership, class relations, origin or source, and so on. The subclass VII should be clear enough from Table 2.1, but VIII is an extremely diverse subclass which includes a large number of relationships. The rewrite rule for this subclass is repeated in [2.10] to facilitate the discussion of the diversity of this group. The descriptors given to VIII in Table

[2.10] $\quad A \quad N \longrightarrow N \quad wh... \quad \{{}^{is}_{V}\} \quad P \quad N^*$

2.1 are at best suggestive. An adequate representation of VIII would

further subcategorize this terminal node into a very complex structure which would not only reflect the semantics of prepositions but would indicate more exactly the relationships among the prepositions of the language and the kinds of verbs (in addition to to be) which can be "recovered" in the reconstruction of these A N's in deep structure (proposals for case grammars like Fillmore (1968) and Anderson (in preparation) seem to be promising directions for such an enterprise). All of this is to say that the representation here is crude, that the A_b and A_c subclasses of Vendler represent a very heterogeneous class of adjectives, and that a considerable amount of work remains to be done to be able to represent the semantics of these adjectives with any sophistication. However, the general form of the structure as given in Table 2.1 is clear enough for now, and both the discussion of the memory assumptions in the last section of this chapter and the examples of stimulus materials for the present research given in Chapter 3 should make it clear the the present level of analysis for VIII is sufficient for the purposes of the present research.

The subclasses of adjectives derived from verbs (IX, X) are fairly straightforward. These adjectives refer to the kind of activity presently characteristic of the noun (IX) or the kind of activity that has occurred in the past that is descriptive of the present status of the noun (X). Such adjectives are typically the present and past participles, respectively, of the verb in question. The rewriting of X is given in the form of a passive or a truncated passive, indicating that in some cases additional information is implicit in the construction

with reference to the agent of the past action. This brings up again the general point that much oversimplification has taken place in order to present Table 2.1 and Figure 2.1. This caveat is repeated only to discourage the serious consideration of the details of the system as some kind of complete and refined Hegelian classification for all those predicates which can be expressed linguistically in English as prenominal adjectives.

The final subclass of noun independents to be considered are the emotives. The explication in Table 2.1 should be sufficient, but the following clarification is needed. In many instances in the surface structure of English sentences it might not be clear whether a given adjective is an example of this subclass or an example of another subclass, like VI or one of Vendler's classes left out here. For instance, awful story (one of Vendler's examples for A_x) might not necessarily be referring to an emotion evoked in someone as in [2.11] but could

[2.11] story which evokes [or causes] awe [in someone]

rather be intended as a synonym for bad (in the sense of poorly written). Such ambiguity would be messy for the research to be outlined later, since one would lose control over which of the structures hypothesized by the model is in fact the one representative of the subject's encoding. Hence, adjectives from this class will only be used in cases where the context clearly disambiguates the intended sense. (This problem of adjectives which can function as members of a number of classes is a general problem since, as Jespersen once put it, "language-makers, that is, oridinary speakers, are not very accurate thinkers [1924, p.81]."

The control of ambiguity through context will be critical in the design of stimulus materials in the present research.)

The terminal nodes XII to XIV are particularly nicely formalized in Table 2.1, and their explication is also fairly straightforward. As discussed earlier, the noun-dependency of these subclasses arises because particular kinds of information must be recoverd in constructing the deep structure representation of the adjective-noun relationship, and this information is critically dependent upon what noun and what adjective are being considered. Moreover, the recovered information is of a sort that restricts the structural relationship to the particular noun and adjective, the modification (or predication) not being transitive to those superordinate classes of which the noun is a member.

Vendler's class A_m is a borderline class between the noun-independent A_1's and the noun-dependent A_2's. I have combined them (i.e., A_m and A_2) into one class, but have connected the node for contrastives to the superordinate node NOUN INDEPENDENT with a dashed line in Figure 2.1 to indicate this combination. The notion Vendler is after is that certain kinds of polar adjectives (e.g., warm - cool, happy - sad fat - lean) have a sense (often, though not always, context dependent) in which the formal structure of A_2 is not appropriate. He does not make the distinction between A_m's and A_2's very sharp, so in so far as possible all tokens in the present research from XII will be of Vendler's A_2 form. The rewrite rules given in Table 2.1 are in fact those for the A_2 class, and the rules bring out nicely the nature of the dependency the adjective has with the noun. Typically a normative value or

a point of comparison along a dimension is implicit in the modification. To use one of the examples discussed earlier with respect to transitivity, the noun phrase the large flea would be represented as in [2.12]. Here it is obvious that information regarding N_{norm} is

[2.12] the large flea ⟶ the flea which is large for being a flea
⟶ the flea which is larger than $flea_{norm}$

needed to interpret the modification. Since in general as we cross class boundaries or move from subordinate to superordinate categories $N_{norm} \neq N^*_{norm}$, transitivity of the kind discussed earlier does not hold. The subclass XII contains the important set of polar adjectives, which have a number of linguistically (e.g., Bierwisch, 1967) and psychologically (e.g., Clark, 1969) interesting properties. In particular, the distinction between the marked and unmarked member of a polar pair (e.g., tall is the unmarked, short is the marked member of that pair, since the question How tall are you? does not imply tallness, while the question How short are you? implies shortness; see Greenberg, 1966, Clark, 1969, for additional criteria) is an example of a kind of property that may be important in the evidence to be considered later.

The subclasses XIII and XIV contain adjectives that serve an explicit adverbial function in the deep structure representation. Linguistically, there is much in common between adjectives and adverbs, and Lyons (1968) has suggested there is little reason to differentiate between them in deep structure (as he writes, "the majority of 'adjectives' in English modify both nouns and verbs in deep structure [p. 327]."). At any rate, these subclasses make the relationship between prenominal

adjectives and adverbials particularly transparent. Adjectives from XIII essentially modify activities of which the noun is the logical subject, while those from XIV modify activities or actions for which the noun is the logical object. As the careful scientist example in Table 2.1 reveals, there is often a considerable amount of flexibility (ambiguity) in specifying the exact activity being modified, although in many cases in ordinary discourse these ambiguities would not be very great. In a number of instances, particularly those which employ contrastive adjectives as adverbials, one might wish to posit normative information of the form found in XII for the deep structure of XIII or XIV. Thus, the fast runner might be explicated as in [2.13]. However, the primary noun-dependency of XIII and XIV arises from the restrictions

[2.13] the fast runner ⟶ the runner who runs fast [for being a runner]

on the class of recoverable verbs for which the surface structure adjective serves as a deep structure adverb.

Additional Theoretical Issues

The preceding discussion was based on an approach to linguistic analysis which, although widely accepted by contemporary linguists, has been coming under attack recently (e.g., Fillmore, 1968; McCawley, 1968, Lakoff, in press; Anderson, in preparation; Postal, 1970; Schank, Tesler, & Weber, 1970; see replies to a number of these proposals by Chomsky, in press). The central theme of most of these proposals has been the reformulation of the notion of deep structure, either by including a

considerable amount of semantically-based information in it or by claiming that the notion of syntactic deep structure is untenable. There has been much interest in accounting for semantic facts without intervening syntactic structures (like Chomskyan deep structure), so that, for instance, fairly radical forms of suppletive relationships like [2.14] and [2.15] can be accounted for by separate transformational

[2.14] Herb sold the book to Eve.

[2.15] The book was bought from Herb by Eve.

derivations from a common underlying source.

There are two aspects of this work that are of relevance for us. First, these investigators have emphasized a number of very important linguistic facts, both for English and for languages in general. We have already mentioned some of these in passing, like the active-stative, surface verb-surface adjective parallelism alluded to in an earlier section. These facts are important to linguists because they are the kind of material upon which a linguistic theory hangs or falls. But these facts are also relevant to the kind of cognitive modeling being done in this dissertation and by the work reviewed in Chapter 1, since often these facts have considerable linguistic universality or point out regularities in English that cut across a wide range of differing surface forms. As emphasized in Chapter 1, these kinds of abstractions are likely to be fruitful sources of ideas for cognitive models. Furthermore, these kinds of facts are not necessarily easily compatible with the type of formalization sketched in the previous section. It is not difficult to see that the form of representation there is not very

abstract. Taking seriously ideas like the collapsing of surface verbs and adjectives or surface adjectives and adverbs in the deep structure would alter a number of the details outlined in Table 2.1 and Figure 2.1.

Second, the work mentioned above has introduced new proposals for the formalization of linguistic regularities that are very different from the kinds of formalization suggested in Table 2.1. The primary reason for choosing the kind of linguistic representation found in Table 2.1 was that this kind of representation is the most well-formulated and understood set of formalizations in contemporary grammatical theory. Some of the formal structures suggested by generative semantics and case grammars may be more appropriate as bases for cognitive models, but so far these structures have only been suggestive, not extensively developed. The incorporation of such formal structures into cognitive models must await further development of these formalisms, but is a project worth attempting when the time is ripe. Thus, we do not intend to ignore the large body of important linguistic work referred to in the first paragraph of this section, but it will not be dealt with systematically.

Memory Assumptions

We have now, by means of Figure 2.1 and Table 2.1 a formal representation for important characteristics of certain subclasses of prenominal adjectives in English. The question now at hand is whether or not the structures just outlined can yield informative propositions

about human cognition. In particular, will the relationships specified by these structures, taken as models of the cognitive structures established during the comprehension of prenominal adjectives in common English sentences, yield testable assertions about the way in which subjects remember adjectival information.

The central proposal of this dissertation is that the formalisms of Table 2.1, along with the roughly sketched interdependencies among subclasses shown in Figure 2.1, will yield predictions about the kinds of errors subjects will make in memory experiments. If (in part, at least) subjects represent prenominal adjectives in some structural form akin to those in Table 2.1, then confusion errors in recognition or production errors in recall should be predictable from the kinds of partial information yielded by degraded representations. The analysis takes on two aspects. The basic notion is that subjects may forget lower order information (e.g., specific lexical items, or, more abstractly, specific markings or features within a formally defined class of components or dimensions) but remember the subclass from which the adjective came and hence make more errors within this subclass. Additionally, they may forget certain delimiting structural features of the particular subclass, but remember more general features that are common to other subclasses, thus tending to make second order confusions to these related subclasses. Let us consider a concrete example. Suppose the subject had studied the sentence in [2.16] and at some later point was tested

[2.16] A small boy kicked the square stone.

for his ability to remember the adjective in the noun phrase, <u>the square</u>

stone. (How such tests will be conducted will be discussed in the next chapter.) Suppose also that the subject has forgotten the exact adjective, in this case square, but has some form of partial information at his disposal. At the time of study the subject constructed and stored in memory a structure analoguous to those in [2.17] as an additional

[2.17] the stone has the shape square
the shape of the stone is square
the stone's shape is square
(the stone whose shape is square, from Table 2.1 strictly)

kernel proposition to be embedded in the higher level proposition, A boy kicked the stone. The first assumption is that the abstract structure of the examples in [2.17], given in [2.18] (where α = shape, N_A = square), is the most resistant to forgetting, that the value of α (shape) is the

[2.18] N wh... [α] is N_A

next most resistant, and finally, that the value of N_A (square) is the least resistant. Thus, if the subject were forced to respond with an adjective, but had forgotten N_A = square, his most likely errors would be other shape adjectives like round or oblong. This means he has remembered what was predicated, i.e., a shape, but has forgotten the specific value of the predication. If he had forgotten α = shape, his errors would be adjectives referring to a variety of physical characteristics (e.g., α = color, α = substance). In this case, he has remembered a more abstract characteristic of the original predication. Errors from "unrelated" subclasses like XII through XIV would arise only if he had forgotten everything.

