DE GRUYTER

I. M. Schlesinger

SENTENCE STRUCTURE AND THE READING PROCESS

JANUA LINGUARUM. SERIES MINOR

SENTENCE STRUCTURE

AND THE READING PROCESS

JANUA LINGUARUM

STUDIA MEMORIAE NICOLAI VAN WIJK DEDICATA

edenda curat

C. H. VAN SCHOONEVELD

INDIANA UNIVERSITY

SERIES MINOR

69

1968

MOUTON

THE HAGUE · PARIS

SENTENCE STRUCTURE AND THE READING PROCESS

by

I. M. SCHLESINGER

THE HEBREW UNIVERSITY, JERUSALEM

AND

THE ISRAEL INSTITUTE OF APPLIED SOCIAL RESEARCH

1968

MOUTON

THE HAGUE · PARIS

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 68-23200

Printed in The Netherlands by Mouton & Co., Printers, The Hague.

To my wife

PREFACE

This study is a revision of a doctoral dissertation submitted to the Senate of the Hebrew University, Jerusalem, in 1964. Most sections have been rewritten and reviews of the literature (especially that in Chapter Three) have been brought up to date till about the middle of 1965. Some sections of lesser interest have been omitted, and descriptions of experimental procedures have been shortened (more detailed descriptions being available in mimeographed form). Every effort has been made to present these descriptions in non-technical language, with an eye on readers with no background of psychological training.

I am very much indebted to my thesis advisers, Professors S. Kugelmass and Y. Bar-Hillel of the Hebrew University, Jerusalem, for all the help they have given me. But for their unfailing patience and forbearing and their continuous encouragement this work would not have reached completion.

My interest in psycholinguistic problems dates back from Professor Bar-Hillel's lectures at the Hebrew University in 1954. Since beginning work on this thesis, in 1960, I have had the benefit of long hours of talks with Professor Bar-Hillel. Besides having been an invaluable experience, these talks have helped me in clarifying many of the problems with which my work has been beset. Thus Professor Bar-Hillel's thinking has influenced, directly or indirectly, almost every part of this study. Finally, it is due to Professor Bar-Hillel's initiative that this work has been published. For all this, I wish to express my gratitude to him.

While this study owes much to Professor Bar-Hillel, the entire responsibility for its shortcomings – of which I am well aware – lies with myself. In part, these shortcomings may be due to the fact that the experiments reported here are among the first to be conducted in this area; I was, therefore, deprived of the opportunity to benefit from the experience of others. In fact, at the time this work was begun, in 1960, not a single psychological study had been published, to my knowledge, in which the influence of a syntactic variable was investigated directly; I was not then aware of the important research which was being conducted elsewhere at about the same time as my own, in the same area and often along similar lines. To the extent that relevant studies have, in the meantime, become available to me, they have been reported here.

On the other hand, I have had the benefit of discussions with persons working in various areas. Of the staff of the Hebrew University I would like to mention the late Professor Irene Garbell, and Professors Haim Blanc, Moshe Goshen-Gottstein, Louis Guttman, Daniel Kahaneman, Chaim Rabin, and E. Shamir, to whom I am indebted for stimulating discussions and for their readiness to help me in many ways. The experiments reported in Chapter Six were stimulated by comments of Professor Noam Chomsky on a short report I had written. By providing me with information about their unpublished work, the researchers mentioned in Chapter Three have made it possible to bring the discussion up-to-date. I want to take this opportunity to thank all these persons for their kindness.

The conscientious work of students of the Psychology Department of the Hebrew University who assisted me in this study is also gratefully acknowledged. Special mention deserve Mr. Benyamin Beth-Halachmi, Mr. Asher Coriat, Mrs. Dalia Etzion, and Mr. David Seidel who carried out several of the experiments and helped with the analysis of results, and Mrs. Rachel Melkman whose able help with the experiments reported in sections 2.2.5 and 5.4.1 was more in the nature of collaboration than mere technical assistance.

Work reported in this study was supported in part by research grants from the following sources: the U.S. Office of Naval Research, Information Systems Branch, Contract No. 62558-4695, NR-049-130; the Rena and Walter Burke Foundation (through the Hebrew University, Jerusalem); and the General Federation of Labour is Israel.

Jerusalem, Israel
July, 1966

I. M. SCHLESINGER

CONTENTS

1

INTRODUCTION

The question how syntactic structure affects the ease of reading is obviously of the greatest importance to teachers, writers, editors, in short – to anyone interested in written communication. The present study is an attempt to deal with this problem experimentally. It is one of the first attempts of its kind, for, in spite of the obvious importance of the problem, empirical research on it is almost non-existent, as we shall see presently.

1.1. PREVIOUS PSYCHOLOGICAL RESEARCH ON THE INFLUENCE OF SENTENCE STRUCTURE

The problem of the influence of sentence structure falls in the province of so-called *readability* research. Work done in this field has been lacking in theoretical orientation; the aim of readability studies has been an applied one – to develop a yardstick by means of which the reading ease[1] of a given text can be conveniently measured. On the basis of correlational studies, formulas were devised which answered this practical need (see Chall's (1958) useful review of these studies). Syntactic structure is, of course, generally admitted to be one of the determinants of readability; however, the only variable to appear in the formulas, which has anything to do with sentence structure, is sentence length. Now, sentence length may perhaps be only a symptom of reading ease,

[1] The term "readability" is often taken to refer also to the amount of interest a given text arouses in the reader (cf. Chall, 1958, pp. 4-7; Klare *et al.*, 1955; Schramm, 1947). But here and in the following the term will not be used in this way, since we will be concerned only with reading ease and difficulty.

reflecting some underlying factor connected with the *complexity* of sentence structure. This is recognized by researchers in this field who give the practical advice to write "simple sentences", so as to ensure comprehension and ease of reading. But what simplicity or, by contrast, "complexity" consist of is not made explicit: here one trusts the judgment of the writer.

This neglect of research pertaining to the relationships obtaining between syntactic variables and readability is paralleled by a general neglect of these variables in psycholinguistic research. Most psychological studies on linguistic variables are concerned with words; psychologists seem to hesitate to come to grips with larger and more complex units. This is perhaps understandable in view of the great difficulties attendant on research with such units (difficulties which will become evident in the following chapters); but as long as syntactic variables are ignored, no understanding of language behavior can be achieved.

The theoretical writings of psychologists bear ample evidence of the fact that the psychological importance of sentence structure is recognized (e.g., Osgood and Sebeok, 1954; Osgood, 1957; Miller *et al.*, 1960). In their research, however, psychologists have usually contented themselves with approaching the problem of sentence structure, if at all, indirectly, for instance via word class (e.g., Aborn *et al.*, 1959). Studies in which sentence structure figures indirectly are those which attempt to relate recall of words to transitional probabilities of the text (Postman and Adams, 1960; Tulving and Patkau, 1962), and intelligibility of words to their position in the sentence (Rubenstein and Pickett, 1958). Mention should also be made of studies concerned with the "psychological reality" of the parts of speech (Barik and Fillenbaum, 1961; Barik and Lambert, 1960; Glanzer, 1962).

A new impetus to psychological work on sentence structure has been given by recent developments in theoretical linguistics. The most prominent single influence here has been the work of Noam Chomsky. Some years after the appearance of his "*Syntactic Structures*" (1957) studies began to be published which dealt with the behavioral consequences of syntactic structure in a more direct manner, investigating its influence on intelligibility (Miller and Isard, 1963; Marks and Miller, 1964), and on learning and recall (Epstein, 1961, 1962, 1963). In line with Chomsky's (1962) remarks regarding the relationship between grammatical models and language learning, a new approach is being taken to the study of language development (e.g., Menyuk 1963a, 1963b; see also Ervin and Miller's review, 1963). Other studies prompted by Chomsky's work will be mentioned in the following chapters.

1.2. THE SCOPE OF THIS STUDY AND ITS THEORETICAL ORIENTATION

Like some of the psycholinguistic work mentioned in the previous section, this study leans heavily on grammatical models developed in modern theoretical linguistics. On the basis of these models, psychological hypotheses will be formulated and an attempt will be made to test these empirically. Specifically, our investigations will deal with:

(a) the psychological reality of the syntactical constituent (chapter 2),

(b) the effect of grammatical transformations on decoding and encoding behavior (chapter 3), and

(c) the effect of sentence complexity as defined in the work of Yngve (1960) and Chomsky (1957) on the ease of reading (chapters 5 and 6).

In addition to these, experiments are reported on other syntactic variables – sentence length (chapter 4) and the location of the subject (chapter 7).

There is perhaps no need to point out that the linguistic theories which serve as the starting point of our investigations do not make any psychological statements, and can therefore not be put to test by psychological experiments. This study is based on the assumption that certain aspects of decoding and encoding behavior can be explained in the light of linguistic theories. Our basic tenet is that *the human user of language incorporates a device which operates* *along the lines of a grammar proposed by theoretical linguists,* such as a Chomskyan grammar. This proposition is not an integral part of any linguistic theory.

The nature of this statement is such that no empirical study can ever refute it; it is not stated in operational terms. It may be looked upon as a *metahypothesis* which gives rise to certain specific *research hypotheses.* One of our research hypotheses, for example, states that self-embedded sentences will be more difficult to read than those which are not self-embedded (5.2.2), and this hypothesis is based on Chomsky's discussion of self-embedding (5.2.1). If this particular research hypothesis fails to be confirmed by an empirical test, the above metahypothesis will in no way be disparaged thereby (and, of course, Chomsky's theory will not be disparaged thereby, as has been pointed out above). This is so, because we can easily formulate a new research hypothesis, which may even be incompatible with the former one, and which is based on the *same* metahypothesis. For instance, one might hypothesize that self-embedding affects not reading ease as measured by reading rate, but as measured by some other criterion, or that it does not affect the reading process at all unless carried to a certain degree, or that it influences only the reading of sentences of a certain kind, and so on. None of these hypotheses can be strictly derived from the metahypothesis, but all are somehow based on it. The value of the metahypothesis lies in that it guides us in the formulation of empirically verifiable research hypotheses; it can not be empirically confirmed but only shown to be fruitful in this respect. The present study, then, is an exploration of the fruitfulness of the above metahypothetical statement.

Although no linguistic controversies can be settled by it, this study should prove of interest to the linguist. Ultimately, a comprehensive theory of language must concern itself with the linguistic processes of the user of language. The question of whether or not a description of the latter on the basis of a given linguistic model is possible, concerns, therefore, linguistics no less than psychology.

1.3. A PROBLEM IN THE INTERPRETATION OF RESEARCH RESULTS

Suppose a specific research hypothesis, which is based on our metahypothesis (1.2) states that sentences of structure S_i will be harder to read than sentences of a different structure, S_j. It may usually be assumed that sentence structures which are more difficult will come to be employed less frequently by speakers of the language. Frequency of occurrence, in its turn, may be a factor in the difficulty of a sentence, when difficulty is measured for instance, by reading rate.[2] Now, if the research hypothesis is confirmed by an experiment comparing the reading ease of S_i and S_j, this cannot

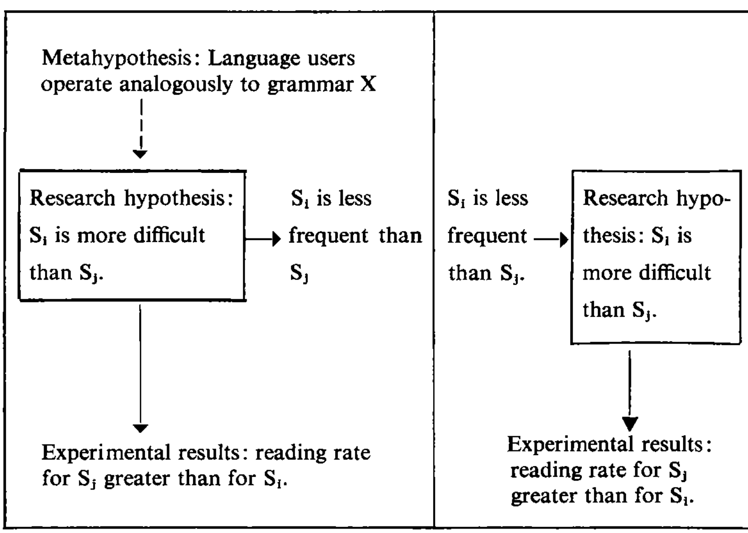

Fig. 1.1. Two alternative explanations for the same experimental results obtained for S_i and S_j ($i \neq j$). Hypotheses appear in frames. The broken line indicates that the research hypothesis is not strictly derived from the metahypothesis (cf. 1.2.).

[2] The recognition threshold is lower for frequent than for infrequent words, whether the words are presented visually (Howes and Solomon, 1951) or auditorily (Howes, 1957). Bloomer (1961) found the difficulty experienced by schoolchildren in reading words to decrease with word frequency. Similar experiments with sentence structures differing in frequency have not yet been conducted, to my knowledge.

be taken as a point in favor of the metahypothesis, because there is an alternative explanation available: sentences of structure S_i may be harder to read than sentences of structure S_j, because the former are less frequent. These two alternative explanations are illustrated in figure 1.1.

To rule out this alternative explanation, one would have to conduct experiments in which the variable of frequency is controlled. Frequency counts of sentence structures are non-existent, and even if the technical problem of the amount of labor required for such a count were solved, a useful frequency count of sentence structures will have to wait for an adequate grammar giving an adequate structural description of any given sentence. Another possibility might be to obtain estimates of frequency by utilizing the knowledge of language statistics presumably built-in in native speakers. In psychological research, such a procedure has been employed for the estimation of transition probabilities between letters and between words. In one of our experiments reported in chapter 3, use was also made of judgments of the relative frequency of sentences. Here a limitation of the use of such judgments needs to be mentioned:

(a) Judges may be influenced by existing stereotypes as to which sentence structures are "too complex", and therefore "wrong", and such structures may tend to be judged by them as less frequent than they really are.[3]

(b) The readability of the sentence may influence the judgment of its frequency. As a consequence, the correlation between judged frequency and experimentally measured readability will be higher than that which actually exists between frequency and readability. Therefore, judgments should not be relied upon too much.

Neither can the above problem of alternative explanations be solved by experiments with artificial languages. From the subject's point of view, the prior frequency of occurrence of sentences appearing in such "languagettes" is, of course, zero. But if the "languagette" follows syntactic rules of natural languages known

[3] Cf. McCracken's (1959) finding regarding judgments of grade school children.

to him, he will inevitably draw analogies with the natural languages and the frequency of sentence structures will again become operative in his performance. On the other hand, if the "languagette" departs radically from usual sentence structure, the experiments cannot tell us anything about the influence on performance of those specific syntactic variables we are interested in.

In the absence of a totally satisfactory way of controlling for frequency of occurrence of sentence types, we must resign ourselves to the fact that our results are amenable to an alternative explanation. Such a state of affairs is, of course, not uncommon in science. The decision in favor of one of the explanations will then be made (if at all) on the basis of such considerations as the generality of the explanation or parsimony. The following consideration should be mentioned in favor of our metahypothesis: Both the experimental results and the relative frequency of occurrence of the sentence structure in question can be attributed to the metahypothesis, whereas the alternative explanation leaves frequency of occurrence unexplained. (The "explanation" that frequency results from difficulty is, of course, circular according to the alternative explanation).

1.4. RESEARCH METHODOLOGY OF THIS STUDY

Previous readability studies have been mainly correlational studies resulting in formulas with which the readability of a given text could be measured (1.1). Readability studies had their heyday in the United States in the forties and the early fifties. This period witnessed a proliferation of formulas and of research papers assessing their validity and reliability, and employing the formulas to measure the readability of educational and other material (cf. Chall, 1958). In the last few years, interest in this area has petered out, and only few readability studies have been published recently. The reason for this appears to us to lie in the fact that the usefulness of the research methodology employed has been exhausted. The formulas have come up with various symptoms of reading difficulty, such as "average sentence length in words", "number of prepositional phrases" and "words per paragraph", but (with very few exceptions which will be mentioned later in this study) no research has been done to find out how these are related to underlying factors influencing the reading process.[4] Even for the limited applied objectives which the researchers had in mind, the results of the correlational approach are far from satisfactory. The formulas may be suitable enough for estimating the level of readability of existing texts – they constitute what has been aptly called (Hebb and Bindra, 1952) "a literary Wasserman test" – but it is doubtful whether they can successfully estimate the readability of texts written according to the formula (Chall, 1958, 97ff.; McCracken, 1959). This is so, because one can very well "beat the formula", intentionally or not, and remove the *symptom* of reading difficulty without removing the underlying *cause*. For instance, when sentences of a given text are shortened, the formulas, which include sentence length as a measure, will decree that the text is now more readable, but it remains to be seen whether this is actually so.[5] This question will be taken up again in Chapter 4.

To find out how linguistic factors affect the reading process, an altogether different research methodology is required. Rather than examine existing texts for variables which correlate with readability, linguistic factors must first be isolated, so that they can be manipulated experimentally. The effect of the linguistic factor

[4] Two studies which attempted to go beyond this by applying a factor analysis (Brinton and Danielson, 1958; Stolurow and Newman, 1959) have come up with the not very illuminating result that both a "word" factor and a "sentence" factor contributed to most of the variance in the difficulty of the text.

[5] A paper by Taylor (1953) is of interest in this connection, in which it is pointed out that a text from Erskine Caldwell is graded as difficult by a formula, whereas a text from Gertrude Stein is graded as very easy! Caldwell (but not Stein) uses long sentences; these, however, do not present any difficulty for the ordinary reader, presumably because they are not involved or complex in some sense which is not accounted for by the formulas. Taylor's own technique of estimating readability by the "cloze" procedure – which is based on guessing words deleted from a text – may have certain advantages over the formulas (see Taylor, 1953, 1956, 1957; and Osgood and Sebeok, 1954, p. 112; but see Chapanis, 1954 for a reservation); but it is open to question whether this technique will do much to further our knowledge of the underlying factors of readability. This is discussed further in 8.1.5. below.

on the reader's performance can thus be assessed. The various experimental techniques which were employed for this purpose are described in the following chapters.

At the same time the present study was begun, not a single experimental readability study had been published which treated of sentence structure. We were thus deprived of the opportunity to benefit from the experience of others, and were compelled to find out for ourselves, often gropingly, what experimental techniques were most suitable for the problems at hand. Often, existing techniques had to be adapted specifically for our purpose. Therefore, it was found advisable to conduct a number of small-scale experiments, each of which was usually based on the previous one. In a more developed field of research, one might wish, and expect, to read about comprehensive multi-variable studies, designed to answer many questions at a time. Such a research design was not thought to be strategic in this field where no previous research experience can be drawn upon and where, moreover, there exists neither a well-developed theory nor a substantial body of data upon which such a theory can be based.

2

THE SYNTACTIC CONSTITUENT AS UNIT OF DECODING

The present chapter deals with the question of the unit of decoding. It may well be incorrect to ask about *the* unit of decoding, since syllables, words, phrases as well as whole sentences may possibly be units at some level (cf. Carroll, 1953, on the hierarchy of units). Here it will be attempted to show, therefore, that at *some* level of decoding, the syntactic constituent figures as unit. This seems to follow from the metahypothesis which says that the human sentence decoding mechanism is built in a manner analogous to a grammar such as, e.g., that of Chomsky (1.2).

2.1. PREVIOUS RESEARCH

Several studies seem to indicate that the unit of decoding is larger than a single word. As early as 1899, an experiment was carried out on the decoding of telegraphic messages, and it was found that practiced operators lag, in their translation of incoming signals, from six to twelve words behind the incoming word, carrying the received signals in their memory (Bryan and Harter, 1899). This rather remarkable feat (ten English words have been computed to contain on the average 237.7 Morse signals) seems to indicate that these operators tend to decode by units of several words.[1] It is not clear from these findings, however, what is the nature of these units and what is their relationship to syntactic constituents.[2]

[1] The process possibly involved here is described by Broadbent, 1958, p. 44. The reorganization of received material into larger units has been shown to occur for letters (Allan, 1961) and for artificial finite state languages (Shipstone, 1960).

[2] Osgood (1954), McClay and Osgood (1959), and Glanzer (1962) have speculated about the nature of the decoding unit on the basis of empirical material.

In later experiments, by Ladefoged and Broadbent (1960) in which listeners judged where in the speech sequence a click had occurred, it was concluded that the listener does not deal with single sounds separately, but with groups of sounds. But no conclusive evidence was obtained regarding the size of the unit.

Miller (1960) has suggested that words of a sentence are decoded by listeners in "chunks" of 2.5 – 3 words. This is based on the finding that subjects can repeat from memory about 15 words from a text but only 5-6 isolated words.

In another experiment, Miller (1962) has compared the intelligibility of sentences (1) with that of pseudo-sentences (2) obtained by reversing the word order of sentences, thus:

(1) Don has no wet things
(2) things wet no has Don.

When there was a pause between the presentation of two sentences, or pseudo-sentences, no differences in intelligibility were obtained; however, when no such pause was made, pseudo-sentences were less intelligible than sentences, and were about as intelligible as any random string of words. Miller explains these findings as follows: Intelligibility decreases with number of alternatives and the number of alternatives is restricted by the context the word appears in (cf. Miller, *et al.*, 1951; Bruce, 1958). In the case of pseudo-sentences, the listener can make use of the context only when he is given sufficient time, i.e. when a pause occurs; when no pause occurs, between pseudo-sentences, subjects have no time to hear each word separately, understand it, and anticipate the next word. But when the words appear in sentences, in their familiar order, subjects can do so even when time is short, presumably by organizing the words into larger units. Miller believes that these units may be the equivalent of syntactical constituents, but there is no direct evidence on this point in his findings.

The tendency to reorganize verbal material in units larger than the word has also been found in experiments by Tulving and Patkau (1962). Their report does not indicate how these units are related to syntactic constituents.

An experiment by Huttenlocher (1964) seems to be relevant to the problem of the nature of the unit of decoding. He found that children were able to reverse the word order of two-word utterances – such as "black-white", "child-lady" – but had difficulty in doing so when these utterances formed common English sequences – such as "man-runs", "red-apple". This difficulty might be attributed to the necessity of breaking down these two-word units when reversing the word-order. While it may be unwarranted to generalize from the two-word units of this experiment to syntactic constituents of sentences, these results are, at least, suggestive.

In an unpublished study by Neal Johnson (quoted by Osgood, 1963), errors in recall were found to be more frequent at the transition between constituents. In the following two sentences, errors were more frequent at the places indicated by an asterisk:

> The tall boy * saved the dying woman.
> The house across the street * is burning.

Also, after "house" more errors were made than after any other word in the second sentence, except "street". These results again, suggest that in decoding the words of the sentence are organized into units corresponding to syntactic constituents. However, these results might be accounted for, at least in part, to the transitional probabilities, which may be assumed to be lower at the transition between constituents:[3] Recall has been shown to be better when transitional probabilities are higher (Miller and Selfridge, 1950; Sharp, 1958).

This alternative explanation apparently does not apply to an experiment on the recall of sentences conducted by Roger Wales.[4] Wales presented to his subjects sentences in three parts, e.g.:

[3] Some evidence for this can be found in the report of Mandler and Mandler (1964).

[4] I am very much indebted to Dr. Roger Wales, of the University of Edinburgh, for information about this as yet unpublished study which he carried out in 1964.

(1) The very old man
was always sitting down
on one of the big chairs.

(2) The very old
man was always sitting
down on one of the big chairs.

The "cuts" between the parts occurred either at the end of a syntactic constituent (1), or elsewhere in the sentence (2). Sentences were significantly easier to learn when presented as in (1) than when presented as in (2). Apparently, it is easier to store in memory units in which a sentence is normally decoded, i.e., the syntactic constituents which form the parts presented in (1). This interpretation seems to be confirmed by an earlier independent experiment on the influence of syntactic structure on the eye-voice span, which is to be described in the next section.

2.2. SYNTACTIC CONSTITUENTS AND THE EYE-VOICE SPAN

The adult reader does not fixate every single word he reads; he can usually take in several words at a glance, and reads a line of print with a small number of fixations, often less than four. In reading aloud, he pronounces a given word, while his eyes move several words ahead. This difference, which is called the *eye-voice span*, can be measured by photographing the reader's eye movements while reading aloud, or, more simply, by extinguishing the light he is reading by, and asking him to continue saying the words he saw while the light was still on.

The EVS (this abbreviation will be used in the following for "eye-voice span") serves to prepare the reader for the text lying ahead. The knowledge which words are to come presumably helps him to understand what he is reading aloud, and to avoid errors in intonation. The EVS is, thus, related to the unit of decoding.

2.2.1. *Previous Findings On the Eye-Voice Span*

Some forty years ago, Buswell (1920) conducted experiments on the EVS, and concluded that it is longer for good than for poor readers, that it becomes longer with age, and that it tends to be longest at the beginning of the sentence, grow shorter at the middle, while being shortest at the end of the sentence. Fairbanks (1937) also found that the EVS was shortest at the end of the sentence but, unlike Buswell, did not find any difference between its length at the beginning and at the middle of the sentence. According to the latter study, the length of the EVS is also dependent on the position within the printed line, being shortest at the middle of the line. Fairbanks also found that the EVS tends to become shorter where difficult words are encountered.

In 1951, it was suggested by Miller (1951, p. 191), that it would be worthwhile to find out whether the EVS was influenced by the amount of information (in the technical sense of the term) in the text. Ten years later, an experiment on this problem was reported by Lawson (1961), who found the length of the EVS to increase with the transitional probabilities in the text up to about the eighth order approximation to English.[5] The latter finding has been replicated recently by Morton (1964). However, a study conducted by myself with Hebrew texts failed to show this effect: Using texts representing second, fourth and sixth orders approximations to Hebrew, no significant differences were evident in the EVS of eighteen Hebrew University students (the mean lengths of the EVS being 3.15, 2.96 and 3.20 words, respectively).[6]

[5] Texts of the *n*th order approximations are prepared by a method – first used by Miller and Selfridge (1950) – which consists in presenting a subject with *n*-l words from a text and letting him supply the *n*th word by guessing; the next subject is then presented with the resulting text *except* for the first word (that is, again with *n*-l words), and guesses the succeeding word, and so on.

[6] The difference between the results of these experiments may possibly be due to the language they were conducted in: Hebrew, contains more morphemes per word than English, and in the above experiments the basis for constructing the texts with differing orders of approximation, as well as for measuring the EVS, was the word. Moreover, Hebrew texts in which vowel signs are often omitted, as they were in our experiment, have shorter words, on the average, than English texts, and this may be a factor in the length of the EVS. Results from experiments on the EVS obtained with English texts may, therefore, not be comparable with those obtained with Hebrew texts.