If structural descriptions of the form given by the rewrite rules in Table 2.1 are representative of the structures in memory, then the

degradation of such structures with delay could proceed as just outlined. This means that the form of the structure, even when degraded, carries information that may be useful to the subject when asked to remember an adjective. By means of the assumptions illustrated in the preceeding example, one should have a mechanism for rank ordering errors of remembering. This analysis assumes, of course, that certain other kinds of semantic factors (e.g., feature representations of lexical items, selection restrictions, associations, etc., depending upon what one's theory of semantics specifies as being important) are controlled. These will be discussed later.

Chapter 3

METHOD

The several experiments to be reported had a general methodology in common. Thus, a unified treatment of method will be given, with the differences among the several experiments discussed in the text. The parameters of the design for each of the experiments are tabulated in Table 3.3 at the end of this chapter, and can be consulted for a summary of the differences among the experiments.

Generation of Stimulus Materials

A large pool of simple English sentences was created by straightforward generation by the author and several independent generators, and by extensive culling of elementary school readers for first, second, and third graders. Such readers are a very useful source of simple sentences for experimental purposes, with the single drawback that the range of subject matter is usually fairly restricted. Approximately 400 sentences were obtained by these methods.

There were a number of important global constraints on the form and structure of these sentences. No strict criteria were imposed to assure complete homogeniety of structure, but the complexity of the structure was not allowed to exceed some informally defined bounds. Table 3.1 gives some examples of the kinds of sentences used in the present research, and suggests the general level of complexity of the sentences. The main criteria were that the sentences make sense, that they be non-

Table 3.1

Examples of the Stimulus Materials

1. The skilled carpenter built three maple cabinets to sell in his shop.

2. The fast boat sank in eight feet of water.

3. A wooden tower stood alongside the old building.

4. A small boy kicked the red wooden ball.

5. Judy wore a silk scarf.

6. The light basket had fresh strawberries in it.

7. The judge issued a warrent for the jewel thief's arrest.

8. The cowboys worked to keep the cattle together.

9. There were some snowy mountains near the boy's home.

10. The cute dog has blistered feet.

anomalous, and that they express a relatively simple, straightforward meaning. One of the more interesting probe recall studies using adjective-noun triples (Lockhart & Martin, 1969) used stimulus materials[2] that contained a number of anomalous or bizarre constructions (e.g., important salty butter, simple timid army). The difficulty with strange or anomalous materials is that, to the extent subjects try to comprehend or make sense of them, one is probably dealing with metaphorical uses of the materials rather than with what Weinreich (1966) has called "humorless, prosaic, banal prose [p. 398]," or what Bolinger (1965) has termed the "enumerable senses [p. 567]" of a word or phrase. The semantics of metaphoric usage is at best poorly understood (see, e.g., Bolinger, 1965; Weinreich, 1966, 1969; Bickerton, 1969, for remarks concerning the semantics of metaphor). The sentences were also chosen so that they contained nouns which could be modified by a variety of adjectives without making the sentences anomalous. The need to have sentences which made non-anomalous sense was important in the construction of the forced-choice recognition tests to be described in a later section. Finally, the sentences involved as little lexical and grammatical ambiguity as possible. Total lack of ambiguity was of course impossible to obtain.

A subset of these sentences was chosen which had meanings as different as possible from each other, and which minimized the duplication of particular nouns between sentences. Sets of adjectives were generated for nouns in these sentences according to the general formal

---

[2]I wish to thank Robert Lockhard for making these materials available.

constraints on the classes of adjectives to be outlined in the description of the recognition test procedure in a later section. Rare or bizarre adjectives were avoided in order to minimize von Restorf-like effects (e.g., one would not want to use a noun phrase like the refulgent jewel). Adjectives were chosen which would not lead to anomalous sentences or to metaphorical uses when inserted in the particular sentence frame. So-called "petrified compounds" (see Vendler, 1968; see also the discussion of compounding at the beginning of Chapter 2) were avoided. These are adjective-noun pairs like high school, Turkish coffee, or dump truck which have come to behave as multiple-word nouns, or pairs like young man, old lady, or little boy which have become so highly associated with each other that they systematically violate the ordinary ordering rules for prenominal adjectives. The latter are often used as forms of address in sentences like [3.1], and Vendler (1968, p. 132) points out that adjectives like old, young, and little are particularly

[3.1] Look here, young man, you should learn to treat your elders with more respect.

susceptible to this kind of relationship.

Only about half the sentences presented to a subject in any of the experiments were later tested, so a large number of filler sentences were used for control purposes. These filler sentences were of the same general nature as the experimental sentences, with a few exceptions to be noted later. Sentences 7 through 10 in Table 3.1 were filler sentences in Experiments 2 and 3.

There were no restrictions on the selection of adjectives for those portions of sentences which were not going to be tested later, nor for

filler sentences which were not tested at all. That is, even though the theoretical considerations outlined in the previous chapter were applied only to a subset of the classes defined by Vendler's (1968) analysis, and only adjectives from this subset were used for the adjective-noun pairings to be studied experimentally, exemplars from the other classes were not excluded from non-tested locations.

A casual survey of spoken and written English, in ordinary day-to-day contexts at least (I exempt Eric Sevareid and William Buckley from these remarks), reveals that adjectives are not as widely used as one might suppose. Perhaps this is the result of years of harping by schoolmarms and rhetoricians against the cumbersome nature of heavily adjectivalized prose. At any rate, it was discovered in running some pilot subjects that many of them notice while studying sentences like those in Table 3.1 that there were more adjectival constructions than usual, and they tended to pay special attention to them during the study session. It was hoped that subjects' use of special rehearsal schemes could be avoided, since the intent of the research was to study memory for adjectives under conditions approximating the comprehension of ordinary English discourse. Therefore, a small number of filler sentences were included among the study materials that were free of adjectives, in order to minimize the tendency for subjects to allocate more processing time to adjectives than they might under non-experimental conditions.

Construction of the Memory Tests

Three types of memory tests were used in the research to be presented here, two forced-choice recognition tests and a recall test. Each of these will be described separately.

Four-choice recognition test. The two recognition tests (including the eight-choice test to be described next) were constructed to test analytically the model outlined in the preceding chapter. The general procedure in these recognition tests was to insert a cluster of adjectives in the test sentence at the position an adjective occupied during the study presentation of the same sentence (occasionally two such clusters were embedded in a test sentence). The subject's task was to select the alternative he thought had been in the sentence as he originally studied it and to check it. Table 3.2 gives several examples of the two kinds of recognition test items as they appeared in the booklets actually given to subjects in these experiments.

An example of one of the four-choice recognition items will clarify the principle behind their construction. The adjectives from sentence 3 in Table 3.2 are given in [3.2]. The four-choice items were constructed

[3.2]
stone
round
wooden
square

with two adjectives from each of two classes, as defined by the theory outlined in Chapter 2. In the example reproduced in [3.2], the adjectives stone and wooden represent class I, and refer to the physical substance of the object being modified. Round and square are from class

Table 3.2

Examples of Recognition Test Items

1. The skilled carpenter built three
   walnut __
   little __
   maple __
   big __
   cabinets to sell in his shop.

2. The
   slow __
   aluminum __
   fast __
   fiberglass __
   boat sank in eight feet of water.

3. A
   stone __
   round __
   wooden __
   square __
   tower stood alongside the old building.

4. A small boy kicked the
   red __
   cracked __
   rolling __
   big __
   bouncing __
   little __
   green __
   dented __
   wooden ball.

5. Judy wore a(n)
   wool __
   new __
   silk __
   small __
   large __
   oblong __
   old __
   square __
   scarf.

6. The
   shallow __
   light __
   deep __
   straw __
   plastic __
   square __
   round __
   heavy __
   basket had fresh strawberries in it.

IV, and describe the shape of the object. Suppose the sentence as it appears in Table 3.1 had been the one the subject studied prior to getting this test item. Then wooden would be the correct alternative. However, if the conceptual classes and memory assumptions outlined in Chapter 2 have any validity, we would expect stone to be the most frequent confusion error made in this example. Relatively fewer confusion errors would be made to round and square. By constructing the test clusters as shown, these predictions of the relative frequency of confusion errors can be made for any particular adjective in a given test cluster. That is, the same test cluster can be used for each of four study sentences based on the insertion in turn of each of the four adjectives. This provided a control for response biases associated with a particular test cluster, insuring that the confusion data would not be confounded with idiosyncratic preferences for certain adjectives in a given sentential context.

The examples of four-choice recognition items in Table 3.2 show that there was considerable variety in which two classes were chosen to construct the set of four adjectives. The main constraint on which classes were used was simply the sentence context itself. For these kinds of test items, then, the only experimental manipulation was the between- versus within-class definition of possible errors. The second-order effects proposed in Chapter 2 were not examined. In principle the between-class "distances" could be recovered empirically by some scaling or clustering technique (such as Kruskal, 1964; Johnson, 1967) applied to the confusion errors resulting from such four-choice

recognition tests. However, the set of four-tuples of adjectives used in these experiments had properties that made this impossible. The four-tuples were constructed in many instances to maximize the theoretical difference between the classes (see Chapter 2) rather than to include all possible types of pairings of classes as would be needed to do a scaling or clustering analysis.

Eight-choice recognition test. An example of an eight-choice item is given in [3.3]. This example is taken from sentence 4 in Table 3.2,

[3.3]

red	green
big	little
cracked	dented
rolling	bouncing

and has been reproduced in [3.3] in a non-random order to facilitate the discussion of the principle behind its construction. The intention with eight-choice items was to test the hypothesized second-order effects discussed in the last chapter. Two adjectives were chosen from each of four classes, and the four classes were chosen so that they fell into two pairs of classes. These pairs of classes were defined analytically by the considerations in Chapter 2 so that the four adjectives in this subset would be more closely related to each other than to any of the adjectives in the other subset of four. In this example, red and green, from class II, and big and little, from class XII, would be more similar among themselves than to the other four items, since they all refer to the physical characteristics of the object in question. Similarly cracked, dented, rolling, and bouncing, all verb derivatives, should be more closely related among themselves, since they all refer

to activities either in the past or the present in which the object was implicated.

Suppose the sentence studied had been the version given in Table 3.1, with red as the correct choice. Then the relative ordering of confusion errors, based on the theoretical discussion in Chapter 2, should be as follows. The most frequent kind of confusion should be to the item in the same class, in this case green. The next most common error should be to items in the related class, either big or little. The least frequent errors should be to the other items. Again, the test clusters have been constructed to yield the same ordering of errors with any particular adjective as the input item, so a fixed test cluster can be used to control for response biases. Each test cluster yielded eight different input sentences.

Recall test. These items simply had a blank in the position of an adjective, and the subject's task was to recall what adjective had been in the sentence when he originally studied it. An example is given in [3.4].

[3.4] A _____ tower stood alongside the old building.

Control Subjects

In order to check the informal criteria used to generate stimulus materials for this research, control subjects were used to obtain various measures on both the sentences and the adjectives chosen to make up the test clusters. Several control procedures were used. First, after a set of sentences had been constructed with the four-choice adjective

clusters in them, a group of 15 subjects were asked to do the following. For each test cluster the subjects were to rank order the adjectives according to how well they thought the particular adjectives fit into the sentence frame. That is, they were to give the highest ranking to that adjective they thought made the best sentence, the lowest to that adjective they thought made the worst sentence. In addition, they were to put a star or asterisk alongside any adjective they thought made the sentence seem "ungrammatical, ill-formed, or bizarre," that is, unusual in any way. In addition, they were to star an entire sentence if it seemed "ungrammatical, ill-formed, or bizarre," regardless of the adjectives in the test position. A second group of 12 control subjects did this ranking task for a subset of sentences containing eight-item test clusters, and in addition, were given sentence frames with a blank in the position of a prenominal adjective and were asked to generate three alternatives that they thought fit into the particular frame. A third control group of 9 subjects did the ranking task for another subset of sentences containing both four- and eight-choice test clusters.

These control data were used to assist in the selection of stimulus materials to be used in the memory tasks. The subjects who generated three adjectives for blanks within sentences provided a pool of adjectives for use in designing test items, and provided a rough normative check on the frequency of particular classes and particular adjectives. Any sentence that was starred by more than two subjects in this ranking task was discarded and not used in the experiment. Also, particular adjectives that were starred by more than two subjects within

a particular sentence were not used in that sentence. Finally, any set of adjectives for which the control subjects showed a marked preference either toward or away from a particular adjective was either discarded or altered. As far as possible, the clusters of adjectives used to make up the test items had no strong biases associated with particular members (there were some minor exceptions to this, since biases could not be controlled completely as long as test items were bound by the constraints of construction outlined in the previous section).

Design and Procedure of the Experiments

The research was conducted in small groups ranging in size from 4 to 16 subjects (median = 12.0). Both the study and the test materials were presented in specially prepared booklets. The construction of these booklets will be described before the procedure is outlined. Table 3.3 can be consulted for a summary of the experimental parameters to be discussed in the next few sections.

Study booklets. In all of the experiments to be reported here, subjects studied 95 sentences in a single study session. The 95 sentences were presented in a five-page booklet, 19 sentences to a page (8-1/2 x 11 pages in all cases). In Experiment 1 there were four possible study lists, corresponding to the choice of one of the four adjectives in each of the test clusters. There were 50 test sentences, and these were combined with 45 untested filler sentences to make a complete study list. The order of the 95 sentences was randomized for each of the four study lists, and sets of 19 consecutive sentences

defined by this randomization were assigned to each of the five pages of the study booklet. A different set of study lists was prepared, and was used in both Experiments 2 and 3. There were eight different study lists for these experiments, corresponding to the selection of one of the eight alternatives in each of the eight-choice test clusters. Each adjective in the subset of sentences having four-choice test clusters (see Table 3.3) in these experiments was used twice in making up the eight study lists. There were 44 experimental sentences, 22 based on eight-choice test items and 22 based on four-choice items. These 44 sentences were combined with 51 filler sentences to make up the 95-sentence study lists. The order of the 95 sentences was randomized separately for each of the eight possible study lists.

In order to control for familiarity effects, most of the adjectives that would appear as distractors in the test items appeared at some time during the study session (76% of the adjectives used in the four-choice items in Experiment 1 appeared in the study lists, 84% of both the four-choice and eight-choice items in Experiments 2 and 3). Most of the distractors appeared in filler sentences, although some appeared in non-tested portions of test sentences or in actual test positions in other test sentences. As a result, the set of filler sentences (but never the non-tested portions of experimental sentences) varied slightly from study list to study list as some of the adjectives were changed in order to expose subjects to as many of the distractors as possible. An informal analysis of errors in Experiment 1, which had the lowest percentage of distractors fulfilling this familiarity control, revealed that it was

in fact an important control. Fewer errors were made to distractors which had not occurred at some point in the study list than to distractors which had occurred. However, adjectives not having the familiarity controls were distributed nonsystematically among the test clusters.

The five pages of each subject's study booklet were randomly permuted before they were stapled together with a cover sheet. This served as an additional control for overall order effects during presentation. Approximately equal numbers of subjects received each of the possible study lists in each of the experiments.

Test booklets. Each of the experiments had a slightly different test booklet. In Experiment 1, 50 sentences containing a total of 57 four-choice recognition test clusters (seven sentences had two test clusters) were presented in a five-page booklet, with ten test sentences per page. The overall order of the 50 sentences was randomly permuted twice to create two test-booklet forms. Within each of these forms the four adjectives in each test cluster were independently and randomly permuted, and were displayed in a top-to-bottom fashion as shown in the examples in Table 3.2. There were thus two test forms, differing both in the overall order of the sentences and in the order of the recognition choices within each of the 57 test clusters. In Experiment 2, both four- and eight-choice recognition items were used in a within-subjects design. There were a total of 44 test sentences, 22 each with four-choice and eight-choice items (one sentence had two eight-choice items). A six-page test booklet was constructed, with two pages containing three eight-choice

items and five four-choice items, and four pages containing four eight-choice items and three four-choice items. Two independent random assignments of the 44 test sentences were made into this booklet structure, yielding two test forms. The order of the adjectives in each of the test clusters was also randomly and independently permuted for each of these test forms, and the adjectives were arrayed top-to-bottom as shown in Table 3.2. For Experiment 3, the same 44 sentences used in Experiment 2 were presented, with a blank as in [3.4] instead of a cluster of adjectives. These 44 sentences were randomly ordered across three pages (15, 15, 14 sentences per page) in the test booklet. Two different randomizations were used to assign the sentences to the three pages, again yielding two test forms.

In all three experiments, then, two different test forms were prepared. Approximately one-half of the subjects having each of the possible study forms received one or the other of the test forms. As with the study booklets, the pages for each test booklet were randomly permuted for each subject before being stapled together with a cover sheet.

Procedure: Study Session. A paced study procedure was used, and each subject had a 5 x 8 index card to help with the pacing. For each page in the booklet the experimenter called out the numbers "1" through "19" in order, according to the particular presentation schedule for the experiment (either a number every five or every ten seconds--see Table 3.3), and subjects moved their index card below the sentence having that number. The subjects were told to study only that sentence whose number had just been called out, not to go back and look at earlier

sentences nor to look ahead to sentences not yet announced. The experimenter counted out each page in this fashion, allowing an extra five seconds between pages for the subjects to turn the page and position their card at the top of the next page. Subjects practiced the pacing procedure with a sample booklet page to be sure they understood it. All 95 sentences were studied in a continuous session. Out of 122 subjects run in the memory experiments, only one had difficulty with the pacing procedure and even then for an unimportant reason (she turned two pages at once).

The subjects were told to study the sentences for meaning. They were explicitly told that "the memory test will not ask you to recall whole sentences from memory, so you need not study for exact wording." In telling them to study the meaning of the sentences, meaning was defined for them "as that which underlies your ability to say something in more than one way, that is, your ability to paraphrase something." To clarify this, the experimenter went on, "Study the sentences as though I were going to ask you to paraphrase each of them later. The memory test will not be a paraphrase test, but it will tap the same ability."

Procedure: Test Session. In the research reported here all testing was done shortly after the study session. The study booklets were collected, the test booklets passed out, and instructions for the test were given, all of this consuming approximately ten minutes. The instructions varied slightly as a function of the kind of test given. The test sessions were unpaced.

For those subjects given a recognition test (Experiments 1 and 2),

a sample page inside the cover sheet of their test booklet contained examples of the kind of test item they would encounter (four-choice recognition items in Experiment 1, four- and eight-choice items in Experiment 2). The subjects were given a sample study sentence and the nature of a correct response was demonstrated for them with this example. Subjects were told that the test booklet contained a subset of the sentences they had studied. They were also told that when they were unsure as to which of the choices was correct they were to make a guess, but that in making this guess they should use any information about the meaning of the sentence they could remember.

Recall subjects were given a sample test item inside the cover of their booklets, and the nature of a correct response was demonstrated by means of an example. Again, subjects were told that if they could not remember the exact word that belonged in the blank they were to make a guess, using whatever they could remember about the meaning of the sentence to help them make this guess.

Design summary. Table 3.3 summarizes the experimental parameters that have been discussed here. Experiment 1 was essentially a between-subjects manipulation of presentation rate, with only four-choice recognition items used during the test session. Experiment 2 used a constant study rate, but had a within-subjects manipulation of recognition test item type (i.e., four-choice versus eight-choice). Experiment 3 used a recall test.

Table 3.3

Summary of Design and Experimental Parameters

Exp. No.	No. Ss	No. Study Sent.	No. Untested Fillers	No. Study Forms	Pres. Rate (Secs.)	Test Item Type	No. Test Sent.	No. Test Items
1a	15	95	45	4	10	4-RECOG	50	57
1b	25	95	45	4	5	4-RECOG	50	57
2	62	95	51	8	10	4-RECOG	22	22
						8-RECOG	22	23
3	19	95	51	8	10	RECALL	44	45

Subjects

All subjects in this research were students enrolled in the introductory psychology course at Stanford, and were participating in the experiments in order to fulfill a course requirement. 159 subjects participated in the research described here, 36 as control subjects and 123 as experimental subjects. Two experimental subjects were lost, one because she lost her place in the study booklet during the group pacing procedure and one because she decided not to participate when her consent was sought (which was a routine procedure followed with all groups of subjects, both experimental and control). The distribution of the experimental subjects by experiment and condition is tabulated in Table 3.3.

Chapter 4

RESULTS AND DISCUSSION

The experiments described in the previous chapter were designed to investigate a proposal for the way subjects would remember adjectives from ordinary English sentences. The focus of the research was the examination of first-order and second-order memory confusion effects proposed by the theory outlined in Chapter 2. The first-order effects are the confusions made within the same class of adjectives as the presented adjective, the prediction being that these should be the most frequent kind of error in the specially constructed four-choice and eight-choice recognition test items. The second-order effects are the result of relations among classes of adjectives. The most frequent errors should be to the class most related to the presented class according to the considerations specified in Chapter 2. The eight-choice items were designed specially to test for these effects. The theoretical rationale underlying the construction of the recognition tests was described in Chapter 2, and examples of their construction were provided in Chapter 3.

Experiment 1 was a between-subjects manipulation of presentation rate, all testing being done with four-choice items. Experiment 2 was a within-subjects manipulation of four- versus eight-choice test items. These two experiments provided the basic tests of the theoretical notions outlined in Chapter 2, and furthermore provided evidence of additional phenomena not directly considered by the theory. In this chapter we

will concentrate our attention on these two experiments, treating both the main effects sought by the design of the test items and some supplementary analyses of the data that proved to be very important. The recall data in Experiment 3 were collected largely to supplement the recognition experiments, to provide additional insight on behaviors obscured by the forced-choice paradigm. These data will be treated where appropriate in the context of the discussion of the recognition data. Table 3.3 in the previous chapter contains a summary of the design of the experiments, and should be consulted as needed during the following discussion.

Procedural Note: Performance as a Function of Presentation Rate in Experiment 1

In general subjects perform extremely well in recognition experiments with complicated verbal materials (e.g., Shepard, 1967), which creates a problem in running experiments like the present one. Subjects tend to make very few errors, a decided drawback when one is attempting to analyze errors. The large number of sentences presented to subjects for study in these experiments was one attempt to increase the number of errors. Experiment 1 also had a manipulation of the rate of presentation during the study session in hopes this would affect the error rate too. The effects of this manipulation turned out to be important, and provided a methodological constraint on the later experiments in this series. Since the results are only indirectly related to our general problem, however, we will explore them separately before examining

the confusion data in detail.