2.2.2. *The Hypothesis*

When a text is read out aloud, the reader's eyes move ahead of the words he is pronouncing so as to prepare his reading. The span his eyes are ahead of his voice represents a unit of decoding. It has been pointed out above that, according to our metahypothesis concerning the analogy between grammar and the human user of language, the unit of decoding should be predicted to be, at *some* level, the syntactic constituent. Yet we shall not simply hypothesize that the EVS will extend to the boundary of the syntactic constituent. Account must be taken of the fact that the reader's eyes, in moving from left to right along the line of print, may encounter a word which, *being ignorant of the subsequent words*, he might take to be the last word in a syntactic constituent. For instance, in the sentence.

> The woman teacher, who had taught him Latin, was very pleased.

the words "teacher", "Latin", and "pleased" are boundaries of syntactic constituents. (The fact that there are others, according to the level of analysis, need not concern us here). Now, a reader who takes in at a glance the first two words, "The woman", may well believe that these are a constituent of the sentence he is going to read; the actual structure of the sentence becomes clear to him only by continuing reading. Likewise, when he has perused the sentence as far as the word "him", he may, in the absence of any knowledge of the following word, construe "him" as being the last word of a constituent.

In the following, such a "would-be constituent" will be called a *chain*, and a real syntactical constituent will also be called a chain. A more formal definition of this term, as contrasted with the term "constituent" might be as follows (using Chomsky's (1961) terminology):

> A section of a terminal string of sentence S_i is a *constituent*, if it is dominated by any symbol in the P-marker of S_i, and is a *chain* if (a) it is dominated by any symbol in the P-marker of *any* sentence S_j, and (b) the part of the terminal string

to the left of it (i.e., preceding it) is made up of consecutive constituents of S_j.

$$\begin{Bmatrix} i = j, \\ i \neq j \end{Bmatrix}$$

It will be seen that a chain which is not a constituent comprises the first words of a constituent.

The hypothesis to be tested is, then, as follows: *The last word in the EVS tends to be the last word in a chain.*

2.2.3. *First Experiment: Method*

To measure the EVS, ten subjects were each asked to read aloud passages from sociological and educational articles of a non-technical nature in Hebrew. The experimenter switched off the light at predetermined places in the text, between two words, and asked the subject which words he had seen, when the light was still on, in addition to those he had read aloud.

The experimental sentences were of four types. To describe these, it will be convenient to introduce a minimum of notation. Let *T* stand for the place (between two words) where the light was extinguished, and *V* for the last word in a chain. The last word of the EVS will be designated by *U*. The hypothesis to be tested is, then, that *U* tends to be a *V*. The experimental testing of this hypothesis must obviously take into account the distance between *T* and *V*, and the four types of experimental sentences were defined according to this distance. These four types are summarized in Table 2.1. In half of the sentences (that is, 20 out of 40) the end of the chain, *V*, was the second word from *T* and in the other half – the third word. There was an *additional* chain,[7] the end of which was in half of the sentences the fourth word from *T*, and in another half – the fifth word. Only such sentences were chosen in which *V* was not marked off by any punctuation mark, because only with these a fair test of the hypothesis is possible. It will be

[7] This might be the constituent of which the former chain was a proper part – cf. 2.2.2 above.

seen from Table 2.1 that exactly one-half of the words by which the hypothesis would be tested were *V*.

TABLE 2.1. *Types of Sentences in the First Experiment*

Type	No. of sentences	Word after *T*			
		Second	Third	Fourth	Fifth
A	13	*V*		*V*	
B	13		*V*		*V*
C	7	*V*			*V*
D	7		*V*	*V*	

T – place (between two words) at which light is extinguished.
V – last word in a chain.

In a previous experiment (2.2.1) the EVS for Hebrew texts was found to be in most cases 2-5 words long, and hence the first word after *T*, as well as all words after the fifth word were not taken into account in the design of this experiment. Each of these words could be either a *V* or a non-*V*. Cases where the EVS was either shorter than two words or larger than five words were not taken into account in testing the hypothesis.

2.2.4. *First Experiment: Results*

(a) *Scoring the length of the EVS*

For the purpose of the analysis, the EVS was defined as all *consecutive correct* words reported by the subject. For instance, if on any trial the subject reported the two first words after *T* correctly and omitted the third word, his EVS was taken to be two words long, even if the fourth word was reported correctly. Similarly, if the third and fourth words were reported in the wrong order, the EVS was taken to include the first two words only. If any mistake was made in reporting a word – even if the mistake was only in an inflexional affix – the immediately preceding word was regarded as *U* (i.e., the last word in the EVS). These rules were adopted so as to simplify the analysis and to avoid arbitrary decisions being made as to the length of the EVS. It will be shown below (section d) that a more "lenient" criterion, i.e., one giving credit for incorrectly reported words, would not have affected the conclusions of this experiment substantially.

(b) *Testing the hypothesis*

Table 2.2 summarizes the results of the experiment. The number of trials – summed over 10 subjects – in which the EVS was 2-5 words long is given in the body of the table. Those trials in which *U* was a *V*, as predicted, are indicated by italicized numbers. From Table 2.1, above, it can be seen that for sentences of type A, for instance, *U* should be predicted to be either the second or the fourth word after *T*, i.e., that the EVS should be two or four words long. Table 2.2 shows that on 33 trials with sentences of type A the EVS was two words long, and on 38 trials – four words long. This contrasts with only 31 trials in which the EVS was three or five words long. Owing to an error made in planning the experiment, the number of sentences of each type differed slightly from the number given in Table 2.1, which accounts for the discrepancy between the two tables.

The number of cases in which the *U* was a *V*, as predicted, and the number of cases in which the prediction was not borne out are summarized in the last two rows of Table 2.2 for each length of EVS. It will be seen that whatever the length of the EVS, there was a strong tendency for *U* to be a *V*. Since half of the words were *V*, *U* would be a *V* in half of the cases by mere chance (2.2.3). Actually, however, the last word of the EVS was a *V* in over two-thirds of the trials: 223 out of 328. (Summing up over 10 subjects, there was a total of 390 trials; in 328 of these the EVS was 2-5 words long, and in the rest, it was either longer or shorter, and could, therefore, not be taken into account. Full details concerning all the 390 trials are given in Table 2.3, below).

TABLE 2.2. *The Length of the EVS According to Sentence Types**

Type	No. of sentences	No. of words in EVS				
		2	3	4	5	
A	12	*33*	25	*38*	6	
B	14	15	*61*	29	*5*	
C	6	*27*	16	3	*8*	
D	7	8	*30*	*21*	3	
U is a *V*		*60*	*91*	*59*	*13*	Total: *223*
U is a non-*V*		23	41	32	9	Total: 105

* The number of trials (summed over 10 subjects) in which the prediction was borne out is given in italics.

For each of the ten subjects tested, *U* was a *V* more often than not. The difference, thus, is a highly reliable one (p = .001) by a sign test (Siegel, 1956), and the hypothesis can be regarded as confirmed.

(c) *Additional findings*

It may be instructive to look at the results for each length of EVS separately: In those trials in which the EVS was two words long, the prediction was borne out for all ten subjects, and the same was true when the EVS was three words long. When the EVS was four words long, there was one subject for whom, contrary to the prediction, *U* was more often than not a non-*V*.[8] Finally, when the EVS was five words long, the difference, though still in the predicted direction, was not any more statistically significant (possibly because of the small number of trials on which the EVS was as long as this). Apparently, the tendency of *U* to be a *V* is stronger the shorter the EVS.

[8] With an EVS of four words, a Wilcoxon test (Siegel, 1956) shows the difference to be significant at the .005 level. However, since this is an additional significance test on the same data, the obtained p-value should be regarded with caution.

TABLE 2.3. *Results of the First Experiment for EVS of All Lengths**

length of EVS (no. of words)	no. of trials for 10 subjects	*no. of trials in which*	
		U is a *V*	*U* is a non-*V*
0	8	8	0
1	29	15	14
2	83	60	23
3	132	91	41
4	91	59	32
5	22	13	9
6	20	20	0
7	4	3	1
8	1	0	1
Total	390	269	121

* Cases in which, owing to technical difficulties, the EVS was not measured appropriately are included under EVS of zero length.

To give a full picture of the results, Table 2.3 includes also trials in which the EVS was smaller than two, or larger than five words long, and which were, therefore, excluded from the main analysis. For these trials, the same tendency is apparent, except that in the single trial in which the EVS was eight words long, *U* was a non-*V*; but in this case, *U* was also the last word in the line, and the EVS could thus not have been longer.[9]

In order to obtain some additional information, the EVS was also measured for one sentence which belonged to none of the above types (Table 2.1). In this sentence – which was, of course, not included in the above analysis – three consecutive words appeared which were non-*V*: the third, fourth, and fifth word from

[9] This was the only case where this occurred. *T* was always located at a point far enough removed from the end of the line to obviate this.

T. In spite of the long "leap" required here, *U* was a *V* for seven out of the ten subjects.

(d) *Subjects' errors*

For the purpose of testing the hypothesis, *U* was defined above as the word preceding the first word after *T* which was not recalled or on which an error was made (see (a) above). The question may be raised whether a different method of scoring would have led to different results.

The 10 Ss who each read 39 sentences made 80 errors in all. These were of the following kinds (in descending order of frequency): substitution of words (34 errors), changes in word form (change in tense, inflectional affix; 25 errors), omission of a word (14 errors), addition of a word (4 errors), and one error of inversion of word order. (Two errors could not be classified as belonging to any of the above types). An alternative scoring procedure would be one according to which all the above errors are disregarded, that is, the last word reported by the subject is taken as *U*, regardless of the correctness of this and the preceding words. It was found that by this scoring procedure, the number of trials which would be counted as bearing out the prediction would be even greater than it was by the procedure adopted above: in 25 additional cases, *U* would have been taken to be a *V*, and in only 15 cases, in which *U* according to the actually adopted procedure was a *V*, it would have been a non-*V*. In the remaining cases, identical conclusions would have been reached by the two systems of scoring. The hypothesis would, thus, have been confirmed by both procedures.

A particular kind of error is worth mentioning here. In some trials, the subject changed the form of the last word he reported in such a way that it became a *V*. The following is an example: In Hebrew, the possessive is indicated by a special inflexional form; thus, *baqqaša* = request, *baqqašat* = request of. The second form appeared in one of the experimental sentences ("... to turn to the [social] worker with the request of ..."), but one of our subjects, who reported this word as the last word in his EVS, turned it into the first form (" ... to turn to the social worker with the request ... "). We might speculate that the tendency to perceive a complete syntactical constituent may lead to a distortion in perception which is congruent with it.

(e) *Problems in the interpretation of the results*

While the experimental findings give support to the hypothesis that the reader tends to decode the text in units which are related to the syntactic constituent,[10] there remain some problems which must remain unresolved for the time being:

(1) Is the tendency of decoding in "chains" a function of a perceptual process, or does it reflect the way the material is stored in memory? The obtained results may have been due either to the fact that the subjects' eyes proceed as far as the end of a chain, or to the fact that the subject remembers as many words as go to constitute a "chain".[11]

(2) Syntactical units are, to a large extent, identical with units of meaning. Hence, one might argue that the experimental findings can also be interpreted as reflecting a tendency to decode in units which are *meaningful* strings of words. It is in the nature of the case that it is impossible to treat syntactical and semantic units

[10] Every effort was made to encourage the subject to report everything he had seen, and not only groups of words which made sense, and there was every indication that these instructions were followed. Therefore, the alternative explanation that the results are due to a tendency to respond in meaningful units seems to me to have very little plausibility. But it is not possible to rule it out entirely, when the hypothesis is tested by means of the EVS, as measured in the present experiment. Cf. also note 11.

[11] The EVS as measured in the present experiment involves also peripheral vision, (i.e. vision by means of peripheral portions of the retina). With photographic techniques (2.2) information may be obtained regarding the location of successive fixation points on the line. It remains to be seen, whether these are also related to syntactic constituents. If they are, which seems to me unlikely, the question would be decided in favor of the perceptual process (and, also, any explanation involving response bias (note 10) would be ruled out). On the other hand, if they are not, the question would remain unresolved, since perception by peripheral vision may account for the effect obtained with our procedure.

completely separately,[12] and hence it is probably an idle question to ask whether the unit of decoding is semantically *or* syntactically determined.

(f) *The EVS and the amount of information in the text*

In discussing Johnson's experiment (2.1), it was pointed out that the fact that errors of recall tend to occur at syntactical boundaries, may be attributed, in part at least, to the greater amount of information at the beginnings of constituents. Evidently, a similar explanation might be suggested for the results of the present experiment: the reader's glance tends to be stopped short by words of relatively high information content which, as a rule, appear at the beginnings of syntactic constituents.

To test the validity of this explanation, word-to-word transition probabilities in all experimental sentences would have to be measured, and this laborious task was not undertaken. Instead, the following less conclusive analysis was made. Of the 80 critical words in the experimental sentences which are non-*V*'s (see Table 2.1), 41 were function words – connectives, prepositions and negations – whereas, *V*'s tended to be almost exclusively content words. Now, function words are much more frequent in the language than content words (cf. Aborn and Rubenstein, 1956; Aborn *et al.*, 1959), and therefore can be assumed to carry less information (cf.

[12] It occurred to us, however, that some information relevant to this question might be obtained by distinguishing between the following two cases in which *U* is a non-*V*: (1) *U* is a content word (i.e., a noun, verb, adjective or adverb); (2) *U* is a function word. In a certain sense, cases of (1) represent, as a rule, units which are semantically more complete than cases of (2), in which a function word is kept dangling at the end. An analysis of all trials in which *U* was a non-*V* showed, that (a) when the EVS was either 1 or 2 words long (1) tended to be the case; (b) when the EVS was 3 or 4 words long, the trials tended to fall under (2).

Taken together with the analogous finding that the tendency of the *U* to be a *V* is weaker when the EVS is longer (see 2.2.4, c, above), these data tempt one to speculate as follows: (a) when the EVS is short, it tends to reach the end of a chain, *or, at least, the end of a semantic unit*; (b) when the EVS is relatively long, the reader has less control over it, and he more often fails to reach the end of a chain or even of a semantic unit. Needless to say, further experimentation is required to substantiate this explanation.

also Nicol and Miller, 1959). Hence it can be concluded that in the present experiment, the last word of the EVS tended to carry a relatively *greater* amount of information, and this conflicts with the above alternative explanation.

(g) *A variable not controlled in this experiment*

In designing this experiment only the number of words was taken into account (2.2.3), and not word length; the latter variable was not controlled for. In tachistoscopic recognition – which is somehow similar to the EVS experimental situation – recognition times have been found to increase with word length (even when word frequencies are held constant, cf. McGinnies *et al.*, 1952; Newbigging, 1961). Although it does not seem likely that in our experiment word length could have systematically influenced the results, it was decided to conduct a further experiment in which this variable would be controlled.

2.2.5. *Second Experiment: Method*

This experiment was intended to supplement the previous one and to test the hypothesis regarding the EVS (2.2.2) with word length controlled.

In addition, the present experiment was designed to test the generality of our findings: Conceivably, only those readers who read the text word by word tend to prepare their reading by glancing ahead till the end of a chain, whereas faster readers, who take in many words at a glance, are unable to control the size of their EVS and thus to plan their preparation. It was, therefore, attempted to find out whether the effect of syntactic structure on the EVS is confined to readers who have a short span by comparing two groups of readers differing in the length of their average EVS.

To enable us to control for word length, the experimental sentences (in Hebrew) were constructed in two forms, which will henceforward be called *type 1* and *type 2*. These types differed from each other in two or three consecutive words, which will be called *critical words*. When one of the critical words in type 1 was a *V*, the corresponding word in type 2 was a non-*V* (cf. 2.2.3), and conversely, when one of the critical words in type 1 was non-*V*, the corresponding word in type 2 was a *V*. Any two corresponding words occupied the same length of the line of print. Tables 2.4 and 2.5 describe the construction of the two types.

TABLE 2.4. *Construction of the Two Parallel Forms of Sentences with* Two *Critical Words*

	first critical word	second critical word	
type 1 ...	...	*V*	...
type 2 ...	*V*	...	...

V = last word in chain.

TABLE 2.5. *Construction of the Two Parallel Forms of Sentences with* Three *Critical Words*

	first critical word	second critical word	third critical word	
type 1 ...	...	*V*	...	...
type 2 ...	*V*	...	*V*	...

V = last word in chain.

The hypothesis was that *U* tends to be a *V*, whichever type of sentence is read. This hypothesis was tested only for the critical words, since only for these word length was controlled. By obtaining at the beginning of the experiment a measure of the subject's average length of EVS, the point at which the light was switched off, *T*, could be determined in such a way that in most cases *U* was one of the critical words. The few cases in which this was not so were discarded from the analysis.

Two versions of a text were prepared, which differed in the type of experimental sentences: each one of the experimental sentences appeared in one of the versions as type 1 and in the other version as type 2. Types 1 and 2 were equally distributed between the two versions, as is shown in Table 2.6. An inspection of Tables 2.4, 2.5 and 2.6 shows that in each version, exactly one-half of the critical words were *V*.

TABLE 2.6. *The Two Versions of the Experimental Text*

No. of critical words	Version I	Version II	No. of sentences
two	type 1	type 2	6
two	type 2	type 1	6
three	type 1	type 2	10
three	type 2	type 1	10
		Total:	32

Subjects were divided into two groups by measuring their EVS at the beginning of the experiment:

Slow readers were defined as subjects whose mean EVS was 2-3 words long.

Fast readers[13] were defined as subjects whose mean EVS was more than 3 but not more than 4 words long.

Eight slow and eight fast readers took part in this experiment. Half of the subjects in each of these groups read version I and the other half – version II.

2.2.6. *Second Experiment: Results*

The results of this experiment are summarized in Table 2.7. The prediction that *U* would be a *V* is borne out for sentences with

[13] The terms "slow readers" and "fast readers" seem justified in that the EVS is related to the number of fixations in a line, and, on the other hand, the number of fixations has been shown (Fairbanks, 1937) to be related to speed of reading.

two critical words as well as for those with three critical words.[14] Since exactly one-half of the critical words in each version were *V* (2.2.5), one would expect by mere chance *U* to be a *V* in one half of the cases just as in the previous experiment; actually, *U* was a *V* in about two-thirds of the cases (245 out of 356). This proportion is strikingly similar to the proportion in the previous experiment (cf. 2.2.4 and Table 2.2).

Moreover, this proportion is almost equal for slow and for fast readers, which shows that the effect of syntactic structure is not confined to the former (2.2.5).

In fact, the prediction was borne out for every one of the fast readers and the results for this group were significant at the 1 percent level by a one-tailed sign test (Siegel, 1956). As for the slow readers, for seven out of eight subjects *U* was a *V* in most of the cases, while for the eighth the number of cases in which *U* was a *V* was equal to that in which *U* was a non-*V*. The difference for this group is also significant at the 1 percent level by the same test. The findings of the previous experiment have, thus, been replicated in this supplementary experiment for both slow and fast readers.

For the purpose of the analysis, the EVS was defined, as in the previous experiment (2.2.4, a), as the number of consecutive words after *T* correctly reported by the subject. An additional analysis showed, that by taking into account incorrectly reported words as well (cf. 2.2.4, d), the difference would have been significant at the same level. (Eleven additional cases would have been counted as bearing out the prediction, as against 12 cases which would have been counted as being contrary to the prediction).

In this experiment, as in the previous one, some errors occurred which lead to a "closing" of the chain (cf. 2.2.4, d).

In measuring the EVS of our subjects at the beginning of our experiment (2.2.5), we came across a subject whose mean EVS was 4.4 words long. Although this was such a large span that he

[14] Owing to the range in length of the EVS of the individual subjects, it was unavoidable that *U* should in some cases fall outside the scope of the critical words. These cases – of which there were 79 for the eight slow readers and 77 for the eight fast readers – were not taken into account in the analysis.

TABLE 2.7. *Results of the Second Experiment for Slow and for Fast Readers*

Group	No. of critical words	No. of cases in which U is a V	No. of cases in which U is a non-V
slow readers	two	42	17
(N = 8)	three	77	41
		119	58
fast readers	two	28	21
(N = 8)	three	98	32
		126	53
Total for both groups		245	111

could not be included even in the group of "fast readers", this person was also asked to read the experimental text (the location of T being the same as for the group of fast readers). It is interesting to note that for this subject, too, U was a V in the majority of cases (20, as against 3 in which it was a non-V).

2.3. SUMMARY AND CONCLUSIONS

Previous research has indicated that decoding proceeds in "chunks" rather than in units of single words, and it has been surmised that these "chunks" correspond to the syntactic units of a sentence (2.1). It follows from our metahypothesis that at *some* level the unit of decoding should, indeed, be the syntactic constituent. To test this hypothesis, the relationship of the eye-voice span to syntactic structure was investigated. It was predicted that the eye-voice span would reach the end of either a syntactic constituent or, of a "chain", which was defined as a group of words that the reader in his left-to-right perusal of the sentence might take to be a constituent.

Two experiments were conducted to test this prediction; in the second of these, word length was controlled for, and separate analyses were made for slow and for (relatively) fast readers. The results of these experiments were as predicted. Together with those of an experiment by Wales on the recall of sentences presented in parts which either did or did not correspond to its constituents (2.1), our results converge on the explanation that decoding proceeds in units which are syntactically determined. It was pointed out, however, that syntactical and semantic units overlap to a large extent, and that it does not seem feasible to keep these two variables unconfounded. At any rate, our metahypothesis, stating that the human user of the language functions analogously to a mechanism built to process sentences by a phrase structure grammar (1.2), has been shown to yield a research hypothesis which has been strongly confirmed. The following chapters will be devoted to the investigation of further hypotheses based on this metahypothesis.

In conclusion, mention should be made of suggestions to facilitate the reading process which have been advanced by several writers, and which are based on the principle that the line of print should be divided into syntactic units (North and Jenkins, 1951; Nahinsky, 1956; Coleman and Kim, 1961). In the light of our findings, it may be suggested that these suggestions hold promise for success. The reading process may, conceivably, be facilitated by arranging the printed words in groups corresponding to the units of decoding.

3

EFFECTS OF GRAMMATICAL TRANSFORMATIONS

This chapter will be concerned with the behavioral effects of grammatical transformations. Chomsky (1957, ch. 7) has shown that a parsimonious description of the sentences of English requires a transformational grammar; from a kernel of simple declarative, active sentences, other sentences are obtained by means of non-obligatory transformations carried out on the phrase-markers of the *kernel sentences*.[1] Thus, from the kernel sentence *John ate an apple*, the following sentences can be obtained by transformations: *Did John eat an apple? John did not eat an apple. What did John eat? An apple was eaten by John.*, and others. The case for a transformational grammar rests amongst others on the claim that while the latter sentences cannot be generated economically by a constituent structure grammar, they are related systematically to the underlying kernel sentence. Moreover, the approach of transformational grammar accords well with our intuitive feeling that the above sentences are somehow related to each other.

The following sections will treat of other psychological facts which accord with transformational grammar. Specifically, a number of studies – by other research workers as well as by the author – will be reported which aimed to show that the speaker of a language creates transformed sentences by applying transformations to kernel sentences, and that the hearer or the reader understands such sentences by applying the reverse operation of eliciting the kernel, which we will call *detransformation.* At the time of writing some of these studies are still in progress, others

[1] For a precise formulation and a fuller discussion see, e.g., Chomsky and Miller (1963, section 5).

are awaiting publication;[2] this chapter, then, is intended as a "state of the art" report of a vigorously developing field in psycholinguistics.

3.1. UNDERSTANDING TRANSFORMATIONS[3]

When reading or listening to a sentence which is the result of a transformation, do we decode it by detransforming it into the underlying kernel sentence? Our metahypothesis (1.2) suggests this to be a plausible hypothesis. To test it, we should predict that transformations take more time to understand than the underlying kernel sentences.

3.1.1. *Evaluation Tasks*

The first experiments with transformations were carried out by Wason (1961) and by Eifermann (1961). These authors employed the negative transformation; they compared affirmative and negative statements – such as "Seventy-eight is an even number" and "Sixty-five is not an even number" – which the subject was asked to evaluate by responding either "true" or "false" to the visually presented sentence. In another experimental condition, subjects were presented with incomplete sentences – such as " ... is an odd number", or " ... is not an even number", which they had to complete so as to turn them either into a true statement, or, according to the instruction given, into a false statement. It was found both for English sentences (by Wason) and for Hebrew sentences (by Eifermann) that negative sentences took longer to evaluate and

[2] I am deeply indebted to the workers in this field who informed me of their work while still in progress, and let me have access to material which was unpublished at the time of writing. Some of their research constitutes an advance beyond my own studies reported here, and some of it represents different approaches. But for the cooperation of these researchers, this chapter would have been less up-to-date that it is at present.

[3] For the sake of convenience, the term "transformation" will be used also for the result of a grammatical transformation, i.e., for a sentence obtained by applying a transformation.

to complete than affirmative sentences, and also that more errors were made with the former.

These results are in line with the hypothesis regarding the greater difficulty of transformations, but it is not at all clear whether they are due to the fact that the negative sentences were detransformed in the decoding process. Introspective reports of the subjects which are quoted by Wason and Eifermann (both of whom do not base their research hypotheses on transformational grammar) show that quite different processes might have been at work.[4] The difficulty of negative sentences might have been a *semantic* rather than a syntactic one; this is also suggested by Eifermann's finding that in Hebrew false affirmative sentences were about as difficult to complete as the true negative sentences, (a finding which, however, did not hold for English sentences in Wason's study).

Also, as Eifermann (1961) and Wason (1959) have pointed out, the negation "not" may have an emotional effect; through past experience, it has become associated with inhibition of responses, and this might lead to the longer latencies observed for negative sentences. Support for this conjecture comes not only from introspective reports in a previous study by Wason (1959), but also from Eifermann's finding that when the negation was expressed by the Hebrew term *eynò* (which is not used in prohibitions), the sentences were easier to evaluate than when the term *lo* was employed (which is used also for prohibitions). In her experiment, *eynò* sentences were still more difficult than affirmative sentences. It seems possible that this effect can also be accounted for by emotional factors which are, conceivably, at work where a negative sentence, however formulated, is concerned.[5]

Research on the understanding of transformation has been developed further by Dan Slobin (in an unpublished study) who also

[4] Needless to say, introspective reports, while being of great value where questions of linguistics processes are concerned, can never be conclusive: the detransformation might have been performed automatically without awareness.

[5] A study by Wason and Jones (1963) is pertinent here. They found that when denials and assertions were indicated by nonsense syllables, the difference between negative and positive sentences became smaller, and they attributed the greater effect of the word *not* to its connotations.

included in his study passive, (P), and passive negative, (NP), transformations.[6] Passive transformations were of two types: (a) subject-object non-reversible sentences, i.e. sentences, where it is clear who is the actor and who is acted upon (e.g., *The boy is carrying the pail.*); interchanging the nouns in such sentences results in semantically anomalous sentences (*The pail is carrying the boy.*); (b) subject-object reversible sentences where there are no semantic cues available in deciding which noun refers to the actor (e.g., *The girl is pushing the boy.*). Slobin's subjects – children in the kindergarten, and in the second, fourth and sixth grades, and adults – were shown pictures and had to decide whether a sentence spoken by the experimenter described the exposed picture correctly or not.