Table 4.1 contains the data for the four-choice recognition items for both conditions of Experiment 1 and for the subset of four-choice items used in Experiment 2. We will only look at the data for Experiment 1 now. It is evident from these data that presentation rate had a marked effect on the overall error rate. When subjects had ten seconds to study each sentence they made an average of 16% errors (an average of about 9 errors per subject out of 57 items), while those subjects who had only five seconds to study each sentence made twice as many errors (an average of 32%, or roughly 18 errors per subject). This difference was highly significant ($\chi^2 = 71.36$, $p < .001$; unless noted otherwise, all χ^2 values reported in this chapter are with one degree of freedom). But this was not the only important difference produced by varying the amount of study time per sentence. The conditional probability of making an error in the same class of adjectives as the correct one also varied as a function of presentation rate (or, more precisely, as a function of overall error rate). This conditional probability is an important statistic throughout this chapter, so let us consider it more carefully. In the four-choice recognition task described in Chapter 3 there was one correct alternative and three incorrect ones. These sets were constructed so that among the three incorrect alternatives or distractors there was one from the same class as the correct alternative and two from a different class. The conditional probability mentioned above was computed by dividing the number of errors made to the distractor from the same class as the correct

Table 4.1

Results for Four-Choice Recognition Items

Exp. No.	No. Ss	No. Items	Overall Errors	Cond. Errors	X^2	
1a	15	57	.16 (855)[a]	.45 (133)	8.83	$p < .01$
1b	25	57	.32 (1413)	.36 (461)	1.89	n.s.
2	62	22	.16 (1364)	.40 (216)	6.16	$p < .02$

[a]Numbers in parentheses in this and all subsequent Tables represent the number of observations used to compute the reported proportion.

adjective by the number of errors made regardless of class. If a subject's choice of a distractor were random with respect to the classes of adjectives this conditional would be .33, since he would choose the adjective related to the correct alternative no more often than he would choose one of the two adjectives from a different class. The hypothesis of first-order effects, of a greater number of confusion errors to adjectives from the same class as the correct one, is supported when this conditional exceeds .33.

The data in Table 4.1 show there was a relationship between the number of errors made overall and the conditional probability of making an error to the adjective in the same class as the correct one. Although twice as many errors on the average were made with the five-second study rate, the conditional probability for these same subjects is lower than that of the ten-second group. The two measures are independent, since a priori any given realtionship between the overall errors and conditional errors was as likely as any other. However, the data in Table 4.1 suggest that as the overall error probability increased the conditional probability of making an error to the adjective in the same class as the correct one decreased. The conditional probability for the ten-second group differed significantly from chance, indicated by the chi-square value in the rightmost column of Table 4.1. The performance of subjects in the five-second group did not differ from chance according to the same statistic. The conditional probabilities for the two groups differed marginally from each other ($\chi^2 = 3.63$, $.05 < p < .10$).

Internal analysis of the data in Experiment 1 supported the conclusion that there was an inverse relationahip between overall error rate and conditional error rate. The overall error rate and the conditional probability of making a related error were calculated for each of the 40 subjects in the experiment. A marginally significant negative correlation was found between the two statistics ($r = -.26$, $.05 < p < .10$, d.f. = 38), again indicating there was a tendency for subjects who made more errors overall to make relatively fewer errors to the theoretically appropriate distractor.

This inverse relationship, although unexpected, is understandable. The theoretical assumptions underlying the construction of the recognition test items had to do with how subjects might perform with incomplete information about the original experience. The memory assumptions outlined in Chapter 2 specified the kind of information a subject might have when he no longer has complete information, and predicted certain kinds of confusion errors as a function of this partial information. In order to provide the subject with partial information the memory representation of the experience must be degradable without being totally destroyed. At least some aspects of the represenation must be encoded well enough to remain accessible to the subject at the time of test. Subjects having only five seconds per sentence may not have been able to represent the sentence well enough to resist total memory loss as aspects of it were lost, and thus may have had little basis for making judgments with partial information. Subjects having ten seconds of study time for each sentence may have been able to construct and store more complete representations of the material, meaning

that fewer overall errors would be made while those errors that were made would be more likely to reflect the presence of partial information about the original experience.

These data influenced the conduct of the subsequent research. All subjects in both Experiments 2 and 3 were run with a ten-second study rate, meaning fewer errors were obtained from each subject. In order to analyze the recognition data from Experiment 2 with any reliability, therefore, a large number of subjects had to be run to provide sufficient numbers of errors.

First-Order Effects: Performance on Four-Choice Recognition Tests

The data from the four-choice recognition tests, shown in Table 4.1, support the hypothesis of first-order effects. With the proviso regarding presentation rate and overall error levels noted in the previous section, subjects consistently made more errors to distractor adjectives from the theoretically defined class containing the correct adjective than they did to adjectives from different classes. In both Experiment 1a and Experiment 2 the conditional probability of making an error to the related adjective was significantly greater than chance, as the chi-square values shown in Table 4.1 indicate. The difference was in the same direction in Experiment 1b, but was too small to be statistically significant. Although the effects attributable to recognition confusions to another member of the same class were small (.45 and .40, compared with the chance level of .33), the effects were remarkably consistent. Table 4.2 tablulates the number of subjects who exceeded

Table 4.2

Conditional Errors by Subjects for Four-Choice Recognition Items

Exp. No.	+[a]	0	-	p[b]
1a	11	2	2	.011
1b	16	1	8	.078
2	30	15	17	.016

[a]The three symbols indicate the number of subjects whose conditional probability exceeded chance (+), the number who fell exactly at chance (0), and the number who fell below chance (-).

[b]Binomial probabilities based on the Sign Test.

chance, were exactly at chance, or were below chance in the number of errors they made to the distractor from the same class as the correct adjective. The binomial probability computed according to the sign test (Siegal, 1956, pp. 68-75) is shown for each of the conditions, and in all cases the values either were smaller than or approached the .05 level of significance, indicating that the effect was consistent from subject to subject.

One reason for the smallness of the conditional probabilities tabulated in Table 4.1 is that the inverse relationship between overall error rate and conditional error rate discussed in the preceding section meant that subjects with the most errors and ipso facto the lowest conditional error probabilities contributed a greater relative proportion of the observations making up the conditionals. This weighting of the conditional probabilities by those subjects with lower individual conditional probabilities but higher error rates means that the values in Table 4.1 tend to underestimate the magnitude of the effect. Despite this, with the slower presentation rates the conditionals differ significantly from chance.

The data from Experiments 1 and 2 support the prediction that the most frequent confusion error with forced-choice recognition tests would be the choice of the adjective in the same theoretically defined class as the correct alternative. This supports the principle developed in Chapter 2 that as information in memory is degraded the nature of the predication is less susceptible to forgetting than the exact value of the predication. For instance, when a subject misremembers bouncing ball

as rolling ball more frequently than as blue ball it would seem he is remembering that the original modification talked about some activity of the object, not about the physical characteristic of its color, even though he cannot remember the precise activity that was originally mentioned. He is remembering a more abstract characteristic of the original verbal experience, as was discussed in Chapter 2.

There are alternative ways of talking about these results. For instance, one might wish to argue that data like those reported in Table 4.1 are due to associative relationships among the verbal units used in the experiments. The version of the associative argument to be discussed here is based on the frequency theory of verbal discrimination (Ekstrand, Wallace, & Underwood, 1966). It is assumed that subjects make implicit associative responses (IARs) while studying a word, and the more highly associated another word is the more likely it will be given as an IAR. Thus, while studying the word table a subject might make chair as an IAR. The occurrence of chair as an IAR increments the frequency count for it. Later, when faced with a forced-choice recognition test, a subject assesses the relative frequency counts of the test words and chooses that one which has the highest count. When high associates of the target word are included as distractors, subjects are more likely to make errors to such associates than to neutral words because of the previous increments in the frequency counts for high associates due to IARs. Data from a variety of recognition tasks have supported the predicted increase in false alarms to associatively related words (e.g., Underwood, 1965; Underwood & Freund, 1968; Anisfeld & Knapp, 1968; Hall,

1969; Hall, Sekuler, & Cushman, 1969).

In the present experiment, the theory just outlined would predict that more recognition errors would be made to adjectives from the same class as the correct one than to adjectives from a different class since on the average the former are higher associates of the correct word than the latter. For instance, round is a relatively high associate of square, red is a high associate of green, big is a high associate of little, and so on. It was impossible to design the recognition test clusters with average associative strength among the adjectives controlled, so a post hoc analysis was used to assess the validity of the associative argument. If the reasoning behind the associative argument is correct, then words more highly associated should be more susceptible to confusion than less highly associated words. Fortunately a very large class of adjectives used in the present research differed from all others in the magnitude of the associative relationships among the members of the class. Contrastive or polar adjective pairs like big-little, fast-slow, and long-short are much more highly associated with each other than pairs taken from non-contrastive classes. If false recognition effects shown in Tables 4.1 and 4.2 were due to associative relationships, the effects should have been greater for polar adjectives than for nonpolar ones.

The adjectives used in Experiments 1 and 2 were partitioned into polar and nonpolar subsets, and the analyses of Table 4.1 were repeated for these subsets. The data about to be reported for polar adjectives represent subjects' behavior when a polar adjective was the correct

choice on a recognition test. Recall that the within-class distractor for a polar was always its antonym. Similarly, the data for nonpolars represent performance when a nonpolar adjective was the correct choice. Often the set of four recognition choices was made up of two polar and two nonpolar items. The results of the analysis of polars and nonpolars appear in Table 4.3.

Let us examine first the effect of this partition on the overall error rate for the two subsets. In both Experiments 1 and 2 the overall error rate (Pr (Err) in Table 4.3) was higher for polar adjectives than for nonpolars, nonsignificantly in Experiment 1 and significantly in Experiment 2. The appropriate chi-squares appear on the line below the relevant probabilities in Table 4.3. Subjects had more difficulty identifying the correct alternative when it was a polar adjective than when it was a nonpolar one. This would be consistent with the associative position outlined before, since with polar adjectives the distractor set would have more associative "pull" than with a nonpolar adjective, making it more likely that an error would be made. However, not only must there be more errors for polars, a greater proportion of errors must be within the same class than with nonpolars. The conditional probabilities in Table 4.3 do not support this. The conditonal probability of making an error to the related distractor, given an error was made, was at chance for polar adjectives in both experiments. On the other hand, the conditional probability for nonpolars was greater than chance for both Experiment 1 ($\chi^2 = 6.12$, $p < .02$) and for Experiment 2 ($\chi^2 = 15.97$, $p < .001$). The important point with respect to the associative argument is that in both experiments the conditional probability

Table 4.3

Polar versus Nonpolar Adjectives with Four-Choice Items

	Experiment 1		Experiment 2	
	Polar	Nonpolar	Polar	Nonpolar
Pr (Err)	.29 (575)	.25 (1697)	.21 (556)	.12 (805)
	$\chi^2 = 2.60$, n.s.		$\chi^2 = 22.11$, $p < .001$	
Pr (Err_w|Err)	.36 (165)	.39 (429)	.32 (118)	.53 (95)
	$\chi^2 = .51$, n.s.		$\chi^2 = 9.06$, $p < .01$	

for nonpolars was greater than the conditional probability for polars. The difference was nonsignificant in Experiment 1, but was significant in Experiment 2. See the chi-square values in Table 4.3.

These data indicate that polar and nonpolar adjectives had very different characteristics with respect to the kind of memory demands made on subjects in the forced-choice recognition tasks in Experiments 1 and 2. The theoretical position outlined in Chapter 2 would, if anything, predict a difference in the same direction as the associative hypothesis. The representation for polar adjectives typically contains information about the normative value of the dimension as well as the dimension itself and the value of the predication on the dimension, suggesting that the representation for polars contains more redundancy than that of nonpolars. For example, the representation for large and red are given in [4.1] and [4.2] respectively. If these are roughly representative of the structures in memory, and if components of these struc-

[4.1] the large book ⟶ the book whose size is large for being a book [with respect to book$_{norm}$]

[4.2] the red book ⟶ the book whose color is red

tures can be lost separately, then the polar adjective large should produce more confusions within the same class since the encoding contains one more reference to the dimension (i.e., a reference to the normative value of the dimension) than does the encoding of red. But by the data of Table 4.3 this reasoning is not supported. On the basis of the data one would suspect that the reasoning about partial information was correct for nonpolar adjectives, since significantly more confusions are made to the adjective in the same class as the correct one than would

be expected by chance. But nonpolars would appear, on the basis of the data presented so far, to have been represented in an all-or-none fashion. Either the original adjective was recognized correctly, or the subject chose among the forced-choice alternatives randomly. Additional data will have to be brought to bear on this, however, before these conclusions can be supported. The relevant data are in Experiment 3, but will be examined only after we have treated the eight-choice recognition tests that were administered in Experiment 2.

First-Order and Second-Order Effects: Performance on Eight-Choice Recognition Tests

The eight-choice test items in Experiment 2 were designed to investigate both within-class and between-class confusions. The assumptions outlined in Chapter 2 and the examples given in Chapter 3 specified the rationale involved in the construction of these items. There was one distractor from the same class as the correct adjective, two distractors from a theoretically "close" category, and four adjectives from theoretically more "distant" classes. The clusters were constructed so this set of relationships would hold with any of the eight adjectives as the correct choice. The predictions were that the most frequent confusion error would be to the single adjective from the same class as the correct one, the next most frequent confusions would be to the two adjectives from the theoretically related class, and the least likely confusions would be to the other four adjectives.

The overall analysis failed to support these predictions. The average error rate for these test items was .22, compared to the error rate of .16 for the four-choice items with the same subjects. This difference in error rate was significant ($\chi^2 = 20.76$, $p < .001$). Two conditional probabilities were computed for the eight-choice items. First, there was the conditional probability that a subject chose the single adjective in the same class as the correct choice, given that he had made an error. Chance for this statistic was .14. The second conditional was computed by taking the number of errors made to the two adjectives from the theoretically "close" class and dividing them by the total number of errors. The chance value for this conditional was .29. The observed values for these two conditionals were .16 and .29, respectively, and they did not differ from chance in either case. There appeared to be no evidence for partial information effects in the overall pattern of responses to the distractors, not even the first-order effects found with the four-choice tests.

Before trying to assess the reasons for this apparent lack of evidence for appropriate kinds of partial information effects in the eight-choice data some additional evidence was examined. As before, the difference between polar and nonpolar adjectives turned out to be very important. The overall error rates and the two conditional probabilities described previously were recomputed for the two subsets of data corresponding to either a polar or a nonpolar adjective as the correct alternative in a test set. The data of interest are in Table 4.4. The conditional probabilities relevant to the second order effects do not

Table 4.4

Polar versus Nonpolar Adjectives in Reduced Analysis of Eight-Choice Items

	Polar	Nonpolar
Pr (Err)	.32 (545)	.17 (880)
	$\chi^2 = 43.18$, $p < .001$	
Pr (Err_w|Err)	.13 (174)	.21 (149)
	$\chi^2 = 3.32$, $p < .10$	

appear in Table 4.4, since for both polars and nonpolars the values did not differ from chance nor did they differ from each other. The reasons for this apparent lack of second-order effects will be discussed later. The data that are entered in Table 4.4 follow a pattern remarkably similar to that found in Table 4.3. The error rate for polar adjectives was significantly greater than the error rate for nonpolars, as was the case for the equivalent analysis of the four-choice items in Table 4.3. Once again subjects had more trouble correctly recognizing a polar adjective they had studied than a nonpolar one. Similarly, when they made an error with a polar adjective as the correct choice they chose the other member of the polar pair no more often than expected by chance. However, when the correct alternative was a nonpolar and subjects made an error they chose the distractor from the same class significantly more often than expected by chance ($\chi^2 = 5.14$, $p < .05$), indicating they retained partial information about the nonpolar when they had forgotten the exact adjective. The difference between conditionals for polars and for nonpolars approached significance, as shown by the chi-square in Table 4.4.

As was the case with the four-choice data, consistent differences were found between polar and nonpolar adjectives in the eight-choice data. The hypothesized first-order effects were present for nonpolar adjectives but not for polars. The magnitude of the effects for nonpolars was smaller in the eight-choice data than in the four-choice data, but this can probably be attributed to the greater number of distractors in the former case. At any rate, the data in Table 4.4 replicate those

in Table 4.3, and support the conclusions reached about the latter.

The data do not contain the proposed second-order effects, either in the overall analysis or in the analyses for polar and nonpolar adjectives. Additional internal analyses failed to uncover evidence in support of the prediction. The failure of the model sketched in Chapter 2 to provide insight into the relations among classes of adjectives in memory can be tentatively attributed to several factors. First, it may have been that the forced-choice design did not provide an adequate test of the effects. Although it is always difficult to ensure that one's measurements are adequate for detecting a predicted phenomenon, the methods employed in the present research seemed adequate. But a more serious problem resided in the design of particular sets of forced-choice test clusters. In order to satisfy the requirements of the theoretically imposed metric on classes while still keeping the sentences in good English large numbers of polar adjectives were used. In fact, of the 23 eight-choice test clusters in Experiment 2, only 5 contained no polar adjectives at all, and 15 contained at least four polars in the set of eight. Since polars and nonpolars have been found to be so different in the research reported here it could be that the second-order effects were obscured by the presence of all these polar adjectives. However, separate analyses of the 5 sets of eight-choice clusters containing no polars at all plus the 13 sets which had four nonpolars contrasted with four polars provided no evidence of second-order effects.

A second reason for not finding second-order effects is theoretical. The metric was informally defined, and moreover was based on

similarities among essentially syntactic (albiet deep structure) descriptions of adjective-noun phrases. The only reason for considering these structures as a model for higher order relationships was that they covaried extensively with "notional" or semantically-based descriptions of adjectival predication. However, by the standard of post-Aspects developments in generative grammars, abstract syntax, and case grammars (mentioned briefly in Chapter 1), the formalisms implicit in the descriptions of adjective classes given in Chapter 2 were not very abstract. As formal developments proceed in this current linguistic work, perhaps more adequate representations for the modeling of higher order relationships will become available.

Supplementary Data: Recall Evidence from Experiment 3

Experiment 3 was run to assess the kinds of errors subjects would make when not constrained by the alternatives of a forced-choice recognition test. Subjects were forced to respond if they could not remember exactly, but were encouraged to use whatever information they remembered about the entire sentence to help them make a guess. As a result there were many intrusion errors for analysis in this experiment.

Not unexpectedly, the overall error rate was much higher with recall tests than with recognition. The data were scored in two ways, according to a strict criterion in which synonymic intrusions were counted as errors and a lax criterion in which synonyms were counted as correct. With the strict criterion subjects recalled 53% of the adjectives correctly (the total number of observations was 855), while with the lax

criterion subjects recalled 58% of them correctly. The conditional probability of a subject's making an error with an adjective from the same class as the correct one, given that he made an error, was .28 for the strict criterion (synonyms counted as errors in the same class) and .20 with a lax criterion (synonyms treated as correct responses, not as errors). There was no way, of course, to compare these values to chance with the recall paradigm.

The important analyses for our purposes come from the partition of the data into those observations for which a polar adjective was correct and those for which a nonpolar was correct. Table 4.5 presents these data, both for the strict and for the lax scoring criteria. First of all, with either scoring criterion significantly more errors were made when polar adjectives were the correct choice than when nonpolars were. This is consistent with the data collected with the recognition paradigms. Similarly, more errors were made in the same class as the correct adjective when the correct choice was a nonpolar than when it was a polar adjective. This difference was significant with the lax scoring criterion but not with the strict criterion. One difficulty with this difference in the recall data is that by the nature of the classes there are many more possible within-class errors with nonpolars than with polars. Even when synonyms are counted as errors, there are usually only three to five possible within-class responses possible for a polar adjective while for nonpolars the number of possible within-class intrusions often cannot be practically estimated.

Table 4.5

Recall Data for Polar and Nonpolar Adjectives

	Polar	Nonpolar	X^2	
$Pr\ (Err)_{strict}$	.56 (327)	.42 (527)	16.12	$p < .001$
$Pr\ (Err)_{lax}$	.48 (327)	.38 (527)	7.77	$p < .01$
$Pr\ (Err_w \mid Err)_{strict}$	.27 (182)	.30 (219)	.37	n.s.
$Pr\ (Err_w \mid Err)_{lax}$	.15 (157)	.24 (202)	3.96	$p < .05$

The data in Table 4.5 show that polars and nonpolars differed in the way in which they were affected by the two scoring criteria. The effects can be summarized succinctly by pointing out that more synonyms were given as intrusions for polars than for nonpolars ($\chi^2 = 3.79$, $.05 < p < .10$). This leads to an interesting result. We can examine only those cases where the subjects have recalled the class or dimension of the correct adjective by summing their correct recalls, their synonymic intrusions, and their intrusions from the same class as the correct adjective. If we examine the conditional probability that a subject recalled the exact adjective he studied, given that he recalled the dimension or class, then subjects do this significantly better with nonpolars than with polars (.83 with nonpolars, .75 with polars, $\chi^2 = 4.87$, $p < .05$). However, if we take synonymic intrusions as reflecting recall of the meaning of the original adjective, then given that a subject recalled the dimension or class of the correct adjective, he did just as well recalling the meaning of a polar adjective as recalling the meaning of a nonpolar one (.88 for polars, .87 for nonpolars).

These results offer some insight into the recognition data of Experiments 1 and 2. Those data indicated that subjects made confusion errors when nonpolars were correct which reflected memory for the functional class of the correct adjective, even though in all these cases the distractor from the same class as the correct adjective was neither a synonym nor an antomyn but an adjective that fulfilled the same function as the original one. On the other hand, those data showed that if a subject made a recognition error when a polar adjective was the correct choice he was no more likely to have picked the antonymic mate of the

correct adjective than if his choices among the distractors had been random. Given the recall data from Experiment 3, some sense can be made of these differences. If we assume that subjects remember the predicative dimension of the original adjective while also remembering something about the denotative value of the original adjective on this dimension, then the lack of first-order confusion effects for polar adjectives would result from the design of the forced-choice clusters. What the subject remembers about the value of the dimension may be very imperfect, but at least it is not likely to be antonymic with respect to the denotation of the original adjective. If a subject were given a recognition test when a polar adjective was the correct choice, and he remembered partial information which reflected the denotative value of the adjective he had studied, he would find that either the correct choice or a choice outside the same class would be consistent with this imperfect representation, but that the distractor in the same class as the correct adjective would not. That is, he would be unlikely to remember exactly the opposite of what he had studied and choose the antonym of the correct adjective. He might remember, for example, that the original adjective predicated something about the physical extent of the modified noun, and remember a rough quantitative estimate of the magnitude of the value predicated. Subjects in the recall task gave evidence of this kind of memory. A number of subjects misrecalled big when huge was correct, or small when tiny was correct. Denotatively, these are inexact correspondences. Yet subjects did not misrecall huge as little or small, values on the other side of the dimensional norm

for the particular noun in question. Thus, if the subject remembered anything, he would recognize the correct adjective when the polar huge was the correct choice. When he could not remember, he would choose randomly among the alternatives.