As in the experiments described above, by Wason and by Eifermann, negative sentences always took more time to respond to than affirmative sentences. The effect of the passive transformation was dependent on the type of sentence used. For *non-reversible* sentences, there was no significant difference between response times to K and to P, and also no significant difference between response times to N and NP. For *reversible* sentences, the order of difficulty was as follows: K, P, N, NP (with the difference between N and NP falling short of significance). These findings for reversible sentences confirm those of McMahon's study (quoted in Miller and Ojeman McKean, 1964), in which the understanding of negatives took 0.3 – 0.5 seconds longer than the understanding of affirmatives (the difference in Slobin's study being 0.3 seconds on the average), whereas the understanding of passives took only 0.1 – 0.2 seconds longer than the understanding of actives.

In discussing these results, Miller and Ojeman McKean (1964) point out that they do not present any evidence for the operation of syntactic factors. In reversible sentences, the difficulty of the passive seems to be a purely semantic one: when non-reversibility obviates semantic confusion, the passive does not affect response times. As pointed out above, the difficulty of the negative may

[6] The abbreviations – N, NP, etc. – used here and in the following will not necessarily be those used by the writers whose work is being discussed. K stands for "kernel", i.e. active, positive, declarative sentence.

also be due to semantic factors. Moreover, Miller and Ojeman McKean remark that it is possible to look upon the negative as a preverb (just as *be*, or *have*) instead of being introduced by a transformation.

In the experiments described above, the time taken to understand a sentence was measured by having the subject check the statement contained in the sentence against a certain state of affairs (that of a number being either odd or even, or the situation described by a picture). There were, thus, two processes involved: (a) decoding the sentence, and (b) verifying whether the statement contained in the sentence corresponds to some referent. The detransformation hypothesis which we are concerned with in the present section pertains to (a), but, to make sure that decoding has actually taken place, these experimenters included in the experimental task (b), which is certainly a *non-syntactic* process. It is the latter process which might account for the difference in response times between true and false sentences that was found in the experiments cited above. To obtain evidence on the nature of the (according to our hypothesis, syntactic) decoding process, an experimental isolation of the two processes (a) and (b), is indicated.

This was attempted by Gough (in two unpublished studies). In his first study, Gough used reversible sentences, which the subject had to verify by comparison to a picture presented at the time the last word of the sentence was heard. The obtained order of difficulty was the same as that found in Slobin's study for reversible sentences: K, P, N, NP. Gough then reasoned that if these differences are due to syntactic processes in decoding, they should disappear, or at least diminish, when a few seconds' delay intervenes between the presentation of the sentence and the presentation of the picture, because this would give the subject ample time to complete the decoding before the picture is shown. Gough, therefore, replicated his experiment with the same procedure and the same materials, introducing only one modification: the picture was presented three seconds after the last word of the sentence was heard. He found that active sentences were still responded to faster than passive sentences, and affirmative sentences faster than negative sentences; in fact, these differences were not any smaller than in his previous experiment. In the light of these results, it appears unlikely that the difficulty of passives and negatives is due to the decoding process; it seems, rather, that it is due to the subsequent verification process.[7] However, the detransformation hypothesis cannot yet be regarded disconfirmed. As Gough points out, it is possible that in this particular experimental situation, where verification cannot be carried out immediately, the subject stores the transformed sentence in his immediate memory and detransforms it only when the picture is presented. Experimental evidence for this explanation is still lacking.

3.1.2. *Reading Task*

The study of grammatical transformation by means of verification tasks has, thus, so far not succeeded in establishing how such sentences are decoded. Verification tasks are not the only method of studying this problem. The eminently practical issue of the readability of texts containing grammatical transformations can be approached by different means. When we read a text, we are usually not called on to verify whether what is said is true; our understanding is evidenced by our ability to give an account of what we have read, and this ability can be measured objectively by multiple-choice comprehension tests. Such an experiment has been carried out by Coleman (1964).

In two of his experiments, Coleman compared two versions of a text, one of which contained passives, nominalizations, and adjectivalizations, and the other the corresponding detransformations. In two other experiments, only nominalizations were used. Comprehension and recall tests showed that the detransformed

[7] Gough offers several explanations for this. One of these, that passive sentences may take longer to verify than active ones because the former contain more words, is ruled out by a second experiment in which active sentences, such as *The girl hit the boy*, were compared with passive sentences, such as *The boy was hit*; the passive, though *shorter* than its active counterpart, was still found to take longer to respond to.

were better understood than the transformed versions. There are obvious practical implications of these findings; however, as far as the process of understanding transformations is concerned, the interpretation of the results is no less equivocal than that of the results of the above experiments. As pointed out by Jespersen (quoted by Coleman), nominalizations are semantically not quite equivalent to their detransformations, because they lack specific references contained in the latter: compare *An inclusion of this is an admission that it was important*, to e.g., *Since she included it she admits*, etc. Moreover, Coleman had to introduce additional words in the transformed version so as to equalize the corresponding sentences in the two versions for length. The obtained differences in comprehension and recall may thus be due to an unknown extent to the differences in content or in words between the two versions.

To confirm our detransformation hypothesis, one would want to replicate Coleman's experiments with transformations for which it is possible to construct two versions which are equal in all respects except the experimental variable. The passive and the negative transformations fulfil this requirement (cf. 3.3, 3.2, below). Yet it is doubtful whether such overlearned sentence patterns will – for the adult reader, at least[8] – present a difficulty which a comprehension test will be sensitive enough to pick up (cf. 8.1.4, below).[9]

Another way of measuring reading comprehension is to measure reading speed: the rate of reading can usually be assumed to reflect the rate of comprehension. Here, again, it appears doubtful whether the difficulty of passives and negatives can be picked up by this measure (cf. 8.1.3, below).

[8] On evidence for the difficulty of transformations for children, see 3.5, below.
[9] Coleman (personal communication) has suggested the following procedure: an exposition concerning, e.g., mathematics or symbolic logic is used as experimental text, and comprehension is measured by having the reader solve problems based on the text. The merit of this approach seems to lie primarily in the nature of the text used: conceivably, the difficulty of the subject matter might interact with the syntactic difficulty in such a way that the latter becomes measurable (cf. also 5.3.5, below).

3.2. THE CONSTRUCTION OF TRANSFORMATIONS

If transformational grammar is taken as a model of the speaker, it should be expected that encoding a sentence containing a transformation will be a more complex process than encoding a kernel sentence: the former is produced by applying a transformation to the already encoded kernel. The present section describes two ingenious attempts by other research workers to test such a model.

3.2.1. *Anagram Task*

The construction of transformations was investigated by John C. Marshall (unpublished Ph. D. dissertation) by means of an anagram task. He used sentences in the past tense of the following types: P (passive), N (negative), NP, Q (question), QP, NQ, NPQ. The words of these sentences, as well as the punctuation mark, were presented to subjects in a scrambled order, e.g.:

N : book not he. the write did
NP: wasn't. him book by written the

For each scrambled sentence, there was only one way to re-order the words so as to form a grammatical sentence. Marshall predicted that the time taken to re-order the sentence would be a function of its transformational complexity, and the results were fully in accord with this prediction: anagrams of sentences which are the result of applying only one transformation to the kernel – N, P, and Q – took 1.9 seconds on the average each to solve; those of sentences resulting from two transformations – NP, QP, and NQ – took 2.3, 2.2 and 2.4 seconds respectively; and those of the most complex sentence, resulting from three transformations, NPQ, took 2.8 seconds.[10] Solution times were, thus, a linear

[10] In a second experiment, each sentence-anagram was based on a different kernel. Solution times still varied with complexity as predicted, but they were significantly longer than in the first experiment, where the same kernel was used for all sentences. Marshall explains this difference as follows: when the kernel is the same for all anagrams, the subject only has to find what is the

function of the number of transformations, each additional transformation requiring 0.4 – 0.5 seconds.

Anagram solution times can clearly be expected to be a function also of sentence length. So as not to confound the complexity variable with sentence length, Marshall did not include in his experiment the grammatically most simple sentence type – the kernel sentence; such a sentence contains fewer words than any one of the transformations employed in this experiment. Also, he used the contracted form of the negative (*wasn't*) for some of the sentences (as in the above example) in order to keep the number of words in the different anagrams as nearly equal as possible.[11] However, this seems to be only a partial remedy, since it may be assumed that in constructing the sentence the subject must search the scrambled sentence presented to him for morphemes and not for words.

An alternative explanation of the differences in solution times obtained in Marshall's experiment is possible, if we are willing to make some assumptions about the way subjects proceed in solving anagram tasks. Intuitively, it seems clear that when he is presented with, for instance, the scrambled sentence NPQ,

written the? by wasn't book him

the subject attends to seven morphemes – including the segmental morpheme *n't* – and to the question mark. (He may also, of course, attend to the difference between *write* and *written*, but this difference is not very conspicuous, and, being redundant – given both *by* and *was* – can safely be disregarded). He puts these morphemes together, and possibly rehearses the sentence before responding. The solution times – which were measured from the exposure of the sentence till the start of the oral response – can be expected

---

syntactic form of the sentence, whereas in the other case, he must in addition discover the underlying kernel. This finding thus accords well with the transformational model.

[11] Still, Q and NQ sentences contained each one word less than any one of the others. (Also, when the question sign, but not the period mark is counted, QP and NPQ contain one "word" more). It is doubtful, of course, whether this difficulty can account for any of the results.

to be a linear function of the number of elements the subject presumably attends to. If we count the question mark (but not the more common period sign, which carries less information) and *n't*, this is indeed the case, as shown in Table 3.1.

TABLE 3.1. *Results from Marshall's Experiment*

Sentence	Anagram Solution Time (secs)	Number of "Elements"
N	1.9	6
Q	1.9	6
P	1.9	6
NP	2.3	7
NQ	2.4	7
QP	2.2	7
NPQ	2.8	8

Marshall's finding that the number of errors also increased with sentence complexity, can be accounted for by the fact that the more complex sentences consisted of a greater number of elements (Table 3.1), and that in these sentences more errors were possible therefore. It seems plausible that in constructing a sentence, the subject is liable to forget some morpheme or other, and thus comes up with a grammatically less complex sentence. Marshall reports that this actually happened, and there is no need to explain this finding on the basis of a transformational model.

Admittedly, these post hoc explanations can easily be challenged, and, in any case, they should not be taken too seriously. It is important to realize, however, that sentence length may be a confounding variable in any experiment of sentence construction. To assess the effect of this variable, further experiments should be carried out with Marshall's promising technique. One possibility which suggests itself is replicating his experiment with sentences in another tense, e.g., the present perfect, where the number of "elements" does not covary in the same way with transformational complexity as it does in the case of the past tense.

3.2.2. *Matching Task*

A different approach to the investigation of the construction of transformations is taken by Miller and Ojeman McKean (1964), who used sentences in the past tense of the following types: K (kernel), N (negative), P (passive), and NP (negative passive). The subject was presented with one of the sentences and asked to transform (or to detransform) it. After doing so, he pressed a button which lighted up a list of sentences, from which he chose the correct sentence; thus a measure could be obtained of matching time *minus* search time. By subtracting from this measure the time taken to read the presented sentence (which was estimated by additional independent trials), the time taken to construct the required sentence was obtained.

In this experiment time taken for a detransformation (e.g., turning N into K), did not differ appreciably from time taken for the corresponding transformation (e.g., turning K into N). Miller and Ojeman McKean arrive at the estimate of 0.41 seconds on the average required for the negative transformation (i.e., turning K into N or P into NP) or detransformation. This is strikingly similar to the additional time taken to understand negative sentences in McMahon's and Slobin's experiments (cf. 3.1.1, above). The passive transformation (or detransformation) took an average of 0.91 seconds, which is much longer than the additional time taken to understand passives (3.1.1).

When two transformations (as in matching K with PN) or one transformation and one detransformation were called for (as in matching N with P), the time required was, on the average, approximately equal to that of the sum of the negative and the passive transformations. This finding of additivity accords well with the transformational model; by contrast, no such additivity was found for sentences which were not related to each other by transformational grammar (*John had warned the old woman. John was warning the old woman.* and *John had been warning the old woman.*).

Miller and Ojeman McKean see evidence for syntactical operations in their finding that whenever the matching of sentences involved a K sentence, it took less time than when the same transformation but no K sentence was involved: Table 3.2 shows that matching K with N (or vice versa) takes less time than matching P with NP; in both cases, a negative transformation (or detransformation) is – presumably – involved. Similarly, matching K with P takes less time than matching N with NP, and matching K with NP takes less time than matching N with P. This is in line with the transformational model, according to which the grammatically more complex sentences (assuming that they are completely analyzed or synthesized syntactically) require more operations in decoding than the grammatically more simple sentence, K.

It appears, however, that the difference between matchings involving K and other matchings can be attributed, at least in part, to another factor, namely, the number of words the subject has to delete and to add in order to change one sentence into the other. Suppose the subject is given a K sentence – *Jane liked the small boy.* – and is asked to construct its passive form. To do so, he has to add two words, *was* and *by*. If, however, he turns an N sentence – *Jane did not like the small boy.* – into its passive form, he has to delete one word, *did*, besides adding the above two words. This additional operation may well account for the additional time taken for turning N into NP. The last column to the right in Table 3.2 shows for each matching the number of words which have to be added and deleted to turn one sentence into the other. For the passive transformation and for the negative and passive transformations this number is smaller where matching involves K, and consequently the construction times are smaller here. For the negative transformation the number is larger where K is involved, and here the difference in construction time, though not in the direction to be expected on the basis of our explanation, is quite

TABLE 3.2. *Results from Miller and McKean's Experiment*

Transformation	Sentences Matched	Construction Time (secs)	Number of Words Added and Deleted
negative	K and N	0.40	2
	P and NP	0.42	1
passive	K and P	0.81	2
	N and NP	1.01	3
negative and passive	K and NP	1.24	3
	P and N	1.82	4

small, much smaller, indeed, than in the other two cases. Further experiments with the matching technique seem indicated in which an attempt should be made to rule out this alternative explanation. As suggested above (3.2.1), experiments with sentences in different tenses might serve to manipulate the number-of-words factor independently of the number-of-transformations factor.

The study of the transformational hypothesis by means of experiments involving sentence construction has resulted in interesting and imaginative techniques – anagram tasks and sentence matching.[12] Yet the evidence for the hypothesis remains at the time of writing still somewhat equivocal. In the following sections, different approaches to the problem will be discussed.

3.3. THE RECALL OF TRANSFORMATIONS

This section describes experiments on the recall of transformed sentences which are relevant to the hypothesis stated at the begin-

[12] Dr. Arthur Blumenthal of the Center for Cognitive Studies at Harvard University has kindly informed me of a study he is conducting, in which he uses, amongst others, a task combining construction with recall: after hearing a set of simple sentences, the subject is required to transform them, from memory, into one long, complex sentence (or vice versa).

ning of the chapter, that transformations are – psychologically as well as linguistically – more complex than their underlying kernel sentences. It will be seen that two different predictions follow from this hypothesis: (a) transformations make greater demands on memory capacity (3.3.1) and (b) errors of recall will be in the direction of the kernel (3.3.2).

3.3.1. *Transformations and Memory Capacity*

If a transformed sentence is the result of an operation performed on a kernel sentence, it might be expected that it is more difficult to remember than the kernel which underlies it: When the transformed sentence is learned by heart, what must be remembered is the kernel, plus some kind of "label" indicating which transformation has been applied to the kernel.

An experiment bearing on the learning of transformations has been carried out by Roger Wales; at the time of writing, this experiment has not yet been published. Wales compared simple sentences and sentences obtained by conjoining transformations (cf. also 2.1):

(1) The very old man was always sitting down on one of the big chairs.
(2) A man killed the cat and his wife fainted and the boy caught her.

Although both types of sentences were of approximately equal length, the transformed sentences (2) were learned in more trials than the simpler ones (1).[13]

A different line of reasoning underlies an experiment by Savin and Perchonok (in an unpublished study). They presented their

[13] A similar experiment has been conducted by Dr. Arthur Blumenthal (cf. note 12), who composed sets of eight simple sentences, which he then conjoined into a long complex sentence by embeddings and conjunctions. One of the experimental tasks was recall. At the time of writing, the outcome of this experiment is not yet known to me. Another (unpublished) study by Blumenthal, concerning the effectiveness of prompt-words in the recall of different types of sentences, also demonstrates the "psychological reality" of transformations, but does not bear directly on the hypothesis stated at the beginning of this chapter.

subjects with a sentence followed by a string of unrelated words, asking them to remember both the sentence and the extra words. They argued that the number of unrelated words which are recalled by the subject would be smaller, the larger the amount of "storage space" taken up by the sentence. They point out that their procedure is analogous of measuring the volume of a body by placing it in a quart bottle and measuring the amount of water which must be added to fill the bottle. Savin and Perchonok employed N, P, Q, QP and NPQ sentences, as well as the emphatic transformation (e.g., *The ball HAS been hit by the boy.*) and questions of the form: *Who has hit the ball?* As predicted, fewer unrelated words were remembered with these transformations than with kernel sentences. Also, sentences resulting from a single transformation, such as N or P seemed to take up less "storage space" than those resulting from two transformations, such as NP or QP. As in previous experiments, sentence length may have been a confounding factor, since, in general, the sentence grows longer with the number of transformations applied to the kernel, and the longer the sentence, the more "storage space" it can be expected to take up. In the case of the present experiment, however, this factor does not account for all the results. Fewer unrelated words were remembered after the "who" question than after the kernel, although the former is actually shorter by one word than the kernel, and this shows – as the authors rightly claim – that the obtained differences are at least partly due to other determinants. Also, for the other sentence types, the number of unrelated words recalled did not seem to be related in any simple manner to the number of words in the sentence.

The above experiments with learning tasks show that the grammatically simpler sentences are easier to learn, and thus give some confirmation to our transformational hypothesis.

3.3.2. *The Direction of Errors*

On the basis of our transformational hypothesis, it may be assumed that a transformed sentence is stored in memory in the form of a kernel sentence plus the appropriate "labels" indicating which transformations were applied to the kernel (3.3.1). If this is so, it may happen that, when the sentence is recalled, the kernel is still intact whereas the labels have been lost; in this case, the sentence will be recalled erroneously in the form of the kernel sentence. If the original sentence is the result of two transformations, such as NP, it may be the case that only one of the labels is remembered along with the kernel and then the sentence will be recalled in a form nearer to the kernel than the original form: as N, or as P. In short, it is to be expected that *errors of recall tend to be in the direction of the kernel.*

This hypothesis was first put to an experimental test by E. B. Coleman (in press) who employed a reconstruction task with nominalizations. As predicted, sentences in which nominalizations appeared were recalled in their detransformed form, and this occurred more often than errors in the opposite direction. Since nominalizations are not semantically equivalent to the corresponding kernel sentences (cf. 3.1.2), we decided to replicate Coleman's experiment with a different transformation: the passive. Passive sentences in the Hebrew language are especially suitable here, since, unlike in English, the passive is expressed by special verb forms and not by an addition of function words. By using a contracted form for the preposition *by*, it is possible to equate the active sentence and the corresponding passive sentence in regard to word length.

The experiment was, in brief, as follows: The subject was given a set of ten Hebrew sentences, five of which were in the passive voice and five in the active voice. Each sentence was exposed to him for a fixed time: half a second for each word in the sentence. The sequence of the sentences was balanced so as not to give any advantage to the active or the passive sentences. After he had thus read the first five sentences of the set, he was asked to reconstruct the first of these sentences; for this purpose he was provided with a set of small cards, on each of which one of the content words of the sentence was printed (except for the main verb which was never presented, so as not to give any clue as to whether the sentence was active or passive). The other four sentences were then reconstructed in the same manner. This procedure was then repeated for the second set of five sentences.

The results were as predicted: The twelve subjects who each read ten experimental sentences made a total of 32 errors involving a change of sentence form; 22 of these involved reconstructing a passive sentence in the active voice, and only 10 errors involved the reverse change from active to passive. This difference was significant at the 1 percent level (one-tailed) by a Wilcoxon test (Siegel, 1956).[14]

It might be argued that such a tendency to reconstruct passive sentences as active sentences is due to the fact that the active form of the particular sentence is more familiar and is, therefore, preferred in reconstructing when the original form has been forgotten. In choosing the sentence material it was therefore decided to cull the passive sentences from newspapers; the fact that the sentences appeared originally in the passive form seemed to indicate that this form might be more "natural" for this particular subject matter and sound more acceptable than the corresponding active form. As an additional safeguard, the sentences were given in both their forms to ten judges, who were asked to indicate for each pair of sentences (a) which one is "more natural" and (b) which one is "more frequent". The sentences finally chosen for the experiment were predominantly judged to be both "more natural" and "more frequent" in the passive form. Thus the dice were actually loaded against the acceptance of our hypothesis.

Yet our findings are still not unequivocal. True, the observed tendency of passive sentences to be reconstructed in the active form follows from the hypothesis that transformations are detransformed in *decoding*. But alternatively, it may be due to the *encoding* of the sentence in reconstructing it, the active voice being preferred whenever the exact grammatical form of the sentence is not re-

[14] Eight cases in which a sentence was recalled in a greatly changed form, including a change of sentence content, were dropped from the analysis. Five of these involved sentences presented originally in the passive voice and only three – in the active voice.

membered. This may be due to either one of the following two reasons: (1) Active sentences are simpler, since, unlike the passive, they do not require an additional transformation. Such an explanation accords well with the hypothesis based on a transformational model, as stated at the beginning of this chapter. (2) Active sentences are more frequent than passive ones, and thus the subject tends to ascribe the active form to the reconstructed sentence.[15] Although, for the particular sentences in the experiment, judges actually preferred the passive to the active, it is still possible that the greater frequency of the active form *in general* played a part here. Now, as already pointed out (1.3), it is very difficult to rule out alternative explanations based on the argument of frequency. Still, some light may be thrown on the matter by experiments with a recognition technique, as will be shown below (3.4).

Additional findings relevant to the process involved here were obtained in an experiment by Mehler (1963). Mehler presented his subjects with eight different sentences, each of which was of a different grammatical form, being either K, N, P, Q, NP, QP, NQ, or NPQ. After hearing the sentences, subjects were required to recall them and write them down. Mehler found that syntactical errors tended to be in the direction of the kernel. An inspection of his published data shows that this effect is due to those cases where K sentences are involved: as in our experiment which was described above, P sentences were more often recalled as K than vice versa, and similarly, all other sentence forms (with the exception of NPQ) were more often recalled as K than vice versa. For all other cases there was no clear tendency of errors to be in the direction of the kernel.[16] There are sentence pairs where the op-

[15] An experiment by Lane and Schneider (1963) is pertinent here. They found that, when a particular structure predominates in a speaker's corpus, listeners who were asked to judge which speaker had uttered a given sentence tended to assign all sentences of the dominant structure to that speaker.

[16] However, Dr. Mehler has pointed out to me that linguists at present tend to treat NQ as not being the result of two transformations but of one (cf. also Mehler, 1963), and that when this is taken into account, there is indeed a greater number of errors in the direction of the kernel than away from it. In another experiment by Mehler (1964) with obligatory transformations, a tendency for errors to be in the direction of the simpler form was also found. The

posite is true; for instance, P was 30 times recalled as QP, while QP was recalled only 27 times as P.

Why should the kernel sentences be thus privileged? Two explanations come to mind. One is, that whenever a sentence is a product of more than one transformation, these are stored as *one* "label"; the QP sentence, for instance, is stored as a kernel plus a label QP indicating the grammatical form of the sentence, and *not* two separate labels, Q and P. If the label got lost, the sentence would be recalled as K; but no prediction can be made regarding errors turning QP into P versus errors turning P into QP. The weakness of this explanation lies in the fact that it accounts only for 160 out of the total of 985 errors observed in Mehler's experiment.

An explanation of Mehler's data proposed elsewhere (Foa and Schlesinger, unpublished paper) is that the transformed sentence is indeed decoded as a kernel plus syntactical labels – one for each transformation – but that labels may *either* get lost *or* be added. This is especially plausible in the case of Mehler's experiment, because there were four sentences for each transformation (passive, negative, and interrogative), and since the subject would probably have been aware of the transformations employed, he may have supplied an additional label. According to this explanation, the tendency of the different sentence forms to be recalled as kernels must be accounted for by other factors, perhaps frequency of occurrence of kernel sentences (cf. the discussion above). What prediction follows from the hypothesis that labels may be either lost or added? The probability that three labels are either lost or added is smaller than the probability that two of these three are lost or added; similarly, the probability that two are lost or added is smaller than the probability that only one of them is lost or added. This is borne out by the data: for instance, NPQ was re-

differences between the results of the latter experiment and that on optional transformations (i.e. transformations of the kind dealt with in the present chapter) were taken by Mehler as showing that this linguistic distinction is behaviorally relevant.

called as K 2 times, as Q – 14 times, and as NQ – 44 times.[17]

This is, then, a case of different and even conflicting hypotheses being based on the same metahypothesis (cf. 1.2). The latter hypothesis is borne out by the findings of Mehler's experiment much better than the former. These findings might be taken as confirming that transformed sentences are detransformed in decoding.[18] However, an alternative explanation of the findings is possible: The frequency of errors can be shown (Foa and Schlesinger, unpublished) to be a simple function of the similarity between sentences in respect to words and content: The more similar two sentence forms are, the greater the probability that they will be confused in recall. It must therefore be concluded that the evidence for the detransformation hypothesis from experiments on recall is still rather equivocal.

3.4. THE RECOGNITION OF TRANSFORMATIONS

Recognition experiments differ from the experiments on recall described above (3.3) in that the previously learnt material is not reproduced by the subject; instead, he is asked to identify a pre-

[17] An essentially similar explanation has been proposed previously by Miller (1962), except that he treats the different transformations as having equally large effects, (cf. also Mehler, 1964). As a consequence, he would predict, e.g., that errors involving negative *and* passive transformations would be less frequent than errors involving only the interrogative transformation. Our hypothesis is a weaker one, and only predicts that they will be less frequent than errors involving either the negative or the passive transformation.

[18] Yet another argument given by Mehler (1963) in support of the hypothesis that the syntactic form of a sentence is stored separately, rests on the finding that syntactical errors by far outnumbered errors of content. However, this seems to be a consequence of the semantic constraints operating in the sentence material employed by him. Consider, for instance, the following two sentences: *The biologist has not made the discovery. The jewel has been worn by the girl.* Clearly, there are better grounds for remembering that it was not the biologist who wore the jewel than for remembering whether he did or did not make the discovery, or that the sentence telling about it was in the active voice. Findings from another experiment by Mehler and Miller (1964) on retroactive interference are also explained by the writers as due to the fact that syntactic information is stored separately from the sentence content. The above objection to the choice of sentence material seems to apply also to the latter experiment.

viously learnt item when presented with it a second time. The experiments described in the following two sections represent attempts to test two predictions derived from the hypothesis that transformations are decoded in terms of their kernels: (a) Errors of recognition will tend to be in the direction of the kernel (3.4.1); and (b) the greater the number of transformations (or detransformations) needed to arrive from one sentence to another, the smaller the probability that these sentences will be confused in a recognition task (3.4.2).

3.4.1. *The Direction of Errors in Recognition*

Our experiment with a reconstruction task (3.3.2) had demonstrated a tendency to reconstruct passive sentences in the active form. The question was then raised, whether this effect must be attributed to a detransformation of the passive sentence taking place in decoding; the alternative possibility which suggested itself was that the experimental results reflected a preference for active sentences in encoding.

To answer this problem it was decided to conduct a recognition experiment with similar sentence material. Subjects were presented with a "learning list" of active and passive sentences. After listening to these sentences, they were given a "test list" which consisted partly of the transformations or detransformations of the original sentences (i.e., sentences appearing in the learning list in the active form appeared in the test list in the passive form, and, conversely, sentences learnt in the passive form appeared in the test list in the active form), and partly of the original sentences. The subject was asked to indicate which of the sentences in the test list he recognized as having appeared previously in the learning list. Since sentences were assumed to be stored in memory in the form of a kernel plus a syntactic "label", and since it may happen that such a label is forgotten, whereas the kernel is remembered, (cf. 3.3.2) the following prediction was made: When the subject encounters an active sentence in the test list, he will tend to recognize it – erroneously – as the corresponding passive sentence of the learning list (because the latter was stored in the form of a kernel, that is, an active sentence). The opposite error – that of recognizing the passive sentence of the test list as the corresponding active sentence in the learning list – was expected to occur much less often.

It was argued that if this prediction would be borne out, this could not be due to any preference of active sentence in encoding (which was one of the explanations advanced to account for the results in the recall experiment), because no encoding takes place in the recognition task: the subject merely indicates whether he has encountered the sentence before or not. The explanation would have to be accepted that in decoding, passive sentences are detransformed. Twenty Hebrew sentences were used in this experiment; the ten sentences employed in the reconstruction experiment were used, and the other sentences were chosen, taking the same precautions to avoid their active form being in some way preferable to the passive (see 3.3.2). The sentences were recorded on tape, half of them in the active form and the other half in the passive form (with the sequence of actives and passives properly balanced). Each sentence was also typed on two cards: on one card in the active form and on the other in the passive form. Each of the twenty subjects participating in the experiment listened to the sentences on the tape and then was given the pack of cards. He was told to indicate for each card – by moving a lever either to the right or to the left – whether the sentence written on it was *exactly* the same as the sentence he had encountered before or somewhat different from it. (That the difference might pertain to the syntactic form was not pointed out to him). Alternatively, the subject might decide that the sentence was quite new and indicate this by an appropriate movement of the lever.

Remembering a passive sentence in its detransformed active form, should result in the subject's tending to indicate recognition when a card is presented to him which bears the detransformed sentence form, and a judgment of "different" when the original (passive) sentence is presented to him. The opposite error – erroneously recognizing the card bearing the passive sentence and judging the original active sentence as being "different" – was predicted to occur much less often. Yet this prediction was not borne out by the findings: the former error was only slightly more frequent than the latter (55 vs. 43 cases, summing over 20 sentences), and the result was not statistically significant.

This failure to find a tendency of errors to be in the direction of the kernel contrasts with the positive results obtained for passives and actives in the reconstruction experiment (3.3.2). At first it was thought that it might have been due to the particular experimental technique adopted, and it was tried to replicate the experiment with the original sentences presented visually and the test list orally, and also with both the learning list and the test list presented by the taperecorder. But since these attempts did not promise any different results, they were discontinued.[19]

The finding that, for passive sentences, the tendency of errors to be in the direction of the kernel is peculiar to recall (in the reconstruction experiment) and does not appear in a recognition task, can be interpreted in different ways. The most obvious one has already been mentioned (cf. 3.3.2): The tendency for errors in recall to be in the direction of the kernel is not the result of a detransformation taking place in decoding but is due to the encoding process. Both active and passive sentences may be stored in memory as a general "sentence content" which is syntactically "neutral", but in encoding the sentence (in the reconstruction task) the active form is preferred, either because it does not necessitate applying a transformation or else because actives occur more frequently in the language.[20] On the other hand, in recognition no encoding is called for and the effect does not occur.

Alternatively, one might retain the detransformation hypothesis and argue that a recognition technique is really not suitable for testing it: the subject probably remembers the *sound* sequence of the original sentence (in addition to remembering the product of

[19] The possibility should be mentioned here that in our attempt to choose sentences which were not preferable in their active form, (cf. 3.3.2) we were actually tipping the scales against the acceptance of the hypothesis, by choosing sentences which could be expected to occur in the passive form.

[20] This explanation does not seem to cover the case of the interrogative and the negative transformations. Cf. Mehler's experiment, section 3.3.2.

the decoding process), and this might serve as a check against recognition errors; the predicted difference might thus be obliterated. Moreover, if there is a tendency to store a general "sentence content", this would also minimize the difference between the number of errors towards the kernel and the number of errors away from it.

Whatever the reason for the failure of our prediction, our results are replicated by Clifton, Kurcz and Jenkins (1964). In an experiment carried out independently of ours, these investigators employed four sentence forms: K, N, P and NP. They report finding no tendency for errors to be in the direction of the kernel. Additional results of their experiment are discussed briefly in the next section.

3.4.2. *Errors of Recognition and the Number of Transformations*

In the study quoted in the preceding paragraph, Clifton *et al.* (1964), found that recognition errors followed a pattern similar to that observed in recall (3.3.2): confusions involving a change in two transformations occurred significantly less often than those involving a change in only one. An exception were confusions between K and N, which occurred about as frequently as those involving two transformations. The writers suggest that semantic or syntactic similarity may account for the findings. They point out that physical similarity between sentences cannot serve as an explanation, because the similarity in words is greater between K and N than between K and P, whereas confusions between the former pair are less frequent than between the latter.

However, an explanation might ultimately be arrived at which is based on a combination of factors, including semantic and physical similarity. Such an explanation would be in line with one of those suggested for data from recall experiments (3.3.2). The settling of this question must await further experimentation.[21]

[21] An extensive study of all the eight sentence forms has later been published by Clifton (1966), and cannot be reviewed here. The problem of psychological effects of transformations has developed into a very active research area since this chapter was written in 1965, but the work carried out since that time cannot be included in this review.

3.5. THE EVIDENCE FROM LANGUAGE DEVELOPMENT

If transformations are behaviorally more complex than simple sentences, it should be expected that the former will appear relatively late in child language. This prediction is borne out by the available evidence from developmental studies. Writers in the field have mentioned the later appearance of passive (Leopold, 1953/54) and of negative sentences (Fisher, 1934). Subsequent experimental studies with children have corroborated these findings by demonstrating the greater difficulty of transformations in comprehension as well as in production (cf. Fraser, Bellugi and Brown, 1963; Slobin, unpublished ms.; Menyuk, 1963/b, 1964).

Menyuk (1963/b) also examined the changes introduced by children when they were asked to repeat sentences. She found that many of these changes involved a detransformation of the transformed sentence presented to them.

The impairment of language abilities in aphasic disorders is often claimed to show similarities to the pattern of ontogenetic development. Goodglass and Hunt (1958) have demonstrated that in aphasia the *'s* of the possessive – which involves a transformation – tends to be more disturbed than the same phoneme *s* when it indicates the plural, i.e., when the sentence is not a result of an optional transformation.

There is perhaps no need to emphasize that findings like these constitute at best only indirect evidence for the hypothesis concerning processes of transforming and detransforming going on in the human user of the language. The developmental sequences mentioned above can be explained quite well without referring to such a hypothesis.

A different approach to the problem of processes at work in handling transformations has been suggested by Ervin and Miller (1963). They argue that, if transformations are acquired as operations, they might be expected to reveal themselves as a sudden increase in the use of the sentence form in question (just as in the case of the acquisition of phonological contrasts). So far, there is only little evidence that this occurs (cf. Bellugi and Brown, 1964, pp. 31-34, 39).

3.6. SUMMARY AND CONCLUSIONS

In the present chapter, the available research evidence was reviewed bearing on the behavioral correlates of transformational complexity. It was hypothesized that encoding a sentence which, from the point of view of the linguist, is the result of a grammatical transformation, involves the operation of transforming applied to a kernel sentence; similarly, the decoding of such a sentence requires the reverse operation of detransforming.

This hypothesis has given rise to more specific research hypotheses, which have been the subject of investigations carried out in recent years by a number of research workers:

(a) Sentences which are the result of a grammatical transformation will be more difficult to understand (3.1.1), to read (3.1.2), to produce (3.2.1, 3.2.2) and to recall (3.3.1) than grammatically more simple sentences.

(b) Errors of recall and of recognition will be in the direction of the kernel (3.3.2, 3.4.1).

(c) The type of confusion errors in recall and in recognition will be a function of the transformational similarity between sentences: e.g., if sentence x can be turned into sentence y by the application of two transformations (detransformations) it will tend to be less often confused with it than with sentence z into which it can be turned by only one of these transformations (detransformations) (3.3.2, 3.4.2).

(d) Developmental trends in child language will evidence the greater difficulty of transformations (3.5).

So far, hypothesis (b) has not received substantial experimental support. As for (a), (c) and (d), the body of research findings corroborating these hypotheses is impressive indeed; the results of well above a dozen experiments carried out by different research workers with different experimental tasks and testing different predictions, all seem to converge on the general hypothesis stated at the beginning of this chapter. Still, it seems indicated to withhold judgement on this issue for the time being, because, as has been shown in discussing these experiments, other factors may have accounted for the obtained results: sentence length and perhaps also the frequency of the sentence form may have been responsible for the greater difficulty of transformed sentences (see *a*, above), and syntactic and semantic similarity rather than transformational similarity may account for the findings on confusion errors (see *c*, above). It is to be hoped that the work in this area which is now being pursued by researchers in a number of places will serve to elucidate more fully the processes at work in the understanding, production and learning of transformed sentences.

But, whatever the underlying process in handling transformed sentences may be, the fact seems to be well-established that these sentences are more difficult than the grammatically more simple kernel sentences. This finding appears to be of considerable practical importance for the writer of texts and for the teacher of language. If simplification of sentence structure is the objective, one of the means of achieving this may be to dispense with transformations.

A word of caution may be in order here, lest this method be used indiscriminately. Transformations often serve important linguistic functions. It has been pointed out above (3.1.2) that nominalizations lack certain specific references which are bound to occur in the simpler version of the sentence, and therefore they might be preferred on certain occasions. Similarly, the passive construction permits of greater conciseness in that the actor does not have to be mentioned (as in *The house was built*). Occasionally, clumsiness of style can be avoided by using the passive transformation. Consider the following sentence:

> This is the fence which was built by the man who was hired by the foreman who is employed by the farmer.

Turning this sentence into the active form, we get the following rather prohibitive construction:

> This is the fence which the man whom the foreman whom the farmer employed hired built.

Sentences of the latter kind will be the subject of the fifth chapter, dealing with sentence complexity. But first, another aspect of complexity, sentence length, will be treated in the next chapter.

4

THE EFFECT OF SENTENCE LENGTH

Long sentences are usually regarded as being more difficult to read than shorter ones. Sentence length was suggested as a measure of readability as early as 1921 (Kitson, 1921), and since then some measure of sentence length has appeared in almost every readability formula. The question dealt with in the present chapter is: does sentence length serve only as a convenient measure of underlying and semantic determinants of readability (cf. 1.4), or does length *per se* also affect reading difficulty?

4.1. SOME CORRELATES OF SENTENCE LENGTH

This section deals with variables which correlate with sentence length. It is important to control these variables, if the effect of sentence length is to be assessed experimentally.

4.1.1. *Sentence Length and Sentence Structure*

Sentences of complex structure are usually relatively long ones. In one study a correlation coefficient of –.775 was obtained between the average sentence length in words and the number of simple sentences in a text, the coefficients for sentence length and other variables pertaining to sentence structure being of a similar order (Nahshon, S., 1957; cf. also Brinton and Danielson, 1958; Stolurow and Newman, 1959). Therefore, the relationship found to obtain between sentence length and readability might be attributable, at least in part, to the influence of syntactic complexity on difficulty.[1] (The notion of "complexity" is, of course, still in need of explication). This seems also to be the impression of readers who find difficulty with a text, at least of Griffin's (1949) subjects, who referred primarily to sentence structure in discussing the difficulty they had experienced. However, no one-to-one correspondence obtains between syntactic complexity and length; this has been pointed out by Bar-Hillel (1964, pp. 195-196).[2]

Yet another aspect of the relationship between sentence length and structure has to be mentioned here. Only a limited number of syntactic structures are possible with short sentences, and this number increases as the sentence becomes longer (Miller, 1951, p. 137). Since difficulty of perception increases with the number of alternatives (see, e.g., Frick, 1953), this fact might account for the increased difficulty of longer sentences.

Finally, it may be that a tendency exists among writers to use longer – and possibly also more complex – sentences when treating of more difficult subject matter (cf. 4.3, below). This might explain, in part, the relationship between length and readability.

4.1.2. *Sentence Length and Redundancy*

Short sentences are, on the average, more redundant than long sentences. Consider, for instance, the following sentence:

> Many readability studies have been published since 1921 indicating that sentence length correlates with difficulty of reading.

[1] Recently, Jones and Carterette (1963) have argued that sentence length does not influence difficulty, if the sentence does not become "too involved".

[2] Perhaps the lack of such a correspondence is the reason of Stevens and Stone's (1947) paradoxical finding that the very readable prose of William James is more "difficult" by a readability formula than the notoriously difficult style of the psychologist Koffka: the long sentences of James are reflected by the formula, but apparently his sentences are not so involved and complex as to make reading difficult.

To break this sentence up into two simple sentences, a pronoun has to be added.

> Many readability studies have been published since 1921. *These* indicate that sentence length correlates with difficulty of reading.

Further examples can be found ad libitum. The facilitating effect of redundancy has been demonstrated in several studies (see the review by Rubenstein and Aborn, 1960, pp. 292-299), and may be assumed to contribute to the greater reading ease of shorter sentences.

The results obtained in an experiment by Coleman (1962) may have been due in part to this factor. Coleman compared the comprehensibility of different versions of a text; one version contained three passages of ten sentences each, and in the other, these passages were rewritten to contain fifteen short sentences each. (A third version was rewritten to contain only six, long sentences). To rewrite the sentences for the fifteen-sentence version, pronouns had to be added in ten places, i.e., in one-third of the sentences of the original version, and it is not clear to what extent this can explain the finding that the fifteen-sentence version was understood, slightly better than the original one in which sentences were longer. This does not detract, of course, from the practical importance of Coleman's experiment, which shows that readability can be improved by shortening sentences.[3] Yet the question remains to

[3] However, Coleman's method of measuring comprehension raises some problems. He employed the "cloze" procedure, which consists of deleting every *n*th word (in his experiment every fifth word) from the text and having the subject supply the missing words after he has read the complete text (cf. 8.1.5, below). Now, the addition of ten pronouns may have the effect of deletions occurring at different places in the parallel versions, and this introduces an uncontrolled source of variation, since it should be expected that some words are easier to supply than others. If care is taken to delete the *same* words in all versions, the addition of pronouns changes the context of some of the deletions: since the deletion following the added pronoun is preceded by more words than the corresponding deletion in the original version, the former was probably easier to fill in. The limitations of "cloze" procedure are discussed also in section 8.1.5, below.

be investigated, whether sentence length affects readability when redundancy is controlled.

4.1.3. *Sentence Length and the Reading Process*

In decoding a sentence, it may occur that its beginning is not properly understood till the end of the sentence is encountered. (Kainz, 1956, vol. IV, pp. 226-227; Chomsky, 1961a, p. 11, footnote 11), that is, " ... the parts of the message are ... either partially decoded or else retained in the memory as undecoded symbols or groups of symbols until the message as a whole can be decoded." (Wonderly, no date). Some confirmation of this description is to be found in Bryan and Harter's classic study (1899) of the decoding of telegraphic messages, which showed that the decoder tends to lag about ten words behind the word being received at the moment (cf. also 2.1). A subject in one of our experiments on sentence length summed up the matter succinctly, saying: "When I come to the end of a sentence, I erase it, as it were ..." Now, as Klare (1963, p. 170) has argued, the fact that with longer sentences the beginning of the sentence has to be kept in memory for a longer time, may account for the greater difficulty of these sentences in reading. It is to be expected that this difference will be especially marked with slow readers (Wonderly, no date).[4]

Another possible influence of sentence length on readability may be caused by the period mark at the end of the sentence.[5] It is

[4] Some fast readers can read two successive lines simultaneously; four of our subjects in the experiment described below (4.2) claimed that they read in this manner. Another subject, who was interviewed on a different occasion, explained that while reading a line she perceived the line beneath it; as a result, the latter was somehow familiar when she came to it, and this facilitated reading. Obviously, such readers can not be expected to experience any difficulty with long sentences on account of the load imposed on their memory.

[5] From a linguistic point of view it is not clear whether a sentence always ends with the period mark; nor needs such a string of words always be the psychologically effective unit. To avoid complications and ambiguities, however, the term "sentence" in this and the following will be taken to refer to a string of words marked off by period marks.

reasonable to assume that the reader pauses habitually on reading this sign at the end of a sentence, because it marks, in most cases, the end of a meaningful unit. If the reading process is indeed punctuated by such short pauses,[6] then reading should take *longer* when the text consists of many short sentences than when it consists of fewer longer ones.

It thus appears, that there are good reasons for suspecting that sentence length affects reading ease – adversely or otherwise – even when variables such as sentence structure and redundancy are controlled. To investigate the influence of sentence length on two measures of readability – (a) reading rate (cf. Kershner, 1964) and (b) comprehension – an experiment was conducted, which will be reported in the following section.

4.2. READING RATE AND COMPREHENSION AS A FUNCTION OF SENTENCE LENGTH

4.2.1. *Method*

Three versions of the same Hebrew text were prepared, which differed in sentence length. Care was taken to control for other variables which usually covary with sentence length, such as sentence structure (4.1.1) and number of words (4.1.2). The three versions contained exactly the same words, with the following exception: the conjunction "and" was dropped in turning a compound sentence of a version containing longer sentences into simple sentences. In Hebrew, this conjunction consists of a prefix of a single letter, and hence it could safely be assumed that this change did not materially affect the results. Each version contained two passages, A and B, dealing with different topics. Table 4.1 shows the number of sentences and the average sentence length in words of the three versions.

[6] This is also more likely to be the case with relatively slow readers. At any rate, such pauses do not occur universally, as is shown by a pilot study on the eye-voice span, in which we found that the span may skip the period mark and include one or two words of the following sentence.

Fifty-eight subjects took part in the experiment. Each subject was randomly assigned one of the versions of passage A, and one of the versions of passage B. (These two assignments were made independently of each other, and there are thus two independent replications of the experiment).

TABLE 4.1. *Number of Sentences and Average Sentence Length in Words of the Three Versions*

Version	Passage A		Passage B	
	No. of sentences	Average length in words	No. of sentences	Average length in words
"long"	4	45.8	9	21.7
"medium"	8	22.9	14	13.9
"short"	11	16.6	21	9.3

First, the subject silently read a practice passage and answered a comprehension test bearing on its content. This was done in order to familiarize him with the experiment.[7] The practice passage and the test were the same for all subjects, regardless of which versions they were assigned to.

Because of the considerations outlined above (4.1.3) it was decided to include only relatively slow readers in the experiment.[8] To obtain a measure of the subjects' reading rates, they were given a "criterion passage" after the practice passage, and a comprehension test bearing on it. Those who read the "criterion passage" at a rate faster than 267 words per minute were excluded from the experiment.

[7] Kershner (1964) found that when subjects did not know that comprehension questions were to be asked, they read at a faster rate than when they had that knowledge. In the present experiment it was thought advisable to let the subjects have full knowledge as to what type of comprehension would be required by exposing them to the practice passage.

[8] Kershner (1964) also found the effect of difficulty of the text to be greater for slow than for relatively fast readers.

Next, the subject silently read passage A in the version assigned to him, and his reading time was measured. Then he was given a short comprehension test on this passage. The same procedure was followed for passage B.

4.2.2. *Results*

To control for the effect of individual differences in reading rate, an "index of response"[9] was computed for each subject's reading of each of the two experimental passages: this was the ratio between the time taken to read the experimental passage and the time taken to read the criterion passage. A Kruskall-Wallis one-way analysis of variance (Siegel, 1956) showed that there were no significant differences between the two passages.

The results for the three versions are compared in Table 4.2. So as to give a more complete picture, results from a pilot experiment, which was conducted with two other passages, are also included in this table.

TABLE 4.2. *Comparisons of the Three Versions in Regard to Reading Rate*

Passage	Reading Rate		
	fastest	intermediate	slowest
A	"medium"	"long"	"short"
B	"long"	"short"	"medium"
pilot 1	"long"	"medium"	"short"
pilot 2	"short"	"long"	"medium"

[9] See Cox (1958, p. 55). Since the ratio between two normally distributed variables does not necessarily have a normal distribution, a non-parametric test was indicated. The very powerful Kruskall-Wallis test gave the following values: for passage A, $H = 2.1$, $.50 > p > .30$; for passage B, $H = 4.4$, $.20 > p > .10$.

It is apparent from Table 4.2 that there is no systematic trend of reading rate to vary with sentence length. Reading rate is fastest sometimes with relatively long, sometimes with medium, or else (in the pilot study) with short sentences. It is noteworthy that in none of these passages did the "long" sentence version require the longest time to read.

Scores on comprehension tests are usually less sensitive measures of difficulty than reading rates (cf. 8.1.4). The mean number of errors made on three questions of the comprehension test is given, for each version, in Table 4.3. The difference in number of errors between the three versions was not statistically significant.

TABLE 4.3. *Mean Number of Errors Made on Three Questions of the Comprehension Test for Each Version*

Passage	Version		
	Long	Medium	Short
A	0.65	0.75	0.79
B	0.05	0.00	0.00

To summarize, in this experiment, in which the short and long sentence versions did not differ in redundancy, the results of Coleman's study (1962) – where the sentence length variable was less rigidly controlled (4.1.2) – were not replicated; neither reading rates nor comprehension scores showed a significant effect of sentence length.

4.3. CONCLUSION AND PRACTICAL CONSIDERATIONS

The results of the above experiment regarding sentence length should not be generalized beyond the population tested and the linguistic material employed. Conceivably, sentence length may have an effect on the readability of much longer texts or of more difficult subject matter, or else, on the performance of beginning readers (cf. also Holland, 1933).

In our experiment, the effect of sentence length per se was investigated. The fact that this variable did not have any effect on reading performance does not exclude the possibility that the shortening of sentences may have a salutary effect on readability. In fact, Coleman's study suggests just this conclusion(4.1.2). Since length of sentence covaries with sentence structure (4.1.1) and redundancy (4.1.2), the advice so often given to keep sentences short may have some justification. However, this advice should certainly be taken cum grano salis; this was brought home to us by introspective reports of some students who commented on their impressions of reading difficulty after reading all three versions. They remarked that the "short" version was difficult to read because the sentences seemed to be "cut up". Reading habits are apparently a factor here: because the use of short sentences is rather infrequent (at least in certain kinds of subject matter), it makes the impression of being out of place. In the absence of any studies determining the optimum sentence length from the point of view of the reader, all that can be said at present is, that attention should be paid to the possibility of sentences being too short, just as they can be too long.

We have just stated that in certain kinds of writing short sentences are relatively infrequent. The reason for this seems to be a simple one: more complex structures may serve an important communicative function in that they make the organization of content more salient, and point to the relationships obtaining between various "elements" of content.[10] That the "splitting up" of sentences may have adverse effects has been shown by Bar-Hillel (1964, pp. 204-205).

Another point to be borne in mind by the writer and editor of texts is, that shortening sentences usually involves an addition of words (4.1.2), and that the redundancy which is thereby intro-

[10] Thus, some of our students objected to the word "but" appearing at the beginning of a sentence in the "short" version. One of them commented: "When I finish a sentence, I "erase" it, as it were. Now, when a new sentence begins with "but", I have to go back and find what this is in opposition to."

duced may make the text easier to read.[11] However, it may be that this factor yields diminishing returns, and that with very short sentences and long texts the repetitions may grow tiresome.

The intricate links of sentence length with other factors do not make it possible to state any hard and fast rules of good writing. Our experiment only serves to show that when it is isolated as far as possible from other factors, sentence length seems to have little effect on readability.

[11] But will the *total* time taken to read it be less than that taken to read a shorter version containing no redundancy? In a study by Miller (1958), subjects were required to learn strings of letters. With redundant strings, *more* letters were learned but *less* information was conveyed (cf. also Aborn and Rubenstein, 1952; Rubenstein and Aborn, 1954; Hogan, 1961). Experiments on discrimination times of figures with redundant cues are also relevant here (Schlesinger 1965a, 1965b). The case of written material with varying degrees of redundancy seems to merit further study.

5

THE EFFECT OF SYNTACTIC COMPLEXITY

Syntactic complexity is regarded to be one of the major determinants of difficulty of reading. However, readability researchers have so far failed to explicate this concept (cf. 1.1). In readability research, sentence complexity has usually been measured by some index of sentence length, but the latter by itself does not seem to be a factor of reading difficulty, as has been shown in the previous chapter. What, then, in the structure of a sentence affects the reading process?

One aspect of syntactic complexity is transformational complexity. The effects of transformations on decoding and encoding behavior have been discussed in Chapter Three. The present chapter deals with other measures of sentence complexity which are based on linguistic models (cf. 1.2). In his study on sentence complexity, Yngve (1960) has pointed out certain behavioral consequences of his theory. The first part of the present chapter is taken up by the discussion of his work and the possibility of testing psychological hypotheses implied by it (5.1). The second part is devoted to a presentation of hypotheses based on Chomsky's (1957) linguistic theory (5.2), and of attempts to put these to an empirical test (5.3, 5.4).

5.1. YNGVE'S MODEL

Yngve's model, which is explained in section 5.1.1, is intended to describe the encoding process, but it can be adapted to the description of decoding, which is the main concern of the present study (5.1.5). The possibilities will be examined of putting this model to an empirical test (5.1.2, 5.1.3, 5.1.6). In the course of this discussion, certain principles will emerge that are of importance for any attempt at building a model of language behavior, and will therefore also be made use of in the discussion of the psychological implications of Chomsky's theory (5.2).

5.1.1. *Description of the Model*

In this section several features of Yngve's model will be discussed which are of special relevance for the psychology of the decoding process. For a detailed description of his model, Yngve's (1960) paper should be consulted.

Yngve has proposed a mechanism, which produces the sentences of the language. One part of this mechanism is a temporary memory store. Its function can be explained by means of an example. Consider the sentence *Jack climbed the bean-stalk*. The "tree" (or P-marker) of this sentence is presented in figure 5.1.