A subject faced with a forced-choice recognition test when a non-polar adjective was the correct choice would have a different stiuation. If he remembered the dimension of predication and an imperfect representation of the original denotation, he might in fact find more than one alternative which was consistent with the information he had. Suppose the correct choice were the adjective bouncing, as in bouncing ball, and that one of the distractors was rolling. These adjectives clearly do not mean the same thing, and given a complete representation of their respective denotations would not appear to be very confusable. But if the subject faced this choice with a degraded representation the differences between the two might diminish. The original adjective, bouncing, predicates something about the current activities of the noun ball. The sentence which was actually used for these adjectives in Experiments 2 and 3 is given in [4.3]. Given this sentential context it would be

[4.3] A small boy kicked the --- wooden ball.

consistent to encode both rolling and bouncing as involving a net horizontal movement of the ball (for instance, if it were encoded in terms of action imagery), and if the distinctive aspects of a bouncing motion as opposed to a rolling motion were lost, both would be consonant with what the subject remembered about the original predication. This could lead to greater-than-chance choice of rolling as an error when bouncing was correct.

What these examples suggest is that if a subject remembers something about the original adjectival predication which retains aspects of the denotative meaning of the adjective, this information could lead to increased confusion errors to adjectives from the same predicative class when the correct choice is a nonpolar but not when it is a polar. If this interpretation of the observed differences in recognition performance (as in Tables 4.3 and 4.4) is correct, then the model presented in Chapter 2 would need to be modified. That model suggested that information about the dimension or class of an adjective and about the value on the dimension could be lost independently, allowing memory for the predicative class of an adjective without any associated memory for information about the value of the predication. This would seem to be incorrect.

Recognition data collected recently by Anisfeld (in press) support the findings of the research reported in this dissertation. He had subjects give yes-no judgments to adjective-noun phrases (e.g., back door) in a continuous recognition task. He constructed distractors using nouns which had appeared before combined with new adjectives which either preserved the meaning of the original noun phrase (e.g., rear door), reversed the meaning of the original noun phrase (e.g., front door), or gave the new noun phrase a meaning which was neutral with respect to the original noun phrase (e.g., screen door). He found that subjects made significantly more false recognition responses to synonymous noun phrases than to the neutral control phrases, while antonymous and neutral phrases did not differ. This was in marked contrast to similar

recognition tasks using single words as stimulus materials. Significantly more false recognitions have been found for both synonyms and antonyms when compared to control words in single-word studies (e.g., Anisfeld & Knapp, 1968; Fillenbaum, 1969; Anisfeld, in press). Anisfeld argued that this indicates that words and phrases are represented in different ways, the latter being encoded in terms of their referential meaning. This encoding would preclude recognition errors to phrases with the opposite meaning since they would have a very different reference. Regardless of Anisfeld's tentative interpretation of the difference between words and phrases, his results for phrases are consistent with the results reported here for adjectives embedded in complete sentences.

The recall data do not provide insight into the difference in overall error rate found between polar and nonpolar adjectives. In all of the experiments subjects consistently had more difficulty recognizing or recalling polar adjectives. This could be due to several factors, although how many are actually involved would be a matter for further research. First, polar adjectives are more abstract than most of the nonpolar adjectives used in this research (see Martin, 1969b), and abstract verbal materials are generally harder to recognize or recall than concrete materials (see Paivio, 1969). Prenominal adjective ordering has been related to concreteness or definiteness of denotation (Martin, 1969b), with more concrete adjectives preferred closer to the modified noun than less concrete ones. Vendler's (1968) analysis of adjective classes was based on these ordering relationships, and the polar adjectives in general are preferred farther from the noun than most of the

other classes used in this research. This difference in relative concreteness might explain the inferior performance with polars, although this has not been confirmed in detail.

A second possible explanation is based on the design of the experiments. Individual polar adjectives tended to be used more often than individual nonpolars, to the extent that several repetitions of particular polars occurred within individual study lists. Such repetitions could lead to more confusions, although supposedly this would be reflected by greater numbers of within-class confusions among polars than were actually found. Furthermore, color adjectives were also used extensively, including some repetitions of particular adjectives within study lists, yet these were the best remembered of all the classes of adjectives.

A third reason is the greater complexity of polar adjectives than nonpolars. As we saw in Chapter 2, polars typically are noun dependent in that the adjectival modification is not transitive to superordinate categories of which the noun is a member. Example [4.4] demonstrates this. This noun-dependency of the polars means in general that the

[4.4] the large flea ⟶ the flea which is large for being a flea
⟶ the flea which is larger than $\text{flea}_{\text{norm}}$

representation of the adjective-noun phrase has implicit normative information in it to capture the lack of transitivity. This additional information could increase the storage demands made in representing these adjectives in memory, yielding more errors because fewer complete representations are constructed than in the case of nonpolars. This

reasoning is extremely speculative, however, and there are no reliable data to support it at this point.

All of these explanations may have some validity, but the first one seems the most reasonable and has an empirical foundation that enhances its plausibility. The data to be presented in the next section are also consistent with a concreteness hypothesis. However, this interpretation is in need of further elaboration and verification before it will be an acceptable account of the persistent differences in memory accuracy found between polar and nonpolar adjectives.

Analysis of Adjectives by Classes

In order to examine the consistency of the effects that have been reported, and to look for systematic differences among the various classes of adjectives, the overall error rates and the conditional error rates for the different classes of adjectives used in the three experiments are presented in Tables 4.6 and 4.7. The classes of nonpolar adjectives were described in Chapter 2, and the category labels from that earlier discussion are presented in a footnote to Table 4.6 for cross-referencing. The miscellaneous category for nonpolars contains adjective classes which were not used in all of the experiments and which have small numbers of observations. Several rough categories were made of the different types of functions served by polar adjectives, yielding the breakdown of the polars shown in these two tables. One large group of polars referred to the spatial extent (involving either two or three dimensions) of the object being modified (e.g., adjectives

Table 4.6

Overall Error Rate by Adjective Classes

Class[a]	Exp. 1	Exp. 2 4-ch.	Exp. 2 8-ch.	Exp. 3 strict	Exp. 3 lax
Nonpolar					
Substance	.29 (279)	.12 (186)	.14 (154)	.41 (104)	.39
Shape	.28 (78)	.13 (61)	.08 (61)	.19 (42)	.19
Color	.21 (160)	.11 (93)	.21 (91)	.24 (59)	.24
Physical Status	.27 (203)	.08 (94)	.10 (62)	.34 (58)	.33
Class Relationships	.23 (459)	.07 (94)	.18 (78)	.42 (48)	.40
Ved	.25 (157)	.13 (155)	.17 (200)	.49 (110)	.45
Ving	.35 (121)	.13 (60)	.21 (125)	.55 (58)	.53
Emotives	.26 (42)	.20 (30)	.13 (77)	.47 (32)	.31
Adv.-S	.22 (139)	.16 (32)	.07 (15)	.81 (16)	.69
Misc.	.12 (59)	---	.71 (17)	---	---
Polar					
Physical Dimensions	.39 (274)	.25 (248)	.30 (298)	.61 (165)	.50
Physical Characteristics	.29 (180)	.23 (155)	.32 (93)	.58 (60)	.52
Misc.	.29 (121)	.13 (153)	.36 (154)	.45 (102)	.42

[a]Nonpolar adjective classes correspond to classes discussed in Chapter 2: Substance (I), Shape (IV), Color (II), Physical Status (VI), Class Relationships (VIII), Ved (X), Ving (IX), Emotives (XI), Adv.-S (XIII). Polar classes are explained in the text.

Table 4.7

Conditional Error Rate by Adjective Classes

Class[a]	Exp. 1	Exp. 2 4-ch.	Exp. 2 8-ch.	Exp. 3 strict	Exp. 3 lax
	(.33)[b]	(.33)	(.14)		
Nonpolar					
Substance	.42 (82)	.70 (23)	.23 (22)	.30 (43)	.27 (41)
Shape	.36 (22)	.63 (8)	0 (5)	0 (8)	0 (8)
Color	.32 (34)	.50 (10)	.05 (19)	.21 (14)	.21 (14)
Physical Status	.46 (54)	.25 (8)	.17 (6)	.30 (20)	.26 (19)
Class Relationships	.34 (107)	.29 (7)	.07 (14)	.40 (20)	.37 (19)
Ved	.30 (40)	.70 (20)	.18 (34)	.35 (54)	.29 (49)
Ving	.38 (42)	.50 (8)	.38 (26)	.19 (32)	.16 (31)
Emotives	.64 (11)	.17 (6)	.30 (10)	.47 (15)	.20 (10)
Adv.-S	.50 (30)	.20 (5)	1.00 (1)	.15 (13)	.09 (11)
Misc.	.43 (7)	---	.25 (12)	---	---
Polar					
Physical Dimensions	.36 (107)	.24 (63)	.11 (89)	.30 (101)	.14 (83)
Physical Characteristics	.37 (35)	.34 (35)	.07 (30)	.14 (35)	.03 (31)
Misc.	.30 (23)	.55 (20)	.20 (55)	.13 (46)	.07 (43)

[a]Classes are defined in Table 4.6.

[b]Chance values are given for the recognition conditions in Experiments 1 and 2.

like wide-narrow, long-short, big-little), while another group referred to other physical characteristics or attributes of the object that did not involve strictly spatial aspects (e.g., wet-dry, heavy-light, cold-warm). These two groups are shown separately in Tables 4.6 and 4.7. The miscellaneous category for polars contains adjectives which served a variety of functions, in general not directly related to the static physical characteristics of the object (e.g., fast-slow, old-young).

Breaking down the nonpolars and polars into the classes shown in these two tables reveals that the general differences found between polars and nonpolars--higher overall error rate for polars, higher conditional error rate for nonpolars--in general held for individual classes as well. There were very few instances of overlap. In general the subclasses of nonpolar adjectives had lower overall error rates than any of the subclasses of polars, while the reverse held for the conditional error rates (although there was more noise in these latter data because there were fewer observations).

If one rank orders the magnitudes of the overall error rates for the nonpolar classes from lowest to highest in each of the experiments (with the four-choice and eight-choice conditions in Experiment 2 separated, and using the lax scoring criterion in Experiment 3), some consistent differences in the error rates among these classes emerge which may offer insight into the kind of representation given to adjective-noun phrases in memory. The average rank order of error rates by classes is shown in Table 4.8. As can be seen there, subjects consistently made the fewest number of errors when the adjective referred to the color,

Table 4.8

Average Rank Order of Error Rate for Nonpolar Adjective Subclasses (From lowest to highest)

Class[a]	Average Rank Order
Color	3.6
Physical Status	3.8
Shape	4.0
Class Relationships	4.3
Adv.-S.	5.0
Emotives	5.3
Substance	5.5
Ved	5.8
Ving	7.9

[a]See Table 4.6 for description of classes.

shape, or physical status of the object (examples of this latter class were muddy, mossy, clear, colorful, the class VI in Chapter 2). On the other hand, subjects made the greatest proportion of errors when the adjective referred to the current activity of the object by means of a present participle like bouncing, rolling, or steaming.