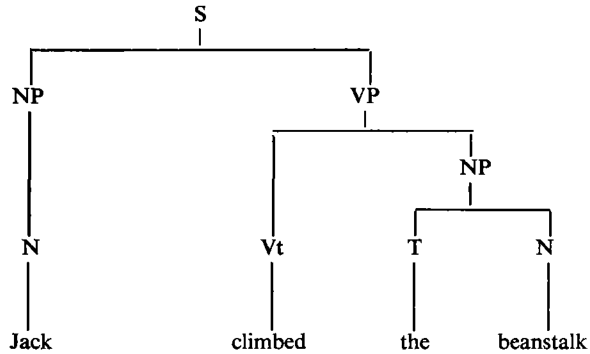

Fig. 5.1. The "tree" of the sentence "Jack climbed the beanstalk".

According to Yngve, sentences are produced within the framework of a simple phrase structure grammar, (a) from top to bottom, and (b) from left to right. Table 5.1 gives a much simplified description of how the words of the above sentence are produced (for a fuller presentation see Yngve, 1960). Thus, "Jack" is produced by starting from S (Sentence), and tracing downwards through NP (Noun Phrase) and N (Noun), and while the mechanism is occupied by this process, VP (Verb Phrase) which is the other symbol derived from S, must be kept in the temporary memory store.

TABLE 5.1. *Production of the Right-Branching Sentence of Figure 5.1 by Yngve's Mechanism*

Mechanism produces	Temporary memory stores
S – NP – N – Jack	VP
VP – V – climbed	NP
NP – T – the	N
N – beanstalk	–

This sentence is an example of *right-branching* sentences, which have the property of requiring minimal storage space for their production. As shown in Table 5.1, when NP enters the temporary memory, the symbol stored previously, VP, has left it in order to produce "climbed". In contrast, left-branching sentences, of which figure 5.2 gives an example, require more storage space.

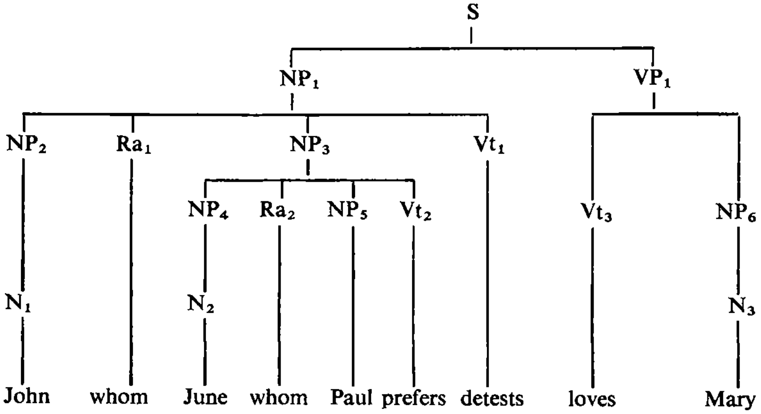

Fig. 5.2. Tree of a left branching sentence (adapted from Bar-Hillel, 1964, p. 199).

For the production of this sentence, a temporary memory which can store five symbols is required. This is the number of items which must be stored while the word "June" is produced, as is shown in Table 5.2.

TABLE 5.2. *Production of the Left-Branching Sentence of Figure 5.2 by Yngve's Mechanism*

Mechanism produces	Temporary memory stores
S – NP_1...	VP_1
... NP_2 – N_1 – John	VP_1, Vt_1, NP_3, Ra_1
Ra_1 – whom	VP_1, Vt_1, NP_3
NP_3 – NP_4 – N_2 – June	VP_1, Vt_1, Vt_2, NP_5, Ra_2
Ra_2 – whom	VP_1, Vt_1, Vt_2, NP_5
NP_5 – N – Paul	VP_1, Vt_1, Vt_2
Vt_2 – prefers	VP_1, Vt_1
Vt_1 – detests	VP_1
VP_1 – Vt_3 – loves	NP_6
NP_6 – N_3 – Mary	–

An important characteristic of the mechanism's memory is that the *sequence* of stored items is preserved: the *last* item to enter the store is the *first* to leave it. Table 5.2 (where new items are entered to the right of the right-hand column) should make it clear why this is essential for the functioning of the mechanism. In the third line, for instance, the symbol Ra_1 must be the first item to leave the temporary memory in order to be further developed, because otherwise a different word would be produced after the word "John". Similarly, VP_1, which entered the store *before* Vt_1, must leave it *after* the latter symbol, so that "loves Mary" appears at its proper place: at the end of the sentence. A finite memory working according to the above principle ("last come – first go") has been called a *push-down store*; the abbreviation PDS will be used in the following.

Any node in the tree from which a symbol branches off to the left, requires *at least* one additional symbol to be stored in the PDS. Now, since the PDS is finite, Yngve's mechanism can produce sentences of only a limited amount of left-branching, or, to use Yngve's term, of a limited *depth*.

Yngve suggests that the human speaker shares the essential properties of this mechanism. It follows that sentences with left-branchings beyond a certain number can not be emitted. He surmises that the number of items the human PDS can accommodate does not exceed Miller's "magical number seven, plus or minus two" (1956). The depth measure, then, can serve as a measure of sentence complexity.

Yngve's interesting discussion of the way language obviates the danger of overstepping the depth limit by providing alternative syntactic structures can not be followed here. One example of this must suffice. The depth of the sentence in figure 5.2 can be drastically reduced by applying the passive transformation:

> Mary is loved by John, who is detested by June, who is preferred by Paul.

Evidently, the sentence is much easier in the latter form (cf. also 3.6).

Yngve's account of the mechanism of a human speaker is, then, based on the following assumptions:

The human speaker

- (a) operates within the framework of a simple phrase structure grammar (with certain amplifications discussed in Yngve, 1960, p. 445);
- (b) generates sentences from left to right, and
- (c) from top to bottom;
- (d) by means of a PDS memory.

It should be mentioned that the plausibility of assumption (c) and the evidence cited in favor of it has been questioned (Chomsky, 1961a; Miller and Chomsky, 1963, pp. 473-475).

5.1.2. *Yngve's Mechanism and the Human Encoder*

It is to be expected that the human organism will differ in some respects from Yngve's mechanism. The latter will be able to produce all sentences of maximum depth with equal facility; unless a certain breaking point is reached, the mechanism is indifferent to the number of left-branchings in the sentence. All that is known about human performance leads us to expect that this will not be true of the human speaker. Even when a task does not exceed the memory capacity of the organism, performance is adversely affected as the limit of this capacity is approached. If the speaker can not emit sentences beyond a maximum depth, it should be predicted that *difficulty* will be encountered well before this maximum is reached.

Other differences are to be expected as well. Results from experiments on immediate memory suggest that in addition to depth other variables may affect the production of sentences. The functioning of immediate memory has been shown to depend not only on the number of items stored (Lloyd *et al.*, 1960; Reid *et al.*, 1961), but also on the interpolation of other stimuli (Brown, 1958; Conrad, 1960), prior information concerning the length of the series to be remembered (Pollack *et al.*, 1959), and the length of time the items have to be retained in storage (e.g., Conrad and Hille, 1958). Now, the length of time an item remains in the PDS will depend, amongst others, on the number of words emitted from the moment the word enters into the store till it leaves it. Hence, various syntactic factors can be expected to interact with depth in affecting the encoder's performance. Consider the following two sentences:

(1) John, who lives next door to us, loves Mary.
(2) John, who lives next door to the man, who married the girl, who knew the boy, who supported the woman, loves Mary.

The clauses interpolated between "John" and "loves Mary" are right-branching, in (2) as well as in (1). Nevertheless, (2) seems intuitively to be more difficult than (1), and may be made still more difficult by the insertion of additional right-branching clauses.[1]

[1] Chomsky (1961a, p. 15) suggests that the difficulty of the following sentence is not accounted for by Yngve's model: "I called the man who read the book that was on the table up." Intuitively, this construction is more difficult

In Yngve's treatment no mention is made of the time variable. There is no need to take this variable into account in discussing the physical possibilities of a *mechanism*, if this is thought of as a mechanical contraption. But any theory of *human* language behavior must relate to what is known about human performance, and therefore certain changes will have to be introduced into Yngve's model, if it is to be applied to the behavior of the speaker. Specifically, it seems that the following predictions ought to be made, which are not provided for by Yngve's presentation:

The speaker's performance (a) will be adversely affected as the (yet to be determined) limit of memory capacity is approached, i.e., as sentence depth increases, and (b) will be influenced by various syntactic variables, which may interact with depth.

5.1.3. *Testing the Model: Depth and Nesting*

In attempting to put Yngve's theory to an empirical test, one will have to cope with the difficulty of keeping the depth variable unconfounded. This is because sentences of greater than usual depth are usually also *self-embedded* sentences. The meaning of this term is explained informally by figure 5.3, presenting a "tree" in which the symbol A is self-embedded.

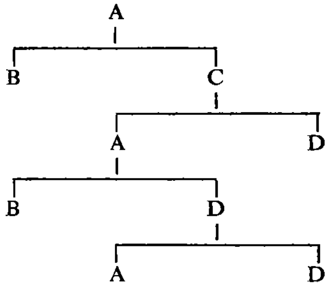

Fig. 5.3. Tree in which the symbol A is self-embedded.

than: "I called up the man who ..." This difficulty may also be due to the amount of time items are stored in memory.

In figure 5.3 a path can be traced from A downwards to another occurrence of this symbol, and from there to a third occurrence of the same symbol; in the latter two occurrences, A is both preceded and followed by other symbols.

Now, it has been shown by Chomsky (1953; 1957) that there exists a degree of self-embedding beyond which sentences cannot be produced by a mechanism with finite memory capacity. Subsequently, the proof has been extended to the case of *nesting*, a more general concept of which self-embedding is a special case (cf. Bar-Hillel, 1964, p. 199). The structure in figure 5.3, for instance, will be a nested structure even if the symbols E and F are substituted for the two last occurrences of A.

Thus, to explain the difficulty of encoding highly nested sentences, only one assumption is needed – the finiteness of memory. Additional, and stronger, assumptions must be made, if one is to predict the difficulty of sentences of great depth (see 5.1.1). If the prediction of difficulty is borne out by an empirical test for sentences which are both of great depth *and* highly nested, the more parsimonious of the two explanations seems to be indicated, namely the one based on the one assumption of finiteness of memory. To provide support for Yngve's stronger hypothesis, proof must be obtained of the difficulty of sentences of great depth, which are *not* highly nested. But, as Yngve (1960) himself points out, such sentences occur only rarely in the language (because, as he suggests, of alternative structures provided by language which make it unnecessary to indulge in too much left-branching). Our experience has shown also that it is hardly possible to construct such sentences for the purpose of an experiment.

There is, thus, a severe practical difficulty in testing Yngve's theory. An approach which suggests itself is the construction of a "languagette", an artificial language, in which the variables of depth and nesting can be kept unconfounded. This possibility is further explored in the next section.

5.1.4. *Testing the Model by Means of a Languagette*

To test Yngve's model by means of an artificial language it is necessary (a) to construct a languagette which permits the production of sentences of the required structures (left-branching, right-branching, etc.), and (b) to create an experimental situation in which the subject is led to utter sentences of different structures, so that (c) various aspects of his performance can be studied as a function of the structure of the sentence.

As for (c), different measures of performance are possible. One of them is the preference of the subject for one sentence form to another; according to the model one might expect that sentences with smaller depth will be preferred to those with greater depth. But a more direct test of the model would be a demonstration of the greater difficulty of sentences of greater depth for the encoder (cf. 5.1.2). Such a test imposes a limitation in regard to (b): the experimental situation must be one in which the choice of the sentence structure (whether left-branching or not) and its depth can be determined by the experimenter.

To meet this requirement, an experimental method was developed by us. An experimental task was devised, and several languagettes were constructed, permitting of right-branching, left-branching (without self-embedding) and self-embedding. The experiment, however, was never carried out, because it became clear to us that an adequate test of Yngve's model would not be possible by our method. Since the difficulty which was encountered here may serve to elucidate an important principle, it may be well to describe the method at some length.

The experiment consists in teaching a subject a languagette, and then present him with some diagrammatic drawings which he is required to describe by means of the languagette. The drawings are such as to require any desired depth or degree of self-embedding (depending on the languagette employed) for their description. As an example, take figure 5.4. on the following page.

This drawing can be verbally described by languagette A (which is a proper part of English), consisting of a very simple vocabulary, the symbol S (Sentence), and four formation rules.[2]

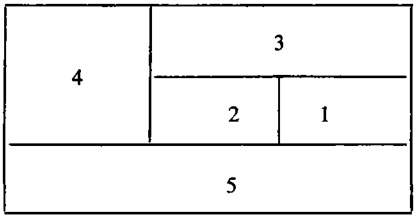

Fig. 5.4. Diagrammatic drawing for an experiment with languagettes. The sequence in which the rectangles are mentioned in a sentence in languagette A is indicated by numbers.

Languagette A – Formation rules

(1) S → a rectangle
(2) S → S + to the right of a rectangle
(3) S → S + in the lower half and a rectangle in the upper half
(4) S → S + in the upper half and a rectangle in the lower half
(Rule 1 is read: "Rewrite *S* as *a rectangle*" etc.)

Figure 5.4 can be described in languagette A as follows (omitting the + sign and adding bracketts for convenience):

> (((a rectangle to the right of a rectangle) in the lower half and a rectangle in the upper half) to the right of a rectangle) in the upper half and a rectangle in the lower half.

Once certain conventions are agreed upon, this description suffices for the reconstruction of the drawing (probably with some distortions – but this need not disturb us here).

It will be seen that languagette A contains only left-branching sentences which are not self-embedded. To enable us to compare the effects of left-branching to those of self-embedding, we construct languagette B which permits self-embeddings in addition to

[2] Greater refinement in describing this and the following languagette may be desirable, but the following is quite sufficient for our purpose.

left-branchings. A few additional formation rules and auxiliary symbols are needed:

Languagette B – Formation rules

(1) S → A
(2) S → B
(3) A → a rectangle
(4) A → S + in the lower half and + S + in the upper half
(5) A → S + in the upper half and + S + in the lower half
(6) B → A + C + a rectangle
(7) C → to the right of

The reader can convince himself that figure 5.4 can be described in languagette B by means of a self-embedded sentence as well as by a left-branching sentence which is not self-embedded. Likewise, a languagette can be constructed in which right-branching as well as self-embeddings are possible.

By means of languagettes and drawings like that in figure 5.4 it was intended to assess the effects of left-branching and of self-embedding. But, as has been stated above, the contemplated experiment was not carried out. The reason is that the use of drawings invalidates conclusions drawn from the experiment regarding the functioning of the temporary memory store, because the subject can make use of the drawing as an additional memory aid. Symbols or words which, according to Yngve, must be stored in temporary memory are represented in the drawing by one of the rectangles. For instance, figure 5.4 is described in languagette A by applying formation rules 4, 2, 3, 2, and 1, in this order. After rule 4 (i.e., S → S + in the upper half and a rectangle in the lower half) has been applied, the symbol S generated by it has to be further developed, and in the meantime the words "in the upper half and a rectangle in the lower half" are stored in temporary memory till the end of the sentence (cf. the sentence describing figure 5.4, above). But the rectangle referred to by the latter expression is represented in the drawing by rectangle 5, and this fact can be made use of by the subject: by checking off mentally which rectangles have been mentioned, he is left at the end with rectangle 5 which he has yet to mention, and this serves as a reminder of the phrase which has yet to be emitted.

Now, the prediction that the difficulty of encoding a sentence will increase with its depth follows from Yngve's model only for such cases where the PDS operates *unaided* by external props to memory. In principle, the capacity of the PDS can be extended indefinitely by external means (such as, e.g., paper and pencil), as Miller and Chomsky have pointed out (1963, p. 467). Therefore, the above method which employs drawings can not be used to test Yngve's model, and it was decided to give up the experiment.[3]

Conceivably, languagettes may be employed in testing the model by means of other experimental techniques like some of those described in chapter 3. Recall and recognition tasks, for instance, may be used in situations which do not involve any additional aids to memory. It is true that by these tasks we do not study only the encoding process, but, as will be shown in the next section, Yngve's model can also make predictions about the decoder.

5.1.5. *Yngve's Mechanism as a Model of the Decoder*

The mechanism described by Yngve produces sentences, and therefore is designed primarily as a model of the encoder. But Miller and Chomsky (1963, pp. 421-422, 465) hold that models of linguistic performance can be interpreted as describing the decoder as well as the encoder. The decoder's task is assumed to be to "reconstruct" the sentence in the same way as the encoder produces it, i.e., to discover by means of successive approximations the process of producing the sentence he has heard (see Halle and

[3] True, if it should turn out that (in spite of these memory aids) left-branching sentences are more difficult to produce than right-branching ones, this would constitute support for Yngve's model. But such an outcome does not seem likely at all: intuitively, one would expect that the subject would learn to make maximal use of the help provided by the drawing and become quite capable of handling the most difficult structures. This is also the impression gained by some informal observations made with this experimental procedure.

Stevens, 1959; and the more general model of Miller, Galanter, and Pribram, 1960). Pursuant to this view, it can be predicted that the depth of a sentence is a determinant of the difficulty not only of its production but also of its understanding, and therefore Yngve's model might be tested by means of experiments on the decoding process.[4]

The difficulty of devising such experiments lies in the fact that most left-branching sentences of considerable depth involve nested constructions; this is the case in Hebrew as well as in English. Experimental evidence of the difficulties attendant on such constructions can not be taken as confirmation of Yngve's model, for reasons which have been stated above (5.1.3). There seem to be two ways to solve the problem of how to construct sentences of great depth which are not at the same time highly nested. One is by means of languagettes such as those described in the preceding section (5.1.4). The other, involves experiments in other languages which permit of different syntactic structures. Slobin (1965) reports an experiment conducted with sentences in Russian by Lushchinkhina, in which it was found that immediate memory for sentences, which were listened to through white noise, was affected by the depth of the sentence. I lack the information whether these sentences were at the same time self-embedded or not.

5.1.6. *The Distinction Between Grammatical and Semantic Decoding*

In exploring the possibilities of testing Yngve's theory, the importance of this distinction was brought home to us. The following example will serve to make this clear. Consider an experiment, in which the subject is required to do mental arithmetic with state-

[4] It should be noted that if negative results are obtained in such experiments, it may still be argued that – contrary to Miller and Chomsky and in line with Hockett (1961; cf. also Osgood and Sebeok, 1954, p. 63) – decoding proceeds on different lines than encoding and, therefore, Yngve's model might still be valid for the description of sentence production.

ments like the following, which he is required to check as being correct or incorrect:

> The product of 6 and a third of the quotient 27 divided by a quarter of the sum of the following two numbers: 8 and the number which results from dividing 20 by the difference between 7 and 2, is greater than 10.

This sentence is obviously very difficult to understand. And yet, its depth is minimal; it is a *right-branching* sentence, as can also be shown by rewriting it:

$$6 \times (1/3\ (27{:}\ 1/4\ (8 + (20{:}(7-2))))) > 10$$

By rewriting this expression so that there are fewer bracketts at the right, it becomes easier to understand. For instance:

$$27{:}\ 1/4\ (8+(20{:}(7-2)))\ 1/3 \times 6 > 10$$

or:

> The quotient 27 divided by a quarter of the sum of the following two numbers: 8 and the number which results from dividing 20 by the difference between 7 and 2 – times one third times 6 – is greater than 10.
> (The dash may be indicated by appropriate intonation).

There is no difference in depth between the two sentences; the reason that the latter one appears to be easier is that there is a smaller number of right-branchings. An experiment with similar material has actually been carried out, and has shown that comprehensibility decreases with the number of right-branchings, a finding which should be obvious in the light of the above example.

The difficulty of right-branching sentences can not be predicted from Yngve's model. At first sight, the above results appear to be damaging to Yngve's theory. Actually this is not so; the results are not even relevant to it. The model accounts only for the syntactic structure of sentences, whereas the solving of an arithmetical problem involves more than the comprehension of syntactic structure. In solving the above problem, symbols which have been encountered at the beginning of the sequence are made use of later in the sequence; for instance, the number 8 must be remembered till the division of 27 by 7 minus 2 has been carried out. In this they are unlike syntactic symbols, which – if Yngve is proved correct – need not be dealt with again once they have been produced.

The need of holding symbols which have already occurred is not peculiar to arithmetical problems. In understanding the predicate of a sentence, it may be necessary to recall the preceding subject (cf. also 4.1.3). A distinction, then, has to be made between semantic encoding (or decoding) and grammatical encoding (or decoding). Only the latter is dealt with in theories like Yngve's or the one discussed in the following section (5.2).

5.2. A HYPOTHESIS BASED ON CHOMSKY'S THEORY

In previous sections (5.1.1-5.1.6) the question was discussed to what extent Yngve's concept of sentence depth can serve as an explicatum of sentence complexity, as far as behaviorally relevant complexity is concerned. The present section deals with a different measure of sentence complexity, which is based on Chomsky's work (1957). The first question to be dealt with is the relationship of Chomsky's linguistic theory to a behavioral theory (5.2.1). Next, a hypothesis will be formulated which will serve as the basis of the rest of the chapter (5.2.2), and the possibilities of testing the hypothesis will be discussed (5.2.3).

5.2.1. *The Relationship of Chomsky's Theory to a Behavioral Theory*

Chomsky's book *Syntactic Structures* (1957) has had a considerable impact on psycholinguistic research (1.1). His grammatical model is not to be viewed as a behavioral model, but only as "a first step towards such a model" (Chomsky 1961a, footnote 14). Various hints of how to develop his model into a behavioral one have been given by Chomsky (1961a, footnote 17) and Miller and Chomsky (1963, pp. 466, 471-472, *et passim*). This and the following sections will be concerned with the behavioral consequences of *nesting* (cf. 5.1.3, for an explanation of this concept).

Chomsky has shown that a device with finite memory capacity is unable to deal with sentences of a degree of self-embedding beyond a certain limit, and the same can be shown to be true of nested structures in general (5.1.3). If we assume that the speaker and hearer of a language incorporate a device which operates along the lines of a Chomskyan grammar (cf. 1.2, above), it follows that they will not be able to produce or understand sentences which are self-embedded or nested beyond this limit. (Miller and Chomsky, 1963, pp. 471-472). There is nothing in Chomsky's theory to specify what this limit is. To formulate an empirically verifiable hypothesis, one might venture a more or less informed guess as to the "breaking point" of the organism in processing nested structures. This is one way of basing psycholinguistic hypotheses on the behavioral model.

That there are other ways, becomes clear from our discussion of Yngve's model (5.1.2). The increasing demands made on memory as the limits of its capacity are reached should lead to the expectation of difficulties in performance. Accordingly, one might predict increasing difficulty of speaking and listening as the degree of nesting increases, and not only a complete breakdown as the (as yet unknown) limit of memory capacity is reached. In addition, one might expect the degree of difficulty to be determined by the length of time items have to be stored in memory (5.1.2). On the basis of these considerations, an empirical hypothesis will now be formulated.

5.2.2. *The Syntactic Decoding Hypothesis*

In presenting the hypothesis it will be convenient to use the term *nested part*, which may be explained by the example of figure 5.5. One kind of nested sentence can be described as

a | John
b | whom June
c | whom Paul prefers
b' | detests
a' | loves Mary.

Fig. 5.5. The nested sentence of figure 5.2, with "nested parts" *b*, *c*, and *b'*.

a b c ... n ... c' b' a'

where *a'* denotes the part "completing" *a*, *b'* – completes *b* etc. All parts *except* the two outermost ones, *a* and *a'*, will be termed nested parts.

The syntactic decoding hypothesis states:

Syntactic decoding of a sentence will be more difficult (a) the greater the degree of nesting, and (b), the greater the length of the nested parts, provided that the decoder's memory is not supplied with external means to increase its capacity.

The latter clause refers to such situations as the one dealt with in section 5.1.4, and the reasons for this exclusion have been given there.

Our hypothesis deals with the decoding process, which is the primary concern of the present study. It can of course be reformulated, mutatis mutandis, to pertain to encoding. Of a more essential nature is the qualification that it deals only with the *syntactic* aspect of decoding i.e., not with semantic processes; the importance of this distinction will have become clear from the discussion in section 5.1.6.

The difficulty of nested constructions becomes intuitively self-evident on considering examples like the one in figure 5.5., above. The operation of memory in decoding a sentence like this may be imagined as that of a PDS (see 5.1.1)[5]: while *a* is being processed, *a'* has to be kept in the PDS; when *b* is processed, *b'* is added to *a'* in the PDS, and so on. In accordance with the basic character-

[5] This is *not* a necessary condition for predicting the difficulty of nested constructions (and, in fact, Chomsky does not mention PDSs in his discussion of the potential difficulties of self-embedding), but such an assumption will be convenient for the purpose of the discussion.

istic of a PDS, *b*' must leave it *before* the item which preceded it in entering the PDS, *a*'. Unless this condition is met, the order of items will not be kept, i.e., the sentence will not be syntactically understood.

The operation of the human PDS can be expected to be affected also by the length of time items are stored in it (cf. 5.1.2), and the length of time is dependent, amongst others, on the length of nested parts. For instance, if part *c* of figure 5.5 were to be changed into something like "whom the well-known journalist and writer of best-selling science fiction detests", the difficulty of the sentence would presumably increase. This consideration leads to the second part of our hypothesis.

The effect of the two variables, degree of nesting and length of nested parts, need not be additive; it is to be expected that the influence of degree of nesting will be greater as the nested parts become longer.

5.2.3. *Ways of Testing the Hypothesis and Previous Research*

One of the approaches to testing the above hypothesis is the customary approach to measuring readability: by means of comprehension and recall tests and the assessment of reading rates. In the following section (5.3) experiments are reported in which texts constituted of sentences differing in degree of nesting were read silently by subjects under ordinary reading conditions so that comprehension scores and reading rates as a function of degrees of nesting could be investigated. It should be pointed out, however, that such experiments do not constitute the most suitable test of the syntactic decoding hypothesis. When the subject is presented with a nested sentence in written form, he is free to revert to earlier parts of the sentence, and by virtue of this fact additional aid is supplied to his memory. As has been pointed out above, supplying such aid to memory may overcome the possible effects of nesting, and our hypothesis does not, therefore, deal with this case (5.2.2).

Still, if *in spite of* the possibility to scan the sentence backwards and forwards degree of nesting is found to affect comprehension and reading speed, this can be taken as a confirmation of the hypothesis. In view of the apparent difficulty of nested sentences (cf. figure 5.5, above), it was expected that this would be the case. This consideration and the obvious practical importance of experiments in ordinary reading situations led to the decision to carry out such experiments. In addition, an experiment was conducted with reading material presented in such a way that reverting to previous parts of the sentence was not possible (5.4).