The differences in memorability among the classes of adjectives shown in Table 4.8 could be explained in a variety of ways, but let us examine some speculations that have ties with some of the research reviewed in Chapter 1. We found there that mental imagery was an important component of the encoding of concrete verbal material. It is likely that imaginal encoding was used by subjects in these experiments to remember the meaning of the simple sentences they studied. But there may be an interesting way in which such imaginal encoding entered into the differences in Table 4.8. A concrete noun like ball can be modified by adjectives from classes at either end of the rank ordering in Table 4.8, and it is not certain (although it would have to be checked) that a noun phrase like bouncing ball is less imaginable than a phrase like red ball. This would mean that an explanation in terms of a simple measure like Paivio's I would be unsatisfactory for explaining the differences among classes of adjectives. Given the dominant effect of the concreteness of the modified noun, it is unlikely that the modifying adjectives could change the overall concreteness of the noun in any significant way. It is true, of course, that as isolated words adjectives like red or round have higher I values than adjectives like bouncing or rolling (see Martin, 1969b; also unpublished norms of Paivio), but

Anisfeld's (in press) research suggests we need to use caution in extrapolating parameters of individual words to the interpretation of coherent phrases. A more plausible difference among these classes of adjectives might be the distinctness or separability of the images constructed for similar noun phrases. For example, it might be that images evoked by adjectives referring to the actions of an object would be more confusable than images evoked by adjectives referring to the color or shape of an object. If memory differences were due to such differences in the discriminability of images, then there would also be a covarying negative relationship with the conditional probability of making an error in the same class as the correct adjective. In the case under discussion this would mean that within-class errors should be relatively more frequent with the Ving class than with Color or Shape. The data in Table 4.7 are inconclusive with regard to this point, largely because of unreliability in the relevant values due to small numbers of observations. But this suggestion deserves further investigation, since it may offer a means for assessing some of the structural characteristics of memory imagery, and may help us learn whether the language of mental imagery can be systematized in some way.

Chapter 5

SUMMARY AND CONCLUSIONS

What is remembered? That question has been the focus of this dissertation, and its rhetorical simplicity belies its conceptual complexity. The network of assumptions, facts, and inferences that must be tapped to speak reasonably in reply to it can easily produce a tangle that masks the kinds of generalizations we seek. Yet this network forms the basis for such generalizations. In Chapter 1 we reviewed a large amount of research that seemed relevant to our inquiry. The succeeding chapters outlined and tested some proposals concerning memory for prenominal adjectives. We must now return to the general questions raised in Chapter 1 and assess our current knowledge. But before doing this let us review what we have learned about the way in which prenominal adjectives are remembered.

The basic unit of analysis in the data presented in Chapter 4 was the performance error in a memory task. We looked at both the frequency and the typology of these events, and on the basis of the specific results discussed in that chapter reached several generalizations.

First, subjects encode in memory some kind of representation of the meaning of an adjectival modification which includes information about the nature of the predication made by the adjective. The information about the kinds of predication is not totally separable from the denotation of the predication, however. This was brought out most clearly in

the differences obtained when polar and nonpolar adjectives were analyzed separately in both the recognition and the recall experiments. Subjects chose the recognition test distractor from the same class or dimension as a correct polar adjective no more often than chance, while subjects consistently made significant numbers of false recognition responses to adjectives of the same class or type as the correct choice when it was a nonpolar. The theoretically defined relationship between a correct choice and a within-class distractor in these experiments meant that the distractor for a polar adjective was always its antonym (e.g., big-little) while with nonpolars the distractor was never a strict antonym (e.g., rolling-bouncing). Subjects also made a significantly greater proportion of synonymic intrusions in recall when a polar adjective was correct than when a nonpolar was correct. In fact, given that a subject recalled the nature of the predication of the correct choice, he remembered the meaning of polars and nonpolars equally well. All of this evidence, which was discussed at greater length in Chapter 4, pointed toward the remembering of denotative meaning.

Adjective-noun relationships were specified in some detail in Chapter 2, based on an analysis derived from syntactic theory, and a metric of relationships among different classes of adjectives was sketched. These proposed relationships among classes were not found in the data, and possible reasons for this were discussed in Chapter 4. This failure to support the prediction of second-order confusion errors, along with the implication of denotative meaning in remembering of adjective-noun phrases, shows that the representation presented in Chapter 2 is incorrect. The partitioning of adjectives into classes in Chapter 2 is

psychologically important, but apparently only to the extent that the classes separate out predicative differences that are perhaps more appropriately called meaning differences.

Finally, there were consistent differences in the accuracy with which subjects remembered different classes of adjectives. Subjects had greater difficulty remembering polar adjectives than nonpolars (even with the correction for recollection of meaning in the recall experiment). This could be due to the greater dependency of polars upon the modifying noun which results in greater complexity of representation, or to the more abstract nature of most polars when contrasted with nonpolars. There are many important linguistic differences between polar and nonpolar adjectives (see, e.g., Bierwisch, 1967; Clark, 1969). But what the relationship between such linguistic differences and the memory performance differences discussed in Chapter 4 might be is still unclear. Among the classes of nonpolars studied, subjects consistently remembered colors, shapes, and quantitatively specified physical attributes (e.g., muddy meaning having much mud, rocky meaning having many rocks) of objects best, while they had the greatest difficulty remembering the current activities of objects in the form of the present participle of a verb acting as a prenominal modifier. One possibility briefly discussed in Chapter 4 was that such differences in the remembering of different classes of adjectival predications might be due to the relative distinctness of mental images, which might provide a means of analyzing the structural details of mental images or forms of representation equivalent to images.

A discussion of the representation of sentences in memory is going to have to entail a discussion of the representation of sentence meaning. The evidence is unequivocal about that. But how are we to talk about meaning? The evidence reviewed in Chapter 1 along with the evidence reported and discussed in Chapter 4 suggests we look for a pluralistic representation of sentence meaning in memory. The variety of factors which have been used to predict memory performance with sentences are probably not unifiable within a single, elegant representation schema. Relationships among sentence constituents specified by linguistic deep structures seem to be important in the representation of sentences in memory, but linguistic deep structures have not provided a sufficient model for what is remembered. Sentences may be encoded in memory as mental images, but imagery has not been a sufficient model for sentence memory either. The same point could be made about a variety of other predictive factors.

Bregman and Stasberg (1968) concluded that memory for the syntactic form of sentences is based on an amalgam of assorted and not necessarily related information in memory, and felt that subjects typically resort to reconstructive processes using scattered menmonic information as input when queried about syntactic form. The same is probably true of meaning. When we perform on the basis of recollections of earlier verbal experiences, we reconstruct the gist of the experience by resorting to snatches of the material we remember verbatim, to images we formed of the objects and actions described in the material, to propositions we derived from separate portions of the text, to our memory of the setting within which

we experienced the verbal communication, to affective responses we may have had to the material, and to social, cultural, and intellectual norms which we have in permanent memory which we use to fill in the gaps or correct inconsistencies in what we have remembered. We may justifiably study these various kinds of mnemonic information separately, but to posit any one of them as the only representation of sentences in memory will be misleading. Furthermore, studying them in isolation may result in our missing the most important aspects of memory, the interrelations among fragments of information. Memory is undoubtedly not as coherent as we might wish.

The facts reported in Chapter 4 are interesting and perhaps even important. But in the context of the current status of our knowledge of memory for meaning they may be superfluous. The greatest need right now in trying to answer the question we have been posing throughout this dissertation is not to provide additional facts but to develop a theoretical language which can be used to talk about the facts. An attempt was made in this direction in Chapter 2, but it was at best only partially successful. The formal descriptions of linguistic deep structure have been inadequate as a theoretical language for memory research, although the development of formal systems along the lines suggested by abstract syntax, generative grammar, or case grammar might be more promising. No general language for mental imagery has been successfully proposed although some attempts to characterize specially restricted domains of mental images have been made (e.g., Baylor, 1969). If we had a way to talk about the details of meaning, a way of specifying gist,

we would be a long way toward a more adequate characterization of what is remembered from sentences and prose. We cannot hope to talk about relationships among the variety of mnemonic factors involved until we can talk about the factors themselves.

It was suggested in Chapter 1 that mental images and propositionally based deep structures might be two different modes of accessing a common stored representation of meaning or gist, or perhaps simply these were two different descriptions of such a common representation. A very important contribution to the study of gist memory would be the development of a theoretical language which would characterize this common level of representation. This language would have to be a relatively abstract semantically-based language of conceptual relations, perhaps somewhat along the lines of case grammars (e.g., Fillmore, 1968; Anderson, in preparation; Schank, Tesler & Weber, 1970), or generative semantics (e.g., Lakoff, in press). However, it is likely this theoretical language would have to be based on verbal entities larger than sentences. Linguistic research has concentrated on developing syntactic models for sentences, to the virtual exclusion of suprasentential relationships, and this emphasis has carried over into research on memory for complex verbal materials. The sentence may be the wrong unit of analysis for studying memory for meaning, but the investigation of suprasentential constituents as factors in memory is hampered by the lack of an adequate theoretical specification of such constituents and their interrelationships. A theoretical language for the representation of gist would also be complicated in that it would have to have a number of

systems relating the fundamental representations to various derived representations (in the linguistic metaphor, a common system of deep structures would be mapped onto several different systems of surface structures). In essence, what we will be after is a competence model of gist memory, in the same sense as the competence models developed for the description of natural languages (see, e.g., Chomsky, 1965, 1968; Fodor and Garrett, 1966).

"'What is truth?'said jesting Pilate; and would not stay for an answer." So wrote Francis Bacon in 1625. What is remembered? So we have been asking throughout this dissertation. Do we too display the impatience bred of flippancy and dalliance, or are we serious in our inquiry? If we are serious, we must realize the question will not have a curt answer.

References

Anderson, B. The short-term retention of active and passive sentences. Unpublished doctoral dissertation, Johns Hopkins University, 1963.

Anderson, J. Adjectives, datives and ergativisation. Foundations of Language, 1969, 5, 301-323.

Anderson, J. Towards a localistic theory of case. In preparation.

Anisfeld, M. False recognition of adjective-noun phrases. Journal of Experimental Psychology, in press.

Anisfeld, M., & Knapp, M. Association, synonymity, and directionality in false recognition. Journal of Experimental Psychology, 1968, 77, 171-179.

Bartlett, F. C. Remembering: A study in experimental and social psychology. London: Cambridge University Press, 1932.

Baylor, G. W. A treatise on the mind's eye. Working paper, Institut de Psychologie, University of Montreal, 1969.

Begg, I., & Paivio, A. Concreteness and imagery in sentence meaning. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 821-827.

Bever, T. G. The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), Cognition and the development of language. New York: Wiley, 1970. Pp. 279-362.

Bickerton, D. Prolegomena to a linguistic theory of metaphor. Foundations of Language, 1969, 5, 34-52.

Bierwisch, M. Some semantic universals of German adjectivals. Foundations of Language, 1967, 3, 1-36.

Bloomfield, L. Language. New York: Henry Holt, 1933.

Blumenthal, A. L. Prompted recall of sentences. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 203-206.