At the time these experiments were undertaken, no studies on the effect of nesting had become known to us. Meanwhile, several such studies have been published. In one of these, by Miller and Isard (1964), subjects listened to sentences of different degrees of self-embedding – 0, 1, 2, 3, or 4 – and were asked to repeat each sentence verbatim. The effect of self-embedding on difficulty of recall was demonstrated in this experiment: the number of errors increased with the degree of self-embedding. Mandler and Mandler (1964) also report that a self-embedded sentence was more difficult to recall. It should be pointed out, however, that the recall of sentences requires encoding (at the time of recall) as well as decoding (at the time the sentence is presented), and the difficulty of self-embedding may be due to either one of these processes. The same holds true for an experiment by Coleman (in press) where the cloze procedure (cf. 4.1.2, note 3) was used to compare sentences of a low degree of nesting with unnested sentences (e.g., *The man who can sell it is Bill.* vs. *Bill is the man who can sell it.*), and in which the unnested version was better understood, as evidenced by the higher cloze scores obtained with it (see also 8.1.5 for a discussion of the use of cloze scores). Thus, our hypothesis which pertains only to the decoding process, is still in need of confirmation.[6]

That the difficulty of nested sentences is mainly due to the un-

[6] However, the prediction of difficulty of self-embeddings follows from the finiteness of memory (5.2.1) and should therefore hold true for both encoding and decoding; see also 5.1.5 (but see 6.2.4, below). – Mention should be made here of Lushchikhina's experiment (5.1.5) which may be pertinent to the problems of nesting.

related verbs at the end of the sentence (see figure 5.5) is suggested by a study of Mackworth and Bruner (quoted in Miller and Isard, 1964), who recorded eye movements of subjects reading such sentences: recursive eye movements began toward the end of the sentence.[7]

5.3. THE EFFECT OF NESTING IN ORDINARY READING SITUATIONS

Two experiments will be described here, in which the effect of nesting on reading rate and comprehension in silent reading was investigated. Sentences with parenthetical clauses were used as reading material, and these were embedded one within the other, so as to obtain different degrees of nesting. This constitutes a special kind of self-embedding, and the experimental results should not be generalized to other kinds without further experimentation.

It was expected that the nested versions would take longer to read than the unnested ones. The reason for such a result might be that in the ordinary reading situation employed in these experiments, subjects were free to revert to parts of the sentence previously read. Nesting might make such reversals necessary because of the demands it makes on the PDS (5.2.2), and these reversals take time. An increase of reading times with degree of nesting might therefore be taken as a confirmation of the syntactic decoding hypothesis.

5.3.1. *First Experiment: Method*

In a pilot study it was found that sentences which included one parenthetical clause did not take longer to read than similar sen-

[7] Doob (1962) experimented with German sentences in which he separated (a) auxiliaries from the main verb in the principal clause; (b) the subject from its verb in the dependent clause; (c) the prefix from its verb; and (d) an article or its equivalent from its noun. In a free reading situation, no effect of such constructions on subsequent recall was found. In Miller and Isard's experiment, unlike Doob's, self-embedded sentences were employed. Of special interest are their nonsystematic observations that nested constructions which are not self-embedded cause less difficulty. At the time of writing, no systematic investigation of the different effects of self-embeddings and nestings without self-embeddings have been carried out yet, to the best of my knowledge.

tences with no such clause. Apparently, higher degrees of nesting were required to obtain an effect on reading rates, but no information was available as to how high a degree. It was decided, somewhat arbitrarily, to use sentences of what will be called henceforward degree of nesting 3, i.e., sentences with three parentheses nested one within another. An example of such sentences is given in figure 5.6 (translated freely from Hebrew). This example should suffice to convince the reader of the fact that such structures occur only rarely, if at all, and that it would have been very difficult to construct sentences of a still higher degree of nesting.

a The defendant's solicitor demanded,
b since he knew that the court would not,
c in view of the attempts
d revealed subsequently under cross-examination
c' to mislead the police officers in the first stages of the inquiry,
b' accept the defendant's statement,
a' that the fact that his client was the head of a large family should be taken into account in giving the verdict.

Fig. 5.6. A sentence with degree of nesting 3.

The same sentence rewritten so as to represent degree of nesting 1 is shown in figure 5.7.

a The defendant's solicitor demanded,
b since he knew that the court would not
b' accept the defendant's statement,
c in view of the attempts
c' to mislead the police officers in the first stages of the inquiry,
d revealed subsequently under cross-examination,
a' that the fact that his client was the head of a large family should be taken into account in giving the verdict.

Fig. 5.7. The sentences of fig. 5.6 in degree of nesting 1.

The experiment included parallel versions of texts (i.e. versions of equivalent content), each with a different degree of nesting. In addition to degrees of nesting 3 and 1 exemplified in the above figures, a version with degree of nesting 0 (i.e. sentences without parentheses) was prepared. Practical considerations precluded the inclusion of an additional version with degree of nesting 2.

Another independent variable in this experiment was length of nested parts. For each experimental sentence a "long" version was constructed, which differed from the "short" version in the number of words in the nested parts (*b*, *b*', *c*, *c*', and *d*, cf. 5.2.2). The sentence in figure 5.6 is taken from the *short* version; the long version of each sentence was from 18 to 36 words longer than its short version. It was predicted that the effect of nesting would be greater in the long version than in the short one (cf. 5.2.2).

Each subject was presented with a page on which eight Hebrew sentences were typed in one of the following forms (the abbreviation *dn* will henceforward be used for "degree of nesting"):

dn 0 – short version
dn 0 – long version
dn 1 – short version
dn 1 – long version
dn 3 – short version
dn 3 – long version

Some of the sentences contained contradictory statements (for instance, a sentence about a person who subsists on a small budget, and which ends by stating that this person is a spendthrift). Subjects were instructed to read the sentences silently, and to indicate after reading each sentence whether it contained a contradictory statement or not. In this manner we made sure that the subjects paid attention to what they read (as contrasted with mere mechanical reading). Moreover, the number of correct judgments served as a rough measure of comprehension. Another measure of reading difficulty was the number of sentences read during four minutes.

In view of the great variability in reading rates which was found in a pilot experiment,[8] it was decided to test a large number of subjects. A total of 415 subjects was tested in groups of varying sizes.

[8] The variation in reading rates was very large not only between subjects but also within subjects: the ratio between the time of reading of a criterion text (cf. 4.2.2) and the experimental text for individual subjects varied between 0.06 and 0.60 (and occasionally even higher).

5.3.2. *First Experiment: Results*

(a) *Reading rates*

Each subject was given a score which was the number of sentences read during the allotted four minutes (i.e., the number of sentences for which he had indicated his judgment concerning contradictions). Table 5.3 shows the median scores for each of the six forms.

TABLE 5.3. *Reading Rates: Median Scores for the Six Forms of the First Experiment*

	Degree of Nesting		
Version	0	1	3
long	6.21	6.23	5.79
short	7.51	7.54	6.44

The scores for dn (degree of nesting) 1 were slightly higher than those for dn 0, for the shorter version as well as for the longer one, but the difference was not statistically significant. The results, thus, replicate those of the pilot study (5.3.1), in which no difference in reading rates between dn 1 and dn 0 were obtained.

On the other hand, sentences of dn 3 were read at a lower reading rate than those of dn 0 in the short as well as in the long version; this is evidenced by the lower scores for dn 3 in Table 5.3. This difference was significant at the 5 per cent level by a Kolmogorov-Smirnov test (Siegel, 1956) for the short version, but failed to reach the 10 per cent level of significance for the long version. Contrary to our prediction, the differences between scores were smaller in the long version than in the short version.[9]

[9] A possible explanation of this result may be sought in the task presented to the subject – to judge whether the sentence contained a contradiction. Some of our subjects in the present experiment and in the pilot study reported that they read the beginning and end of the nested sentence and then scanned the nested parts. For detecting contradictions this was actually sufficient. Presumably, the tendency to adopt this technique of reading is stronger with long and complicated sentences, and this may have reduced the difference between

(b) *Number of errors*

Each subject judged eight experimental sentences as to whether they contained a contradiction or not; he could therefore make from 0 to 8 errors. The mean number of errors as a function of dn and length is shown in Table 5.4.

TABLE 5.4. *Mean Number of Errors for the Six Forms of the First Experiment*

	Degree of Nesting		
Version	0	1	3
long	1.09	1.12	1.13
short	1.30	1.44	1.42

The mean number of errors was greater for the short version than for the long one. This seems to be accounted for by the fact that more sentences were read in the shorter version (cf. Table 5.3). No effect of dn on number of errors was apparent. (The difference between dn 0 and dn 1 in the short version falls short of significance by a Kolmogorov-Smirnov test, two-tailed [Siegel, 1956]). Although fewer sentences were read in dn 3 than in dn 1 (cf. Table 5.3), the number of errors made in dn 3 was not correspondingly smaller. The reason for this may be that the highly nested version was read with less understanding.

It should be pointed out that the finding of no difference between dn 0 and dn 1 has a parallel in Miller and Isard's (1964) study on the recall of nested sentences (cf. 5.2.3). These writers report that for half of their subjects dn 1 was as easy to recall as dn 0.

5.3.3. *Second Experiment: Method*

In the previous experiment, a difference in reading rates was found between dn 3 and dn 1, while no difference was found between

dn 3 and dn 1 (or 0) in the long version. While not too much importance should be attached to such an ex post facto explanation, it is suggested that the task of the subjects was not suitable for our purposes. A second experiment was therefore designed, in which a different experimental task was employed.

dn 1 and dn 0. Since dn 2 was not included in the previous experiment for technical reasons, the problem which remained to be investigated was, at what degree of nesting would an effect on reading rates make itself felt: only at dn 3 or already at dn 2. The present experiment attempted to answer this question by using three degrees of nesting: 0, 1, and 2. In addition, a replication of the previous experiment was intended in respect to the comparison between dn 0 and dn 1.

Eight experimental sentences were constructed in three forms: dn 0, dn 1, and dn 2. Each of the sixty subjects participating in this experiment read one of these forms. A change was introduced in the experimental procedure (cf. also note 9): instead of judgments of contradictions, subjects were given a comprehension test after reading the eight sentences. Amount of material read was measured by instructing subjects to mark the exact place reached in reading, at the close of one minute.

To control for individual differences in reading rate, a criterion text was presented at the beginning of the experiment and an index of response was computed for each subject (cf. 4.2.2), which was the ratio of the number of words read in the experimental text in one minute to the number of words read in the criterion text in one minute.

5.3.4. *Second Experiment: Results*

The results of this experiment are summarized in Table 5.5, which shows (a) the mean reading rate scores – i.e., the mean of the indexes of response of the individual subject – where a higher score indicates a faster rate (cf. 5.3.3); (b) the mean number of correct answers in the comprehension test. Since subjects answered only those items of the test pertaining to the part of the text which they had read, the latter measure reflects the amount read as well as the quality of reading.

It appears from this table that degree of nesting had little effect on reading rate, and that the effect on number of correct answers was not in the expected direction. None of the differences was statistically significant.

TABLE 5.5. *Mean Reading Rate Scores and Mean Number of Correct Answers on Comprehension Test for the Three Versions of the Second Experiment*

Degree of Nesting	Mean Reading Rate Score	Mean No. of Correct Answers
0	1.02	10.46
1	1.03	15.30
2	1.01	14.08

Sentences of dn 1 were read slightly faster than those of dn 0 in both this and the previous experiment. This finding was replicated in another experiment which investigated the degree of nesting on the eye-blink rate,[10] in which sentences of dn 1 were also read faster than unnested sentences. In none of the three experiments was the difference statistically significant, but the consistency of the finding that nested sentences have a slight advantage suggests that the matter should be followed up by further experiments.[11]

5.3.5. *Conclusions and Practical Considerations*

Our experiments show that in ordinary reading situations nesting has practically no effect on reading rates and comprehension. The

[10] In this experiment, eye-blink rates of dn 0, 1 and 3 were compared. Blink rate was found to increase with dn, but the differences failed to reach statistical significance by a Friedman two-way analysis of variance (Siegel, 1956). Cf. section 8.1.8 on the use of eye-blink rate in readability research.

[11] If the finding proves to be a reliable one, its explanation may have to be sought in the reading process. When fast reading is required, the adult reader presumably tends to skip parenthetical clauses or to scan them rapidly (cf. note 9, above). In this manner he will grasp the main sense of the sentence, and do so faster than by reading all of the sentence at an equal speed. One may speculate, that with unnested sentences, where the material of the parentheses is presented at the end of the sentence, skipping is less efficient. A test of this explanation would necessitate photographic recording of eye movements during reading.

only effect that could be demonstrated was that of dn 3 (5.3.2). Even here the effect was much smaller than might be expected, considering that these sentences (cf. figure 5.6) strike one as being rather monstrous and fairly incomprehensible. Up to dn 2, nesting does not seem to make any difference (5.3.4), and this finding is hardly in accordance with the prescription of readability experts that we should make our sentences as "simple" as possible. Sentences of dn 2 – to say nothing of dn 3 – hardly ever occur,[12] and there seems to be no point in advising us against a vice which we are apparently not likely to indulge in. The finding that sentences of dn 1 are read at the same speed and understood just as well as those of dn 0 also runs counter to our intuition; witness the rather tortuous impression made by the sentence in figure 5.7. Some remarks seem to be in order, therefore, regarding the limited generality of our findings:

(a) The population tested: All our subjects were high-school graduates, who had a good command of the language. Conceivably, nested constructions may be an impediment to fluent reading with younger subjects, people of lesser intelligence, or those who know the language less well. Further study is indicated to test this conjecture.

(b) The measures of readability employed: It may well be that reading rate and comprehension scores are not affected by degree of nesting, because the reader makes up for the greater difficulty of the sentence by investing more effort. Unfortunately, there are no well-perfected techniques of measuring effort in reading (cf. 8.1.8).

(c) The reading situation: The readability expert might be right

[12] Writers in the German language are notorious for the complexity of their sentences, but upon inspection it turns out that the impression of complexity is not due to the *degree* of nesting – which is hardly ever higher than 1 – but, possibly, to the *length* of the parentheses or the nested constructions and to the fact that several of these may be contained within one sentence (one alongside the other, not nested within another). Likewise, complaints against the obscurity of legal jargon do not seem to stem from the degree of nesting; our impression is, rather, that such writing uses predominantly paratactical constructions. Perhaps the complexity of legal writing is primarily one of content and not of sentence structure.

in warning us against "complex" structures, such as nestings, because these may be potential sources of difficulty. Conceivably, this difficulty does not make itself felt in ordinary situations but only when there is "overload" of the system, as for instance under suboptimal perceptual conditions,[13] anxiety,[14] or fatigue. With much longer passages than those used in the above experiments, the cumulative effect of nesting might become evident. Finally, the difficulty of subject matter may conceivably interact with nesting in such a way that the effect of this variable makes itself felt much before dn 3 is attained.

While the possibility can not be ruled out that nesting may affect the reading process adversely under certain conditions, it should be recognized that nested constructions may fulfil important functions. One of these may be *organization* of the subject matter (cf. 4.3). Also, *ambiguities* may occasionally be avoided by nesting. Bar-Hillel (1964, p. 202) has advanced the thesis that some things "cannot be expressed in sentences with a low degree of syntactic complexity, without a loss being incurred in other communicatively important respects". The question what price has to be paid for the simplification of sentence structure cannot be answered summarily, and, in fact, is still in need of investigation. Here only a few examples will be given.[15]

[13] Thus, Pierce and Karlin (1957, p. 3, footnote) report that the number of alternative possible words influences reading rates only when the words are read in very dim light, and Miller *et al.* (1951) have found the influence of context and number of alternatives on intelligibility to be greater under more difficult reading conditions (lower signal-to-noise ratio). An interaction between the latter variable and linguistic variables was also found in an experiment by Miller and Isard (1963). Another possibility of increasing the "overload" of the system would be by using multichannel listening tasks (cf. Broadbent, 1958). – A first step towards exploring the effect of suboptimal reading conditions has been made by us in an experiment in which sentences of dn 0, 1, and 2 of the second experiment (5.3.3) were employed, and the quality of printing was very poor. No effect of dn on reading speed was obtained. However, this technique may well be worth further exploring.

[14] Anxiety has been shown to lower reading speed in an experiment by Chansky (1958).

[15] These have been taken from exercises written by students of a course in "Philosophy of Language and Communication" at the Hebrew University in 1963/64 by Professor Y. Bar-Hillel. They were required to "denest" a sentence

Consider the following sentence:

> Anyone$_1$ who feels that if$_2$ so-many$_3$ more$_4$ students$_5$ whom we$_6$ haven't$_6$ actually admitted are$_5$ sitting in on the course than$_4$ ones we have that$_3$ the room had to be changed, then$_2$ probably auditors will have to be excluded, is$_1$ likely to agree that the curriculum needs revision.

Nesting of dependencies is indicated here by subscripts. Let us look now at different ways of rewriting the sentence so that its degree of nesting becomes lower:

(a) *Splitting up the sentence.* e.g.:
Many more students ... *This may hold to such an extent* that the room had to be changed. If *so*, auditors ...

or:

(a) So many more ... that the room had to be changed;
(b) probably auditors will have to be excluded.
Anyone who feels that if (a) then (b), is likely to agree that the curriculum needs revision.

In splitting up the sentence, words are added (as in the first example – indicated by italics) or auxiliary symbols are introduced (as in the second example). The sentence thus becomes somewhat longer, as is generally the case when sentences are shortened (cf. 4.1.3). This is a price which it will not be always worth while to pay.

(b) *Rewriting the sentence in a lower degree of nesting*, e.g.,

a b ... n ... b' a' as *a a' b ... n ... b'*:
Anyone is likely to agree that the curriculum needs revision, who feels that ...

But note that in this rewritten version "who" is separated from "anyone" by nine words. This may be a new source of difficulty or, in certain cases, even of ambiguity.

---

without, as far as possible, introducing further changes. The material was kindly supplied by Professor Bar-Hillel, and permission to quote the following example of a nested sentence by Chomsky (1963, p. 286) has kindly been granted by the publishers, John Wiley and Sons.

> Alternatively, the wording may be changed thus: ... the room had to be changed, *because* so many ...

This results in a change of emphasis which may be contrary to the writer's intentions. Possibly, every change in word order may introduce some change in emphasis.

(c) *Transformations* may sometimes be employed to decrease the degree of nesting:

> That the curriculum needs revision is likely to be agreed to by anyone who ...

In addition to the shift in emphasis attendant on the change in word order, some difficulty might possibly be introduced by the use of the transformation (cf. chapter 3 on the effects of transformations).

(d) *Paraphrasing* may be an efficient way of "denesting" sentences, at least partially, as is evidenced by the following example:

> ... if the number of students whom ... exceeds that of those we have, to such an extent that the room had to be changed ...

There is, of course, no single answer to the question what changes in shades of meaning and emphasis are introduced by paraphrases. All depends on the skill of the writer and the "price" he is willing to pay.

No claim is being made that the above list is exhaustive. Moreover, nothing is known so far about the conditions under which simplifying sentences in any one of the above ways is "worth" the price which has to be paid. The answer to this question will of course be anything but simple and straightforward and will be obtained only by extensive parametric experiments.

5.4. TESTING THE SYNTACTIC DECODING HYPOTHESIS

The experiments reported in the previous sections with ordinary reading situations were not designed primarily to test the syntactic decoding hypothesis. As pointed out previously (5.2.3), an adequate test of this hypothesis requires a situation in which the reader is not free to revert to the beginning of the nested sentence, because otherwise he is provided with external memory aids and the hypothesis does not apply. In the present section an experiment is reported, which was designed to meet the above condition by projecting sentences on the wall part by part (5.4.1).

According to the syntactic decoding hypothesis, an increase in degree of nesting results in increased difficulty of decoding (5.2.2). In the present experiment this difficulty was predicted to influence the following measures:

(a) *Judgments of grammaticality*: The higher the degree of nesting, the more likely will the subject be to make errors in syntactic decoding, and consequently he will be more likely to judge a perfectly well-formed sentence as being ungrammatical.

(b) *Comprehension scores*: Difficulty of syntactic decoding will affect the subject's ability to understand the sentence correctly.

(c) *Sentence reconstruction*: The ability of the subject to reconstruct the sentence was assumed to be dependent on correct syntactic decoding.

The hypothesized influence of length of nested parts (5.2.2) was also tested in this experiment.

5.4.1. *Method*

Two sets of Hebrew sentences were constructed, differing in length of nested parts. The "short" sentences consisted each of seven parts (cf. 5.2.2 for an explanation of symbols):

a a' b b' c c' d.

They could attain a maximum dn (degree of nesting) 3:

a b c d c' b' a'.

The "long" sentences consisted each of nine parts:

a a' b b' c c' d d' e.

These could be rearranged so that the sentence attains a maximum dn 4:

a b c d e d' c' b' a'.

As in the previous experiments (5.3.1, 5.3.3), nesting was introduced by means of parenthetical clauses; the results of the present experiment are therefore limited to nesting of this kind. An example of the kind of sentences used is given in figures 5.6 and 5.7 (in section 5.3.1, above) which show a "short" sentence (i.e. one consisting of seven parts), in dn 3 and dn 1.

For each part of a sentence (*a*, *b*, etc.) a slide was prepared) – seven slides for each "short" and nine slides for each "long" sentence – and these were projected by means of an automatic slide projector. The sequence of slides varied according to the degree of nesting. The parts of a "short" sentence, for instance, appeared in either one of the following sequences:

a a' b b' c c' d	(dn 0)
a b b' c c' d a'	(dn 1)
a b c c' d b' a'	(dn 2)
a b c d c' b' a'	(dn 3)

In a sense, "long" and "short" sentences differed in the length of "nested parts". This is illustrated by taking, for example, sentences of dn 2:

"short": *a b c c' d b' a'*
"long" : *a b c c' d d' e b' a'*

The length of time that *b'* remains in the PDS is different for "short" and for "long" sentences: With "short" sentences *b'* has to remain in the PDS only while *c*, *c'*, and *d* are being processed, whereas with "long" sentences, *e* has to be processed as well. In accordance with considerations mentioned previously (5.2.2), the effect of degree of nestings was predicted to be greater for "long" sentences.

In order to acquaint the subject with the task of judging the grammaticality of a sentence, he was first presented with a practice sentence which was not well-formed. Translated from Hebrew, this sentence reads:

> Influenza, the epidemic which, because of its power to subdue, as has been known ever since it showed itself to be a most dangerous enemy of Man, at the end of World War I,

when Europe's population suffered from hunger, strong and healthy people.

Three "short" and three "long" sentences constituted the material of the experiment proper. Each subject read all six sentences, in different degrees of nesting. Sequence of sentences and of degrees of nesting was varied so as to control for possible practice and fatigue effects.

After reading each sentence the subject was asked:

(a) To judge whether the sentence was well-formed and to indicate how certain he was of his judgment by checking the appropriate answer in a questionnaire. Care was taken to impress upon the subject the need to base his judgment only on syntactic considerations, and to disregard the truth value of the statement contained in the sentence.

(b) To answer twelve true-false questions pertaining to the sentence.

After reading the last of the six sentences he was required, in addition:

(c) To reconstruct the sentence with the aid of a list of content words contained in the sentence.

Owing to the length of the experiment, it was not thought practicable to require the reconstruction task after each sentence. A rigorous test of the hypothesis by means of this task was, therefore, not possible, since the different degrees of nesting were not equally often represented. Still, it seemed desirable to obtain additional information by means of this task.

A pilot experiment indicated that the number of sentences presented to subjects at a single sitting ought to be limited to three "long" and three "short" ones. Since it was intended to investigate the effect of five different degrees of nesting (0, 1, 2, 3, and 4), it was decided to conduct the experiment in two stages. In stage I subjects read three "short" sentences of dn 0, 1, and 2, and three "long" ones of dn 0, 1, and 2. In stage II a *different* group of subjects read three "short" sentences of dn 1, 2, and 3, and three "long" sentences of dn 1, 3, and 4. (It will be remembered that the maximum dn for "short" sentences was 3, and for "long" sentences – 4).

Thus it was expected to find out at which dn an influence on judgments of grammaticality on comprehension and on reconstruction would make itself felt. In the light of the results obtained, the choice of these particular degrees of nesting for the two stages was perhaps not too fortunate.

5.4.2. *Results*

In designing the experiment, we felt quite confident that our subjects would not be able to handle the more highly nested sentences efficiently. It seemed quite obvious, that sentences like the one in figure 5.6 (in section 5.3.1, above) would not be understood when presented in such a way that reversal to earlier parts of the sentence is impossible. The results of the experiment came as a surprise; not only was there no breakdown in communication even with sentences of dn 3 and dn 4, but subjects performed, on the whole, remarkably well with these sentences. A detailed report of the results for the three experimental tasks is given below:

(a) *Judgments of grammaticality*

Subjects' judgments were assigned scores from 0 to 7 as shown in Table 5.6. (It will be remembered that *all* experimental sentences were well-formed).

TABLE 5.6. *The Scores for Judgments of Grammaticality*

Judgment	Degree of Confidence	Score
	quite certain	0
well-formed	fairly certain	1
	not so certain	2
	not at all certain	3
	not at all certain	4
	not so certain	5
not well-formed	fairly certain	6
	quite certain	7

The median scores for each dn and for "short" and "long" sentences are given in Table 5.7. This table shows that, except for "long" sentences of dn 3, the median score was below 4, i.e., it fell in between judgments of "well-formed" and of "not well-formed". The score rises with dn (though not very consistently), but this difference was statistically significant[16] (at the 5 per cent level) only for "long" sentences in stage II. The explanation for the drop in scores with dn 4 is not clear.

As predicted, the increase in scores with dn tended to be greater for "long" than for "short" sentences.

TABLE 5.7. *Median Scores of Judgments of Grammaticality According to Degree of Nesting and Length of Sentence*

Degree of Nesting	"Short" Sentences		"Long" Sentences	
	stage I	stage II	stage I	stage II
0	0.5	—	1.94	—
1	0.3	0.4	0.4	0.8
2	1.36	1.28	3.5	—
3	—	1.36	—	4.5
4	—	—	—	1.5
Significance Level	$p < 0.2$	$p < 0.1$	$p < 0.2$	$p < 0.001$

(b) *Comprehension scores*

A comprehension test of twelve true-false questions was given for each sentence. Table 5.8 shows the median number of errors made, for each type of sentence.

[16] The differences, for each sentence length and at each stage, were tested by a Friedman two-way analysis of variance (Siegel, 1956). The p-values in Table 5.7 should be accepted only with reservation, since the several tests were carried out on scores which are not independent of each other: at each stage "short" and "long" sentences were read by the same subjects. No test for significance of interaction between sentence length and dn was carried out, because this lack of independence precluded the use of analysis of variance.

TABLE 5.8. *Median Number of Errors for Twelve True-False Questions According to Degree of Nesting and Length of Sentence*

Degree of Nesting	"Short" Sentences		"Long" Sentences	
	stage I	stage II	stage I	stage II
0	1.7	—	0.9	—
1	1.4	1.3	1.3	1.28
2	2.21	2.5	1.5	—
3	—	1.25	—	1.79
4	—	—	—	1.5

The median number of errors was small for all degrees of nesting, and in no case did it reach a quarter of the number of questions. No systematic increase of number of errors with degree of nesting is apparent in Table 5.8, and the differences were not statistically significant. The table also shows that comprehension of "long" sentences was no worse than that of "short" ones, except for dn 3.

(c) *Sentence reconstruction*

One sentence was reconstructed by each subject. Each reconstruction was later categorized by the experimenter as being either "correct" or "incorrect" according to whether the content of the sentence had been preserved. The number of correct and incorrect reconstructions according to degree of nesting is given in Table 5.9.[17] Here, again, no systematic decrease in correct responses as a function of dn is apparent.

[17] The numbers in Table 5.9 are to be accepted with caution, since the decision as to whether a reconstruction is correct or not is sometimes made not without an element of arbitrariness. These data do not lend themselves to a formal testing of the hypothesis, anyhow, because of the experimental design employed (cf. 5.4.1). A few of our subjects were not required to fulfill the reconstruction task, because of lack of time; however, since this task was carried out at the end of the experiment, other results could not have been affected by this omission.

TABLE 5.9. *Number of Correct and Incorrect Reconstructions According to Degree of Nesting*

Degree of Nesting	Correct	Incorrect
0	1	2
1	7	5
2	6	4
3	2	5
4	0	2
Total	16	18

In summary it may be said that the syntactic decoding hypothesis receives only meager support by the results of the present experiment. Only in the case of "long" sentences was a statistically significant difference due to nesting obtained, and even this with only one of the experimental tasks. It is recognized that the experimental procedures of this experiment leave ample room for refinement and that it may reasonably be expected that further experiments may come up with more positive results. However, one conclusion seems to be well established: the influence of dn on performance, if any, is extremely small, even with dn 3 and 4. At the time, it was so surprising to realize this, that instead of trying to hunt down the effects of nesting by further, improved experiments, we were led to a reconsideration of the syntactic decoding hypothesis. The conclusions which were reached eventually are presented in the next chapter, but first some additional findings of the present experiment will be reported, which may serve to shed some light on the decoding process.

5.4.3. *Additional Findings*

(a) *Judgments of grammaticality*

(1) The practice sentence presented at the beginning of the experiment was not grammatical, but about two thirds of our subjects

judged it to be well-formed, and half of these were even "quite certain" that this was so. A look at this sentence (see 5.4.1) will convince the reader that its faulty construction would not have escaped detection if only our subjects would have been given ample time to read and reread the sentence. Yet, since the experimental conditions were such that they could not reread the sentence, they came to the wrong conclusion more often than not.

(2) When a subject judged one of the sentences to be not well-formed, he was asked to state his reason. The reasons given seem to indicate that many of the subjects were guided in their judgment not only by syntactic considerations but also by semantic ones, despite the very explicit instructions given on this point.

The above two observations seem to throw some doubt upon the subjects' ability to carry out this task appropriately. This point will be taken up in the next chapter (6.2.2).

(b) *Sentence reconstruction*

The implications of the following two findings will also be discussed in the next chapter (6.2.2).

(1) As expected, subjects usually succeeded to reconstruct only those sentences which they had judged as being well-formed. This is a rather obvious result; of special interest are the exceptions to this rule: Three subjects judged a sentence as ungrammatical (two of them were "quite certain" of this), and then proceeded to reconstruct the sentence correctly. One might well ask how they did succeed to reconstruct the sentence if they did not understand it (as evidenced by their judgment).

(2) A tendency was apparent to reconstruct a sentence in a lower dn than that of the original sentence. This occurred in one case where the sentence was originally presented in dn 1, six cases where the sentence was presented in dn 2, one in dn 3 and one in dn 4. Of the sentences presented in dn 2, two were reconstructed in dn 1; in the rest of the above cases the reconstruction was in dn 0. It should be noted that this tendency might also be attributed to the encoding process.

5.5. SUMMARY

The problem treated in this chapter was that of the effect of sentence complexity on reading ease. For a measure of complexity we turned to linguistic theories. Two possibilities were discussed: Yngve's "depth" measure (5.1) and the degree of nesting (5.2). The empirical testing of Yngve's model was found to be impracticable, because most left-branching sentences, which are predicted to cause difficulty, are at the same time self-embedded sentences, and the difficulty of the latter is based on fewer assumptions, as has been shown by Chomsky.

Pursuant to this line of thought, a hypothesis was formulated which is based on Chomsky's linguistic theory. This *syntactic decoding hypothesis* states that the difficulty in the syntactic decoding of a sentence increases with degree of nesting and the length of nested parts. This difficulty presumably stems from strain imposed on memory in decoding nested sentences. In investigating the possibilities of testing Yngve's theory it became clear that a test of this hypothesis requires (a) that no external aids to memory be given to the subject; and (b) that a distinction be made between syntactic and semantic decoding (5.2.2). Little is known about the latter process, and our hypothesis pertains only to the former.

In ordinary reading situations the effect of nesting was found to be negligible. Only with degree of nesting 3 could an effect of nesting on reading rate be demonstrated for a population of adult readers, when nesting was introduced by embedding parenthetical clauses one within the other. Such highly nested sentences occur seldom, if ever, outside the laboratory (5.3.5). While these findings should not be generalized beyond the population tested, the measures taken, and the reading material and reading situation employed, it seems safe to state that nesting is a much less powerful variable than is commonly supposed. On the other hand, it was pointed out that indiscriminate simplifying of complex sentences may not only fail to increase readability, but may actually have adverse effects (5.3.5).

Ordinary reading situations cannot provide a test of the syntactic decoding hypothesis, because there the subject is free to revert to the beginning of a sentence (and, thus, to make use of external aids to his memory). Therefore, an experiment was conducted in which such reversals were obviated by presenting the sentences in sections. The results of this experiment were rather disappointing in that no effect of nesting on comprehension and sentence reconstruction could be demonstrated, and the effect of nesting on judgments of grammaticality was small and did not make itself felt in the relatively shorter sentences (5.4.2). These results, and some additional findings which seemed demanding of attention (5.4.3), led to a reconsideration of the decoding process. The nature of this process is the subject of the next chapter.

6

SYNTACTIC AND SEMANTIC DECODING

An implicit assumption underlying the syntactic decoding hypothesis (5.2.2) was that syntactic decoding is not influenced by semantic factors. Such an assumption is in line with much of current thinking among linguists (6.1). Following our failure to find appreciable experimental confirmation for this hypothesis (5.4), a different description of the decoding process was developed which will be presented in this chapter (6.2). Experiments were carried out, the results of which lent support to this description (6.3).

6.1. THE POSTULATE OF THE SEPARABILITY OF GRAMMAR AND SEMANTICS

In his paper *Grammar for the Hearer*, Hockett (1961) proposes as one of his postulates that "Any fact about a sentence used by a hearer in parsing a sentence is itself a grammatical fact" (p. 221). He advances this postulate so as to preclude the possibility that the hearer first understands the meaning of the sentence and then uses this information in inferring the syntactic structure of the sentence.[1]

Such a postulate seems to underlie attempts to describe the human user of the language as incorporating a device which operates on the basis of a grammar and to formulate behavioral hypotheses on this basis (see 5.2.1). The task of the psycholinguist would, indeed, be so much easier if he could treat syntactic de-

[1] Hockett (1961) does not seem to imply that a syntactic process taking place in isolation from semantics is indispensible for an understanding of the sentence; he only describes how syntactic decoding takes place, when it does.

coding and encoding in isolation. So little is known about semantic processes that it would greatly complicate his task, if the latter would turn out to be inextricably linked to syntactic processes. Thus, Katz and Fodor (1963), in their outline of a semantic decoding theory, assume a sentence to be first analysed syntactically with semantic decoding taking place after this. This is in line with Hockett's separability postulate.

The conception of syntactic decoding taking place in a semantic vacuum does not seem to accord with our intuitive notion of the process of understanding a sentence. The discussion in the following section and the experimental studies presented subsequently will show that, as far as decoding is concerned, complete separability of syntactic and semantic processes is an untenable proposition.[2]

6.2. THE SEMANTIC-SYNTACTIC DECODING PROCESS

The decoding of nested sentences was assumed to take place by means of a PDS and without resorting to the meaning of the sentence (5.2.1). In the following, an alternative description, the semantic-syntactic decoding process, is presented (6.2.1, 6.2.3), in view of which, previous findings regarding the effects of nesting appear no longer surprising (6.2.2). Finally, the effect of nesting on encoding will be discussed briefly (6.2.4).

6.2.1. *Description of the Process*

It will be remembered that the syntactic decoding of a nested sentence was assumed to require a memory of the PDS type (5.2.2). An essential feature of a PDS is that it also retains the *sequence*

[2] It is perhaps not unnecessary to emphasize that this and the following have no direct bearing on the controversy about the place of meaning in grammatical description (e.g., Chomsky, 1957; Putnam, 1961). Grammar may be treated without having recourse to meaning, and at the same time the process of understanding may include semantic factors. This psychological process is the sole concern of the present chapter.

of the items stored. This is necessary when decoding a sentence such as

$$a\ b\ c\ d\ c'\ b'\ a',$$

because syntactic understanding of the sentence requires that *c*' be recognized as "belonging" to *c*, *b*' to *b*, and *a*' to *a*.[3]

Storing of what has been called *order information* (Crossman, 1961), thus makes the PDS a more powerful tool. It also imposes additional strain on the organism. Retention of order information is particularly difficult, as is shown by such familiar experiences as, for instance, recalling telephone numbers, where one often can remember all digits but not their correct sequence.

It is suggested that nested sentences may be decoded *without* resorting to a PDS. The temporary memory store merely retains the requisite items, and no order information. After all parts (*a b c* ...) of the sentence have been received, the decoder "pieces them together" (that is, pairs *a* with *a*', etc.) with the help of *semantic* cues.[4] Take, for instance, the sentence in figure 5.6 (section 5.3.1, above). There are semantic constraints operating in this sentence; *accept the defendant's statement* quite obviously complements *since he knew that the court would not*, and any other pairing, as, e.g., *in view of the attempts accept the defendant's statement* would be quite meaningless (even without taking into account the omission of *to*). As the reader proceeds through the sentence, expectations (of a semantic nature) are built up; it is quite out of question that *the court would not* should be complemented by *mislead the police officers*, because that is anyway not what one expects courts to do. The content of the sentence thus points the way to correct pairing of *c* with *c*'.

Nested sentences were expected to cause difficulty, because it was assumed that order information must be retained. But, given sufficient semantic cues, this is not necessary, *unless*, of course,

[3] Note that no PDS is necessary for sentences of dn 1; in *a b a*', *a*' can only be paired with *a*.

[4] This process is carried out on parts of the sentence and not, as is generally assumed to be the case in syntactic decoding, on the symbols of the P-marker.

we insist on the separability of grammar and semantics (6.1). The point made here is that, for nested sentences to be understood correctly, syntactic decoding need not be carried out to the end; when syntactic decoding becomes too difficult, the process described above serves to decode the sentence. It is proposed to call this process *semantic-syntactic decoding.*

Stated differently, this means that syntactic structure is to a large extent redundant. There is nothing novel about this. It has long been known that a sentence may be correctly understood when only its "content words" are supplied (Cherry, 1957, p. 119; Wisseman, 1960). According to the explanation presented above, this redundancy may be made use of not only in the particular situation when part of the sentence is deleted, but also when the sentence becomes too difficult because of, e.g., nesting.

No claim is made that this semantic-syntactic decoding process occurs invariably. Quite to the contrary, we conceive of it as a *possible* alternative to purely syntactic decoding (and, perhaps, only supplementing it). This point will be taken up again below (6.2.3).

In the following it will be shown how semantic-syntactic decoding may account for some of our perplexing experimental findings with nested sentences.

6.2.2. *Previous Findings in the Light of the Semantic-Syntactic Decoding Process*

Various findings reported in the previous chapter will be quoted here, and an explanation will be given in terms of the suggested semantic-syntactic decoding process.

(a) *Comprehension is not appreciably affected by degree of nesting* (5.3.2, 5.3.4, 5.4.2, b). If order information is not retained, the number of parts in the sentence, but not the sequence in which they are presented, can be expected to determine the difficulty of decodings and, thus, nesting should not have any effect. It should be pointed out, however, that even with semantic-syntactic decoding a certain amount of difficulty *may* be attendant on higher degrees of nesting, because the pairing of parts by semantic cues may become more difficult the more parts have to be paired. This, perhaps, accounts for the slower reading rates of sentences of dn 3 (5.3.2).

(b) *Judgments of grammaticality are guided by semantic considerations* (5.4.3, a, 2). When syntactic decodings is not complete, judgments of grammaticality can be guided only by the following two considerations: (i) "Do I understand the sentence?" and (ii) "To what extent did I experience difficulty in understanding it?"[5] Semantic considerations, thus, influence the subjects' judgments.

(c) *An ungrammatical sentence tends to be judged as being well-formed* (5.4.3, a, 1). In the light of the above, any sentence will be judged as being grammatical, when it is understood with relatively little difficulty.[6] Most of our free speech has been found to be ungrammatical (McClay and Osgood, 1959; Goldman-Eisler, 1964; Hockett, 1961, pp. 235-236; Cherry, 1957, p. 120), and the fact that this goes unnoticed most of the time is also in line with the above.

(d) *Well-formed sentences which were judged as being ungrammatical, were in some cases correctly reconstructed subsequently* (5.4.3, b, 1). As stated above, judgments of grammaticality may reflect the comprehension of the sentence content. It may there-

[5] In an experiment to be reported below (6.3.2) one subject ventured the information, without being asked about it, that this is the way he judged grammaticality. In a pilot study in preparation for the experiment described in 5.4, one of our subjects, when instructed to make his judgments solely on the basis of grammatical considerations and not to let himself be influenced by the content, claimed that it was impossible to keep the two apart. In retrospect, we tend to agree with him, but at the time his comment went unheeded, because we still believed, that syntactic decoding could, and should, be studied in isolation.

[6] McClay and Sleator (1960) also found that University students often judged quite obviously ungrammatical sentences as being well-formed (see also Hill, 1961, and Chomsky's [1961b] criticism). In our experiment, judgments of grammaticality were still more difficult to make, as there was no possibility to revert to the beginning of the sentence (5.4.1). On the understanding of ungrammatical sentences see also the studies by Ziff and by Katz, in Fodor and Katz (1964 pp. 390-399, 400-416).

fore have happened that a subject failed to recall one or more of the "parts" constituting a sentence, and consequently did not see how the sentence "made sense". This sentence he judged as being "not well-formed". If, after making this judgment, he is presented with a list of words contained in the sentence, this may help him to recall the forgotten sentence parts and, thus, to understand the sentence and to reconstruct it correctly. (It will be remembered that a reconstruction was regarded as being correct even when only the content of the sentence, and not its structure, was preserved). Of course, the subject may want to reverse his judgment of grammaticality after succeeding with the reconstruction task; this was, in fact, the reaction of some of the subjects after completing the reconstruction of a sentence previously judged as being ungrammatical.

(e) *Even highly nested sentences were regarded as being well-formed; and only with the relatively longer sentence material could an effect of nesting on judgments of grammaticality be demonstrated* (5.4.2, a). If the semantic-syntactic process takes place in decoding, even highly nested sentences may be comprehended and, hence, (according to *b*, above), be judged as grammatical. Semantic-syntactic decoding does not necessarily preclude the greater difficulty of understanding nested sentences (cf. *a*, above), and this may account for the lower scores of grammaticality obtained with such sentences.

(f) *Degree of nesting does not seem to affect the ability to reconstruct a sentence correctly* (5.4.2, c). This is a corollary of *a*, above: semantic-syntactic decoding makes possible the correct understanding of highly nested sentences, and, likewise, the reconstruction of their content.

(g) *A tendency was apparent of reconstructing sentences in a "denested" form* (5.4.3, b, 2; this finding was replicated in a later experiment, 6.3.3). With semantic-syntactic decoding, the sequence of the sentence "parts" (*a*, *b*, etc.) is not retained, and it is to be expected, therefore, that in reconstructing the sentence the more usual "denested" form will tend to be preferred.

In the above, some plausible post hoc explanations of previous experimental findings have been presented which are based on the hypothesized semantic-syntactic decoding process. A direct experimental test of the hypothesis was also conducted, but before reporting on this (6.3), some issues pertaining to the semantic-syntactic decoding will have to be discussed.

6.2.3. *The Redundancy of Syntactic Structure*

It has been stated above (6.2.1) that, according to the hypothesized semantic-syntactic decoding process, the syntactic structure of a sentence is to a large extent redundant. This redundancy may be assumed to fulfil an important function, namely, that of combating "noise". In less technical terms, when part of the message fails for some reason to reach the decoder intact, the redundant aspects of the message ensure that communication is successfully consummated. This is generally known to be the function of redundancy as far as physical properties of a message are concerned, but the principle may be extended to hold for other aspects as well.

Tentatively, the decoding process may be described as follows: As a rule, both semantic and syntactic decoding takes place. When the incoming information (in the technical sense of information theory) is more than the organism can deal with, there are several possibilities of dealing with this "overload" (cf. J. G. Miller, 1962). One of these, is the *filtering* of information, which may take the form of retaining information pertaining to the sentence content, while discarding, in part at least, that related to the structure of the sentence, "Overload" in decoding may be due to a variety of factors, such as fatigue, distracting stimuli, difficulty of subject matter, and also, of course, any combination of these. One such factor may be a high degree of nesting. When this factor operates, filtering out syntactic information reduces the overload, and because of the redundancy of this information, understanding is not impaired thereby (cf. 6.2.1).

Now, overload is a matter of degree, and, furthermore, is not specific to nested structures. Hence, some filtering may be expected to occur also with other kinds of sentences. It may well be that with other structures, too, filtering takes the form of discarding part of the syntactic information and relying on semantic cues in decoding the sentence. Should further research corroborate this conjecture, the postulate of separability of grammar and semantics (6.1) will have to be further qualified.

6.2.4. *The Encoding of Nested Sentences*

Highly nested sentences occur only very rarely in the language (5.3.5). According to the syntactic decoding hypothesis (5.2.2), this is just what should be expected: such sentences are difficult to decode and therefore users of the language refrain from using them. In the light of our experiments, which failed to show any appreciable effect of nesting on comprehension (5.2.2, 5.3.4, 5.4.2), this explanation must now be dismissed.

The infrequent occurrence of highly nested structures should, rather, be attributed to difficulty of *encoding*. While decoding of such sentences may be possible without the operation of a PDS (6.2.1), it seems inconceivable that they can be encoded without such a mechanism. Unless order of nested parts is retained, the emitted sentence is likely to be ungrammatical. As a result, sentences of a degree of nesting greater than 1 are very difficult to construct.[7] With dn 1 the situation is different, since retention of order information is not necessary (cf. 6.2.1, note 3), and, in fact, such sentences are not uncommon.

In this connection it should be pointed out that in those studies in which an effect of nesting could be demonstrated, recall tasks were employed (cf. 5.2.3). The subject was required to reconstruct

[7] After years of experimenting with nested sentences, the writer has to admit to experiencing much difficulty in constructing highly nested sentences; usually one or two corrections are necessary to make the spoken or written sentence grammatical. The difficulty of encoding such sentences, then, can serve to explain why they do not occur more frequently. But the rarity of occurrence seems to be due also to the fact that most of the things one speaks or writes about do not lend themselves to being expressed in the form of highly nested sentences.

the sentence in its original form (whether by repeating it verbatim as in Miller and Isard's [1964], experiment, or by filling in deleted words, as in Coleman's study [in press]), and his success in doing so was, thus, dependent on his ability to *encode* the nested sentence. The tendency to "denest" sentences when reconstructing them (5.4.3, b, 2) may also have been due in part to difficulty of encoding.

There seems to be no process of encoding analogous to the semantic-syntactic decoding process. Even if our syntactic decoding hypothesis should be ultimately disconfirmed, the assumption that the human user of the language incorporates a device built along the lines of a Chomskyan grammar (5.2.1) seems to be a fruitful one, since it leads to the hypothesis that nesting causes difficulty in encoding. Here we have an instance of two different hypotheses being based on the metahypothesis which served as the starting point of the present study (1.2).

The description of the decoding and encoding processes suggested in this chapter implies that these two processes differ in their reliance on semantic factors. This means that, contrary to what some researchers assume (5.1.5), they are not to be looked upon as mirror images of each other.

6.3. EXPERIMENTAL VERIFICATION OF THE SEMANTIC-SYNTACTIC DECODING PROCESS

Findings of previous experiments (6.2.2) led to the suggestion of a semantic-syntactic decoding process which has been described above (6.2.1). Yet, to obtain direct evidence for this hypothesized process, independent experiments had to be carried out. The present section reports on two such experiments.

6.3.1. *Predictions*

According to the semantic-syntactic decoding process, the parts of nested sentences are "paired" with the aid of semantic cues.

These make it unnecessary to preserve the sequence of nested parts, and, hence, nesting has little effect on decoding (6.2.1). This process is, thus, dependent on the availability of semantic cues. In fact, all sentences employed in our previous experiments (5.3., 5.4.) provide such cues (see figures 5.6., 5.7 in section 5.3.1 above), and this is most probably true of practically all sentences which actually occur in the language. But it is possible to construct sentences where this is not the case. For instance:

(1) This is the boy, that the man, whom the lady, which our friend saw, knows, hit.

Here there are only minimal (if any) semantic cues which the hearer can make use of in decoding: any one – the boy, the man, the lady, or the friend – might have seen or known or hit either one of the other persons mentioned. Semantic-syntactic decoding can not take place. Either "order information" is retained (and, as a consequence, degree of nesting may be expected to affect decoding, as stated by the syntactic decoding hypothesis, 5.2.2), or else there is a likelihood of the parts of the sentence not being paired off correctly, i.e., the sentence is not properly understood.

It is proposed, then, to distinguish between two types of sentences: *C – sentences*, (cue-sentences) which do contain semantic cues, and which permit, therefore, of semantic-syntactic decoding, and *N-sentences* (no-cue sentences) containing little or no such cues, of which (1) above, is an instance, and for which an effect of nesting is predicted.

The following is an example of a C-sentence:

(2) This is the hole, that the rat, which our cat, whom the dog bit, made, caught.

Intuitively, this sentence is much easier then the N-sentence (1) above (although both are of the same degree of nesting), because here we have semantic constraints operating: It is quite "obvious" that the dog bit the cat, and not vice versa.

It should also be obvious that the rat might be expected to make the hole and the cat to catch the rat; yet it may have escaped the attention of the casual reader that this is *not* what is stated in the above C-sentence, which is, in fact, nonsensical, because it says that the rat caught the hole and the cat made the rat. To make sense, the order of the verbs "caught" and "made" ought to be interchanged. As it stands, sentence (2) is an example of sentences in which reliance on semantic cues leads to incorrect construing of the sentence. From the hypothesized semantic-syntactic decoding process it follows that such sentences will be understood incorrectly, because only if decoding is *not* influenced by semantic factors will understanding of such a sentence be in accord with its syntactic structure.

When order is restored amongst the verbs at the end of the sentence, semantic cues are conducive to understanding the sentence in a way which is compatible with its syntactic structure:

(3) This is the hole, that the rat, (which our cat, (whom the dog bit,) caught,) made.

It is proposed, then, to subdivide C - sentences into two types: *C_c - sentences*, like sentence (3), and *C_i - sentences*, like sentence (2); the subscript indicates that semantic cues leads to an understanding of the sentence which is *correct* – or *incorrect*, respectively – in terms of its syntactic structure.

The types of sentences discussed – N, C_c and C_i – lend themselves to different degrees of nesting. The examples given above are of dn (degree of nesting) 2.

The hypothesis concerning semantic-syntactic decoding leads to the following predictions:

Prediction 1: For any dn above 1, C_c-sentences will be easier to understand than N-sentences.

Prediction 2: For any dn above 1, C_i-sentences will be understood in terms of semantic cues and not of syntactic structure, i.e., the meaning of the corresponding C_c sentences will be assigned to them.

In these predictions the qualification "dn above 1" is introduced, because even for purely syntactic decoding no PDS is required for dn 1 (cf. 6.2.1, note 3). The description of the experiments (6.3.2, 6.3.4) will make it clear how ease of understanding, referred to in prediction 1, and assignment of meaning, referred to in prediction 2, are to be measured.

Only C-sentences lend themselves to semantic-syntactic decoding. In the case of N-sentences the reasoning which led to the formulation of the syntactic decoding hypothesis (5.2.1) seems to be valid, and degree of nesting ought to influence ease of decoding. This leads to the last prediction:

Prediction 3: Degree of nesting will have a greater effect on ease of decoding with N-sentences then with C_c sentences.

Since highly nested sentences may present some difficulty even with semantic-syntactic decoding (6.2.2, a), it is not predicted that dn will have an effect *only* in the case of N-sentences, but rather that this effect will be greater when no semantic cues are available. How this effect is to be measured will be shown in the description of the second experiment (6.3.4-6.3.5).

6.3.2. *First Experiment: Method*

The first experiment was designed to test predictions 1 and 2, above. Two N-sentences and two C-sentences were constructed in dn 2. Each of the C-sentences was employed in two forms: C_c and C_i. In this experiment the sentences were in English.

Each of the eight subjects (four men and four women) participating in the experiment read silently one N-sentence, one C_c sentence, and one C_i-sentence, the sequence of these sentence types being systematically varied so as to counterbalance possible fatigue and practice effects. After reading a sentence, the subject was given the following three tasks: (a) judgment of grammaticality; (b) recall of sentence content; (c) sentence reconstruction:

(a) He stated whether the sentence was grammatical or not. Three practice sentences were given at the beginning of the experiment to acquaint him with this task.

(b) He rendered orally the content of the sentence, in his own words. If this rendering was incorrect, he was told to read the sentence again and then to perform tasks (a) and (b) again till he understood the sentence correctly (up to four readings, at least, if necessary). Slight alterations of the content, such as omission of pronouns, were disregarded for the purpose of this experiment.

(c) He orally reconstructed the sentence, as far as possible verbatim.

The experimenter took a complete protocol of the experimental session.[8]

6.3.3. *First Experiment: Results*

(a) *Results pertaining to prediction 1*

The number of readings required for correct comprehension of C_c sentences was either 1 (four subjects), 2 (one subject), or 3 (three subjects). In contrast, *none* of the subjects succeeded in understanding an N-sentence even after the fourth reading; two gave up the attempt to understand the sentence after the fifth reading, and two others – after the sixth reading; only one of the eight subjects succeeded in rendering the content of an N-sentence correctly and this only after the fifth reading.

The C_c -sentences employed in this experiment were, thus, very much easier to understand than the N-sentences. The difference was significant by a sign test ($p = 0.004$), and prediction 1 is therefore borne out. Semantic cues contained in a sentence thus clearly affect ease of decoding.[9]

[8] The procedure here is evidently less standardized than that of other experiments in this study (e.g., time taken for reading a sentence was not controlled for). This experiment was intended as an exploratory study on the basis of which the more rigorous experiment reported below (6.3.4) was designed.