Blumenthal, A. L., & Boakes, R. Prompted recall of sentences. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 674-676.

Bolinger, D. The atomization of meaning. Language, 1965, 41, 555-573.

Bower, G. H. Organizational factors in memory. Cognitive Psychology, 1970, 1, 18-46.

Bower, G. H. Mental imagery and associative learning. In L. Gregg (Ed.), Cognition in learning and memory. New York: Wiley, in press.

Bregman, A. S. Forgetting curves with semantic, phonetic, graphic, and contiguity cues. Journal of Experimental Psychology, 1968, 78, 539-546.

Bregman, A. S., & Strasberg, R. Memory for the syntactic form of sentences. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 396-403.

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

Chomsky, N. Aspects of the theory of syntax. Cambridge, Mass.: M.I.T. Press, 1965.

Chomsky, N. Language and mind. New York: Harcourt, Brace & World, 1968.

Chomsky, N. Deep structure, surface structure, and semantic interpretation. In R. A. Jacobs & P. S. Rosenbaum (Eds.), Readings in English transformational grammar. Waltham, Mass.: Blaisdell, in press.

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

Clark, H. H., & Card, S. K. Role of semantics in remembering comparative sentences. Journal of Experimental Psychology, 1969, 82, 545-553.

Clark, H. H., & Clark, E. V. Semantic distinctions and memory for complex sentences. Quarterly Journal of Experimental Psychology, 1968, 20, 129-138.

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

Davidson, R. E. Deep structure differences and transition error patterns. Proceedings of the 76th Annual Convention of the American Psychological Association, 1968, 3, 11-12.

Dolinsky, R., & Michael, R. Post-integration in the recall of grammatical and ungrammatical word sequences. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 26-29.

Ekstrand, B. R., Wallace, W. P., & Underwood, B. J. A frequency theory of verbal-discrimination learning. Psychological Review, 1966, 73, 566-578.

Epstein, W. Recall of word lists following learning of sentences and of anomalous and random strings. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 20-25.

Fillenbaum, S. Memory for gist: Some relevant variables. Language and Speech, 1966, 9, 217-227.

Fillenbaum, S. Words as feature complexes: False recognition of antonyms and synonyms. Journal of Experimental Psychology, 1969, 82, 400-402.

Fillmore, C. J. The case for case. In E. Bach & R. T. Harms (Eds.), Universals in linguistic theory. New York: Holt, Rinehart & Winston, 1968. Pp. 1-88.

Fillmore, C. J. Types of lexical information. In F. Kiefer (Ed.), Studies in syntax and semantics. Dordrecht, Holland: D. Reidel, 1970. Pp. 109-137.

Foa, U. G., & Schlesinger, I. M. Syntactical errors in sentence recall: A reanalysis of Mehler's data. Unpublished paper, Isreal Institute of Applied Social Research, 1965.

Fodor, J., & Garrett, M. Some reflections on competence and performance. In J. Lyons & R. J. Wales (Eds.), Psycholinguistic papers. Edinburgh: Edinburgh University Press, 1966. Pp. 133-154.

Glucksberg, S., & Danks, J. Grammatical structure and recall: A function of the space in memory or of recall delay? Perception and Psychophysics, 1969, 6, 113-117.

Gomulicki, B. R. Recall as an abstractive process. Acta Psychologica, 1956, 12, 77-94.

Greenberg, J. H. Language universals. The Hague: Mouton, 1966.

Hall, J. W. Effects of IAR occurrence during learning on confidence judgments during recognition. Journal of Experimental Psychology, 1969, 79, 578-580.

Hall, J. W., Sekuler, R., & Cushman, W. Effects of IAR occurrence during learning on response time during subsequent recognition. Journal of Experimental Psychology, 1969, 79, 39-42.

Hebb, D. O. Concerning imagery. Psychological Review, 1968, 75, 466-477.

Hockett, C. R. A course in modern linguistics. New York: Macmillan, 1958.

Jespersen, O. The philosophy of grammar. London: Allen and Unwin, 1924.

Jespersen, O. A modern English grammar on historical principles. Part VI: Morphology. Copenhagen: Ejnar Munksgaard, 1942.

Johnson, R. E. Recall of prose as a function of the structural importance of the linguistic units. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 12-20.

Johnson, S. C. Hierarchical clustering schemes. Psychometrika, 1967, 32, 241-254.

Katz, J. J., & Fodor, J. A. The structure of a semantic theory. Language, 1963, 39, 170-210.

Katz, J. J., & Postal, P. M. An integrated theory of linguistic descriptions. Cambridge, Mass.: M.I.T. Press, 1964.

King, D. J. Retention of connected meaningful material as a function of modes of presentation and recall. Journal of Experimental Psychology, 1968, 77, 676-683.

Kruskal, J. B. Multidimensional scaling by optimizing of fit to a nonmetric hypothesis. Psychometrika, 1964, 29, 1-27.

Lakoff, G. On generative semantics. In D. D. Steinberg & L. A. Jakobovits (Ed.), Semantics: An interdisciplinary reader in philosophy, linguistics, anthropology and psychology. Urbana: University of Illinois Press, in press.

Lambert, W. E., & Paivio, A. The influence of noun-adjective order on learning. Canadian Journal of Psychology, 1956, 10, 9-12.

Lees, R. B. The grammar of English nominalizations. Bloomington: Indiana University Research Center in Anthropology, Folklore, and Linguistics, 1960.

Lockhart, R. S. Retrieval asymmetry in the recall of adjectives and nouns. Journal of Experimental Psychology, 1969a, 79, 12-17.

Lockhart, R. S. Retrieval asymmetry and the criterion problem in cued recall. Journal of Experimental Psychology, 1969b, 81, 192-194.

Lockhart, R. S., & Martin, J. E. Adjective order and the recall of adjective-noun triples. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 272-275.

Lyons, J. Introduction to theoretical linguistics. London: Cambridge University Press, 1968.

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

Martin, J. E. Some competence-process relationships in noun phrases with prenominal and postnominal adjectives. Journal of Verbal Learning and Verbal Behavior, 1969a, 8, 471-480.

Martin, J. E. Semantic determinants of preferred adjective order. Journal of Verbal Learning and Verbal Behavior. 1969b, 8, 697-704.

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

McCawley, J. D. The role of semantics in grammar. In E. Bach & R. T. Harms (Eds.), Universals in linguistic theory. New York: Holt, Rinehart & Winston, 1968. Pp. 124-169.

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

Miller, G. A., & Selfridge, J. A. Verbal context and the recall of meaningful material. American Journal of Psychology, 1950, 63, 176-185.

Morris, C. W. Foundations of the theory of signs. Chicago: University of Chicago Press, 1938.

Oldfield, R. C. Memory mechanisms and the theory of schemata. British Journal of Psychology, 1954, 45, 14-23.

Oller, J. W., Jr., & Sales, B. D. Conceptual restrictions on English: A psycholinguistic study. Lingua, 1969, 23, 209-232.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. The measurement of meaning. Urbana: University of Illinois Press, 1957.

Paivio, A. Learning of adjective-noun paired-associates as a function of adjective-noun word order and noun abstractness. Canadian Journal of Psychology, 1963, 17, 370-379.

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

Paivio, A. Imagery and natural language. Symposium address, American Educational Research Association meeting, Minneapolis, March 1970.

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. Journal of Experimental Psychology Monograph Supplement, 1968, 76, No. 1, Part 2.

Paul, I. H. Studies in remembering. Psychological Issues, 1959, 1, Whole No. 2.

Polzella, D. J., & Rohrman, N. L. Psychological aspects of transitive verbs. Paper presented at the meeting of the Eastern Psychological Association, Philadelphia, April 1969.

Pompi, K. F., & Lachman, R. Surrogate processes in the short-term retention of connected discourse. Journal of Experimental Psychology, 1967, 75, 143-150.

Postal, P. M. On the surface verb "remind." Linguistic Inquiry, 1970, 1, 37-120.

Quillian, M. R. Semantic memory. In M. L. Minsky (Ed.), Semantic information processing. Cambridge, Mass.: M.I.T. Press, 1968. Pp. 216-270.

Richardson, A. Mental imagery. New York: Springer, 1969.

Rohrman, N. L. The role of syntactic structure in the recall of English nominalizations. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 904-912.

Rohrman, N. L. More on "the recall of nominalizations." Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1969.

Rohrman, N. L., & Polzella, D. J. Recall of subject nominalizations. Psychonomic Science, 1968, 12, 373-374.

Rosenberg, S. Recall of sentences as a function of syntactic and associative habit. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 392-396.

Rosenberg, S. Association and phrase structure in sentence recall. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 1077-1081.

Rosenberg, S. The recall of verbal material accompanying semantically well-integrated and semantically poorly-integrated sentences. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 732-736.

Ross, J. R. Adjectives as noun phrases. Paper presented at the Winter Meeting of the Linguistic Society of America, New York, 1966. (Reprinted in D.A. Reibel & S.A. Schane (Eds.), Modern studies in English. Englewood Cliffs, N. J.: Prentice-Hall, 1969. Pp. 352-360.)

Sachs, J. S. Recognition memory for syntactic and semantic aspects of connected discourse. Perception and Psychophysics, 1967, 2, 437-442.

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

Schank, R. C., Tesler, L., & Weber, S. Spinoza II: Conceptual case-based natural language analysis. Artificial Intelligence Project Memo AIM-109, Computer Science Department, Stanford University, 1970.

Shepard, R. N. Recognition memory for words, sentences, and pictures. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 156-163.

Shepard, R. N., & Chipman, S. Second-order isomorphism of internal representations: Shapes of states. Cognitive Psychology, 1970, 1, 1-17.

Siegal, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

Smith, C. S. A class of complex modifiers in English. Language, 1961, 37, 342-365.

Smith, C. S. Determiners and relative clauses in a generative grammar of English. Language, 1964, 40, 37-52.

Turner, E. A., & Rommetveit, R. Focus of attention in recall of active and passive sentences. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 543-548.

Underwood, B. J. False recognition produced by implicit verbal responses. Journal of Experimental Psychology, 1965, 70, 122-129.

Underwood, B. J., & Freund, J. S. Errors in recognition learning and retention. Journal of Experimental Psychology, 1968, 78, 55-63.

Vendler, Z. The grammar of goodness. Philosophical Review, 1963, 72, 446-465.

Vendler, Z. Adjectives and nominalizations. The Hague: Mouton, 1968.

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

Weinreich, U. Explorations in semantic theory. In T. A. Sebeok (Ed.), Current trends in linguistics. Volume III: Theoretical foundations. The Hague: Mouton, 1966. Pp. 395-477.

Weinreich, U. Problems in the analysis of idioms. In J. Puhvel (Ed.), Substance and structure of language. Berkeley: University of California Press, 1969. Pp. 23-81.

Weisberg, R. W. Sentence processing assessed through intrasentence word associations. Journal of Experimental Psychology, 1969, 82, 332-338.

Welborn, E. L., & English, H. Logical learning and retention: A general review of experiments with meaningful materials. Psychological Bulletin, 1937, 34, 1-20.

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

Yuille, J. C., & Paivio, A. Abstractness and the recall of connected discourse. Journal of Experimental Psychology, 1969, 82, 467-471.

Yuille, J. C., Paivio, A., & Lambert, W. E. Noun and adjective imagery and order in paired-associate learning by French and English subjects. Canadian Journal of Psychology, 1969, 23, 459-466.