[9] It may be argued, that there is no evidence in the experimental findings that *syntactic* understanding of N-sentences is more difficult than that of C_c sentences. The subjects may have grasped the syntactic structure of the N-sentence (*a b c d c' b'*) and only have had difficulty in remembering which noun served as subject in *a* and which in *b*. By referring to the protocols of the experiment, however, it became clear that this was not the case, since errors were never limited to an interchange of subjects and predicates. A more rigorous comparison of the syntactic difficulty of C_c and N sentences was carried out in the second experiment (6.3.4).

Furthermore, subjects had more confidence on the grammaticality of C_c -sentences than in that of N-sentences: All eight subjects judged the C_c-sentence as being well-formed (seven of them on the first reading, and the eighth only after the second reading), whereas two judged the N-sentence as being ungrammatical. (Actually, of course, all these sentences were grammatical). Interestingly, the remaining six subjects, who after their first reading of the N-sentence judged it as being grammatical, began to waver in their verdict after subsequent readings, when they realized that they failed to comprehend it. This is in line with previous observations on the influence of comprehension on judgments of grammaticality (5.4.3, a, 2; 6.2.2, b).[10]

(b) *Results pertaining to prediction 2*

After reading the C_i-sentence, seven of the eight subjects rendered its content as that of the corresponding C_c-sentence, thus confirming prediction 2. Only one subject realized that the C_i-sentence does not make sense when its meaning is construed in accordance with its structure. It appears, therefore, that these sentences were decoded on the basis of semantic cues.

(c) *Reconstruction of sentences*

As in the previous experiment (5.4.3, b, 2), a tendency was apparent to reconstruct sentences in a "denested" form. As stated above, the experimental sentences were in dn 2. Three sentences were reconstructed in dn 1 and one sentence in dn 0. It has been shown above (6.2.2, g), that this finding is in accord with the hypothesized semantic-syntactic decoding process. However, this tendency to "denest" sentences may also be a function of the en-

[10] For the sake of curiosity an ungrammatical sentence was also presented to the subjects. This was the N-sentence quoted in 6.3.1, with the words "which our friend" omitted. Five of the eight subjects realized that this sentence was not well-formed. It may be that the decreased length of the sentence was responsible, in part, for the success in judging the grammaticality of the sentence; the overload may have been decreased to such an extent that syntactic decoding became easier (cf. 6.2.3).

coding process, especially in the present experiment, where (unlike the experiment described in 5.4.1) the subject was not provided with any cues in the reconstruction task.

6.3.4. *Second Experiment: Method*

In this experiment it was intended to test prediction 1 (6.3.1) in a more controlled manner than was the case in the previous experiment. In addition, prediction 3 was submitted to an experimental test. Previous findings pertaining to prediction 2 (6.3.3) appeared to be unequivocal, so that a further experiment to test it seemed unnecessary.

Two degrees of nesting were employed in this experiment – dn 0 and dn 2 – and two types of sentences – N and C_c. Two N-sentences and two C_c -sentences were constructed, in Hebrew.[11] Each sentence was formulated in dn 0 and dn 2. Thus, there were four kinds of sentences: N-0, N-2, C_c-0, and C_c-2.

Sixteen people (eight man and eight women) served as subjects. Each subject read one sentence of each of the above four kinds, the sequence being systematically varied so as to counterbalance possible practice and fatigue effects. If the subject read one N-sentence in dn 0, the other N-sentence was presented to him in dn 2, and the same was true for C_c-sentences.

Prediction 3 is based on the syntactic decoding hypothesis, and therefore it was necessary to present the sentences in such a way that the subject could not refer back to parts of the sentence which he had already read (cf. 5.2.3). This was done by means of a simple apparatus in which a cardboard strip with the sentence printed on it was inserted. An apperture exposed about two words of the sentence at a time. The subject could read the sentence at his own

[11] In the previous experiment, in which all sentences were in dn 2, English sentences were used. To reformulate sentences like the N-sentence of 6.3.1 in dn 0 one must either apply to it the passive transformation, or else put up with an awkward (and, possibly, ambiguous) construction like: "This is the boy that the man hit whom the lady knows." The use of Hebrew sentences, which in dn 0 are of a more acceptable form, was therefore decided upon.

speed by drawing the cardboard strip past the apperture. He read the sentences silently. Reading time was measured by the experimenter by means of a stop-watch.

After reading a sentence, the subject was asked to render its content in his own words. If he failed to do so correctly, he had to read the sentence again and to try again to render its content, and so on, up to five times if necessary, till the experimenter was satisfied that he had understood the sentence correctly (cf. the procedure of the previous experiment, 6.3.2). Four practice sentences were given at the beginning of the experiment so as to acquaint the subject with the experimental task. The length of this experiment precluded the use of additional measures of comprehension, such as judgments of grammaticality and reconstruction, which were employed in the previous one. A full protocol of the experimental session was taken by the experimenter.

6.3.5. *Second Experiment: Results*

(a) *Defining syntactic and semantic comprehension*

Although our predictions deal primarily with semantic comprehension, an attempt was made to obtain also a measure of syntactic comprehension. For this purpose the rendering of the sentence by the subject was categorized as being *identical in content* with one of the following:

(i) the original sentence;
(ii) a sentence of the same *structure* as the original sentence, but not having the same content;
(iii) a sentence differing from the original sentence in structure as well as in content.

To illustrate, take the following (fictitious) examples of the way a subject might render the content of the C_c sentence given above (6.3.1):

(i) This is the rat which made the hole. The rat was caught by the cat which was bit by the dog.

(ii) This is the mouse which ate the cheese. The mouse was bit by the cat, which was bit by the dog.

(iii) This is the rat which made the hole. The rat was caught by the cat and also bit by the dog.

The *content* of the sentence can be said to be understood only in example (i). Both in (ii) and in (iii) the content is misunderstood; but there is an important difference: Example (ii) can be *reformulated* in the form of a sentence in dn 2:

> This is the cheese that the mouse, which the cat, whom the dog bit, bit, ate.

The latter sentence, which says the same as (ii), differs in meaning from the original C_c-sentence, but has the same structure, in that both are in dn 2. In contrast, (iii) can not be reformulated in dn 2 ("the rat" being the object of both "caught" and "bit"), and it differs, thus, from the original sentence in both content and structure.

Rendering the original sentence as in (i) will, therefore, be taken as evidence of *semantic comprehension*, and rendering it as in (ii) – as evidence of *syntactic comprehension.* Semantic comprehension, as defined here, presupposes syntactic comprehension.

An obvious difficulty with the above categorization is the absence of an operational definition of "identical in content". The exigencies of the experiment were such, that the decision in each case of what is and what is not to be regarded as identical in content to the original sentence had to be left to the experimenter, with the understanding that the same rule-of-thumb be employed as in the previous expeiment (6.3.2). A recheck of the experimenter's decision on the basis of the full protocol of the experimental sessions revealed that, in fact, the cases were clearcut, and the experimenter's decisions could be upheld.

(b) *Results pertaining to prediction 1*

Table 6.1 shows the number of subjects who understood each kind of sentence on the first reading and on the fourth reading, as well as the mean number of readings required for each type of sentence. According to all three measures – which are given here for syntactic as well as for semantic comprehension – C_c-2 was easier to understand than N-2.

TABLE 6.1. *Number of Readings Required for Sentences of Different Kinds*

Comprehension	Sentence Type	No. of Subjects* Comprehending on First Reading		No. of Subjects* Comprehending on Fourth Reading		Mean Number of Readings to Comprehension	
		dn 0	dn 2	dn 0	dn 2	dn 0	dn 2
Syntactic	N	8	3	16	13	1.63	2.08**
	C_c	15	12	16	16	1.06	1.38
Semantic	N	5	1	16	9	1.94	3.22**
	C_c	14	12	16	16	1.13	1.38

* Total number of subjects – 16.
** For those subjects who comprehended the sentence on the fourth reading at the latest.

C_c-2 sentences required a smaller number of readings for syntactic understanding than N-2 sentences for 11 out of the 16 subjects; only two subjects required more readings of C_c-2 sentences, with the four remaining subjects requiring the same number of readings for both types. The difference is significant by a sign-test ($p = 0.011$). The differences for semantic comprehension are even more marked.[12] Prediction 1 is, thus, confirmed.

[12] The significance test was carried out for syntactic comprehension, because of the argument raised in note 9, above; it should be noted that the differences for semantic comprehension were even greater. An alternative measure which might have been used is time required (over all readings) till the sentence is understood. This is a more sensitive measure, and for some subjects who required an equal number of readings (for syntactic comprehension) for C_c-2 and N-2, there was a difference in reading time in the expected direction. Difference in reading time would also have been significant at the 1 per cent level.

(c) *Results pertaining to prediction 3*

According to this prediction, the difference in ease of comprehension should be greater between N-2 and N-0 than between C_c-2 and C_c-0. This was the case for all three measures of comprehension given in Table 6.1. The difference in number of readings required for semantic comprehension was in the predicted direction for nine subjects, and in the opposite direction for only two (with equal differences being obtained for the remaining five subjects) and the result is, thus, statistically significant by a sign test ($p = 0.033$; one tailed).[13] It can therefore be concluded that the size of the effect of nesting is determined by the absence or presence of semantic cues.

6.3.6. *Observations Pertaining to the Nature of Syntactic Decoding*

Semantic-syntactic decoding can hardly have been applied to N-sentences, since these lack semantic cues. Such sentences were therefore assumed to be syntactically decoded (6.3.1), and the results of this experiment have borne this out. The fact that some (though few) subjects did comprehend an N-2 sentence on first reading (cf. Table 6.1) seems to suggest that their memory was operating along the lines of a PDS (cf. 5.1.1, 5.2.2).[14] But several observations made in the course of the experiment serve to throw some doubt on such a conclusion. Thus, some subjects reported after the experiment that they were aware of the fact that the first verb is the predicate of the last subject and the last verb – of the first subject (in sentences of dn 2). With the help of such a rule, subjects may have decoded the sentences without a PDS by simply

[13] The difference would also have been significant ($p = 0.003$) for syntactic comprehension and for time required till syntactic comprehension was achieved ($p = 0.004$). In this experiment, C_c-2 sentences required a greater number of readings than C_c-0 sentences (which is quite compatible with the hypothesized semantic-syntactic decoding process, cf. 6.2.2, a), but, for semantic understanding this difference was not statistically significant.

[14] Cases where the sentence was understood only after repeated readings ought to be disregarded for the purpose of this discussion, since there are good grounds for looking upon such repeated readings as providing additional aid to memory (cf. 5.2.3).

labelling the parts of the sentence ("first verb, last verb, first subject, last subject") and storing them with their labels.[15]

Yet another way which may have helped subjects in dealing with the nested sentences is repeating the sentence after reading it. Some subjects reported doing so. Apparently, it was easier for them to recall the sentence verbatim (or nearly so) than to decode it syntactically. Now, repeating the sentence, like rereading it (cf. 5.2.3) may have provided additional aid to memory.

There are also a priori grounds for questioning the operation of a PDS mechanism. N-sentences are very infrequent, and so are sentences of a degree of nesting greater than 1 (cf. 5.3.5). A PDS is required, if at all, only for sentences combining both these characteristics, such as the N-2 sentences of our experiment. It seems unlikely that a complex process should have been incorporated in the human behavior repertoire for which there is practically no use whatsoever (except for serving as subject in experiments on degree of nesting).

6.4. SUMMARY AND CONCLUSIONS

The current assumption that syntactic decoding proceeds in isolation from semantic factors has been questioned in this chapter. This assumption was implicitly made in formulating the syntactic decoding hypothesis (5.2.2), which failed to receive substantial experimental support (5.3.2, 5.3.4, 5.4.2). These negative results, along with some additional experimental findings (5.4.3), led to the hypothesis of a semantic-syntactic decoding process, according to which use is made of the partial redundancy of syntactic structure (6.2.1), at least in those cases where the syntactical decoding is beset with difficulties (6.2.3), as is the case with deeply nested constructions.

[15] In the second experiment it was attempted to minimize the chances of subjects to find out about this rule by presenting the practice sentences in dn 0 or in dn 1 (i.e., there was no first and last verb belonging to the last and first subject, respectively). However, previous experience and exposure to training in syntax may have facilitated the discovery of such a rule in the course of the experiment.

Support for this notion came from experiments in which it was shown: (a) that nested sentences were much easier to decode if they contained semantic cues by means of which the corresponding parts (*a* and *a*', *b* and *b*') could be correctly paired (6.3.3, 6.3.5); (b) that these semantic cues might lead to construing the sentence in a way incompatible with its syntactic structure (C_i-sentences of 6.3.3); and (c) that degree of nesting has a larger effect on the comprehensibility of sentences in which no semantic cues for decoding are supplied (N-sentences) than on that of sentences which contain such cues (C_c-sentences; see 6.3.5). None of these effects would have been obtained, had syntactic decoding proceeded without any influence of semantic factors.

This conclusion regarding the operation of semantic factors in decoding is in line with those arrived at by other research workers (see 3.1.1; cf. also 3.4.2; Miller and Isard, 1963; and the paper by Jenkins and Palermo in Bellugi and Brown, 1964, pp. 165-168).

This chapter has not brought us any nearer to the solution of the problem what makes for (behaviorally relevant) complexity of sentence structure. N-sentences, which show a clear effect of nesting, are the exception rather than the rule, and for the kind of sentences which one is likely to encounter, the effect of nesting is either small or non-existent (cf. 5.3.5). If nesting fails to make reading difficult, one might well ask, *what* does. The massive evidence from readability studies for a sentence-structure factor in addition to a word factor in reading difficulty (cf. 1.4) does not permit of any doubt as to the effectiveness of such a factor. The conclusion seems to be forced upon us that the much decried "complexity" of sentences which makes reading difficult, is, in part at least, a function of *content* (cf. also 5.3.5, note 12). A special case of an interaction of semantic and syntactic factors in producing sentence complexity is being discussed in the next chapter.

7

SENTENCES BEGINNING WITH SEMANTICALLY INDETERMINATE WORDS

The preceding chapters dealt with various syntactic variables. Predictions as to the effects of these were based on the metahypothesis which states that the human user of the language incorporates a device analogous to a grammar of the type developed by, e.g., Chomsky (1.2). The argument presented in Chapter Six implies that in some cases it may be unwarranted to base behavioral hypotheses on this metahypothesis. Some of the current linguistic models, which exclude semantic considerations (6.1), may not be directly translatable into a description of the behavior of the decoder, because it is these semantic factors which play an important role in the comprehension of syntactic structure (6.2.1).

The hypothesis investigated in the present chapter bears no such relationship to grammatical theory. It is based on considerations about the operation of memory in decoding a sentence, but takes into account both syntactic and semantic factors. (Cf. also 4.1.3 for similar considerations). It pertains to a stylistic factor which appears to affect the difficulty of the sentence, without making the sentence more "involved", in the usual sense of the term.

7.1. THE PROBLEM

Consider the following sentence:

> After addressing the House of Commons, Mr. Churchill left for a meeting with the Chief of Staff.

The subject of "addressing" remains unknown to the reader till after he has read the opening phrase; he has to store this phrase in his immediate memory till the subject of the sentence ("Mr.

Churchill") appears. In the words of Hebb and Bindra (1952): "The response to one of the stimuli (words) arriving at any moment may be quite indeterminate and become determinate only after the arrival of another stimulus".[1]

Now, if the non-finite verb form is not too far removed from its subject – as is the case in the above example – the effect of this stylistic factor on reading ease will probably be quite negligible. The problem arises when the subject is farther removed; then the additional demand on the reader's memory may conceivably make the sentence more difficult and it may prove to be advisable to rewrite the sentence with the subject preceding the verb.[2] For the above example, this would be:

> After Mr. Churchill addressed the House of Commons, he left for a meeting with the Chief of Staff.

Note that the sentence is *not* indeterminate in terms of *grammar*; what is in question here is the effect on the decoder's performance of his having the store information of a *semantic* nature. In contrast, it was assumed in the case of nested sentences that *syntactic* information (the sequence of sentence "parts") has to be stored (5.2.1).

The question might be raised what linguistic function does this

[1] Hebb and Bindra exemplify this with a somewhat different type of sentence:
Whether it is large or small, you should
always keep your dog on the leash.
Here the pronoun stands before the noun to which it refers.

[2] The use of Hebrew sentences in which the verb form precedes the subject has been condemned by some authorities as being due to the influence of foreign languages. While such normative statements are outside the scope of a psycholinguistic study, we may take issue with one of the arguments advanced for them. Isaac Epstein (1947, p. 273) claims that the indeterminate phrase at the beginning of the sentence rouses the reader's curiosity and that "every tension requires immediate relief". Therefore, says Epstein, the subject of the sentence must be maximally close to the predicate. This argument leaves the question open, why the speaker of English should have a higher tolerance for "tension" than the speaker of Hebrew, as evidenced by the frequent use of such sentences in the English language. Note also that the factor of curiosity and tension may actually be an argument in favor of this sentence structure; cf. the use of a deviant style in poetry (Riffaterre, 1959; Fônagy, 1961). Should empirical evidence be obtained that this structure reduces readability, an explanation in terms of memory load appears to be more plausible.

potentially difficult stylistic form fulfil. One function may be that of providing stylistic diversity. In addition, this form may serve to avoid *ambiguity*, as can be seen when the sequence of events of the above two sentences is reversed, as in the following:

(1) After meeting the Chief of Staff, Mr. Churchill left for the House of Commons.
(2) After Mr. Churchill met the Chief of Staff, he left for the House of Commons.

The ambiguity inherent in (2) (Who left?) is resolved by the alternative formulation (1). This may perhaps serve to explain the development of the alternative structure.

Still another way of resolving the ambiguity is by repeating the subject. This is usually done by paraphrasing it (probably, for esthetic reasons). Thus, instead of (2) one might write:

(3) After Mr. Churchill met the Chief of Staff, the Prime Minister left for the House of Commons.

Here there is neither the ambiguity of (2) nor the need to keep a phrase of the sentence in memory, as in (1). Thus, if one is willing to pay the price of adding a few words (and of looking for a suitable paraphrase for "Mr. Churchill", and perhaps also of risking that the hearer does not know that the Prime Minister is Mr. Churchill), one can avoid both these pitfalls.

The need of keeping in memory semantically "indeterminate" words leads to the hypothesis that, when the number of intervening words is sufficiently large, form (1) will be more difficult than either form (2) or form (3) (when no ambiguity is involved). In the next section a relevant experiment is described.

7.2. AN EXPERIMENT WITH PRONOUNS STANDING BEFORE THE NOUNS TO WHICH THEY REFER

7.2.1. *Method*

In the experiment to be described here, Hebrew sentences were used. The semantically indeterminate form – which will be referred to as form (1) – includes a pronoun standing before the noun to which it refers. (E.g., instead of (1) in section 7.1: "After he met the Chief of Staff, Mr. Churchill ... ". Cf. also note 1, above). The pronoun was removed from the noun much farther than the indeterminate phrase in the examples of section 7.1. This form was compared to form (2), in which the noun came first, and to form (3), in which the subject was repeated in paraphrase.

Eight sentences were each formulated in three parallel forms, (1), (2), and (3). Care was taken to avoid ambiguities of the kind discussed above. Sixty subjects were assigned randomly to three experimental groups, and each group read the eight sentences in one of these forms. Only relatively slow readers were included, for reasons similar to those discussed in a previous experiment (4.1.3, 4.2.1). Subjects reading at a rate of not more than 267 words per minute were defined as slow readers.

In order to make sure that the subject read the sentences with understanding, he was required to judge each sentence as to whether it was "logical" or not. Three "illogical", i.e., incoherent, sentences were interspersed among the eight experimental sentences (cf. 5.3.1 for a similar procedure). Time taken to read each sentence was measured by a stopwatch.

7.2.2. *Results*

For each subject and for each experimental sentence an index of response was computed which was the ratio between the time taken to read a criterion passage – which was the same for all groups of subjects – and the time taken to read the experimental sentence. This measure controls for individual differences in reading rate (cf. 4.2.2). Each subject was then assigned a score, which was the sum of these indexes of response for eight experimental sentences, where a larger score indicated a faster average reading rate.

The average score of the group of twenty subjects reading form (1) was smaller than that of either one of the other two groups, as predicted. Moreover, a comparison of the average indexes of response for each one of the eight sentences showed that for each sentence, form (1) took longer to read. However, the differences between the scores of the three groups failed to reach statistical significance (by a Kruskall-Wallis one-way analysis of variance [Siegel, 1956]; $H = 3.33$; $.20 > p > .10$). Thus, though the results are in the expected direction, further studies will be required to provide substantial support for the hypothesis.

As to the difference between form (2) and form (3), there were four sentences in which form (2) took longer to read, whereas for the remaining four sentences form (3) took longer. It occurred to us that these differences between sentences might be related to the number of words intervening between the two occurrences of the subject, or of the subject and the pronoun (in the examples of 7.1: between "Mr. Churchill" and "he" in (2), and between "Mr. Churchill" and "the Prime Minister" in (3)). Conceivably, the repetition of the subject in (3) is an advantage when the number of intervening words is large, so that the subject is helped by being reminded what the subject was, whereas when this number is small, the repetition serves no useful purpose. The data provide some indication of the validity of this conjecture: The average number of intervening words was 17.2 for those sentences where form (3) took longer to read, and 24 for those where form (2) took longer. But the small number of sentences precludes a definitive test of our explanation.

7.3. SUMMARY

Reading performance may be assumed to be affected by the requirements made on the decoder's memory. During the process of decoding two kinds of information may have to be remembered, information about the structure of the sentence and information of a semantic nature. When linguistic models are taken as a starting point, certain hypotheses can be formulated as to the kinds of sentences which demand an unusual amount of syntactic information to be stored in immediate memory; such was the hypothesis about the effect of nesting (5.2.1, 5.2.2). In contrast, we have only common-sense considerations to go by in predicting difficulties due to the need to remember semantic information.

In this chapter such considerations have been discussed (7.1). Furthermore, an experiment was described which aimed at comparing the difficulty of three kinds of sentences, differing in the demands made, presumably, on the reader's immediate memory. The results of this experiment have not been conclusive (7.2).

8

SUGGESTIONS FOR FUTURE RESEARCH IN READABILITY

One of the main themes of this study has been the relevance of modern linguistic theories to psycholinguistic research. Most of the hypotheses which were put to an experimental test were based on a metatheory linking linguistic models to linguistic behavior (1.2). It will have become clear in the preceding chapters that only a beginning has been made in pursuing this line of research and that many questions have remained unanswered, and often even unasked.

While the main orientation of the study has been a theoretical one, our experiments usually dealt also with a practical problem: the influence of various syntactic factors on the readability of texts. This applied field of readability will be the concern of the present final chapter. Some of the lessons which seem to emerge from the present study will be presented and some suggestions will be offered regarding future research in readability. The first section (8.1) contains remarks on research methodology, and in the second (8.2), a general outline will be presented of a proposed area of research.

8.1. SOME REMARKS ABOUT RESEARCH METHODOLOGY

8.1.1. *The Experimental Approach*

In the introductory chapter it was indicated that the methodology of the present study differs from that of classical readability research (1.4). Instead of "fishing for correlations" between readability and various easily countable attributes of a text, attempts were made to manipulate linguistic variables experimentally so that their effect on the reader's performance could be measured. The difficulties attendant upon this approach are amply illustrated by the preceding chapters. Nonetheless, the research described there seems to justify the conclusion that the experimental approach is both feasible and fruitful.

It is suggested, then, that future research should concentrate on a systematic exploration of variables by experimental manipulation, rather than follow the well-worn path of analyzing texts in order to devise bigger and better formulas. As stated before (1.4), the latter approach seems to hold little promise for further progress in this area.

8.1.2. *Readability and the Reading Process*

In our search for behavioral correlates of syntactic variables we often did not content ourselves with customary measures of reading difficulty. Our objective was to study the reading process in general; hence, techniques were employed which prima facie do not serve to measure readability. Ultimately, of course, a better understanding of the reading process may be expected to contribute to the theory of readability. Hence, it is suggested that a more theoretical orientation will, in the long run, be more fruitful for readability research than the pursuit of short-range applied goals. This implies a wider conceptual framework for research (see below, 8.2), as well as a greater variety of research techniques.

The techniques developed in this and in other studies have been described in various places (e.g., 2.1 - 2.2, 3.1 - 3.4, 5.3.1, 5.4.1, 6.3.2, 6.3.4, 7.2.1). Needless to say, these still require further refinement. The usefulness of more customary techniques is, of course, far from being exhausted; measures of comprehension and reading rate, for instance, will probably continue to be useful. A few comments concerning the use of these measures will therefore be in order here.

8.1.3. *Reading Rate*

This measure has been employed in the present study, as well as in various other readability studies (e.g., Klare *et al.*, 1955). Our experience shows that individual differences in reading rate are quite large even in relatively homogeneous samples like college students.[1] In comparing the readability of different texts by means of this measure, these differences ought to be controlled for by analysis-of-covariance or by using an index of response (cf. 4.2.2). An additional difficulty consists in the great intraindividual variability of reading rates.[2] To overcome this difficulty, at least in part, the subject should be familiarized with the experimental task by means of a practice passage (cf. 4.2.1).

Reading rate seems to be a very sensitive measure. While it is most readily employed in individual tests, this measure can also be used in group experiments (cf. 5.3.1, 5.3.3). A factor which ought to be taken into account in employing this criterion in readability, research is the lack of flexibility found for some readers by Letson (1958; but conflicting findings have been obtained by others, e.g., Tinker, 1939).

A new technique involving reading rate has been suggested by E. B. Coleman (personal communication). The linguistic material consists of instructions for a task which the subject is called on to perform, (e.g., doing a particular problem in mental arithmetic, drawing a particular figure), and the measure is the amount of time the instructions must be presented for the subject to perform the task correctly.

8.1.4. *Comprehension and Recall Tests*

Comprehension tests are easily administered to groups of subjects, and scores on these tests have usually been found to correlate

[1] Data which have been kindly supplied by the Institute for Productivity of the Israel Ministry of Labor show that the range of reading rates of participants in a course for reading efficiency given several years ago was between 74 and 388 words per minute at the beginning of the course. The variability was even greater at the end of that course: 76 to 777 words per minute.

[2] See 5.3.1 note 8. Robinson and Hall (1941) found reading rates within an experimental text to vary also with type of content.

positively with reading rates.[3] To ensure reliable scoring of answers, multiple-choice tests are employed.[4] Since the number of possible scores is the number of questions asked plus one (for "no correct answer"), and the number of questions which it is feasible to construct for a short text is rather limited, the comprehension score will usually be found to provide a less sensitive measure than reading rate.

The concept of comprehension which is measured in these tests is still in need of clarification. Several studies have shown that different questions of a test do not measure the same "thing" (Davis, 1944; Thurstone's reanalysis of his data, 1946; Derrick, 1953; Vernon, 1962). An overall score on such a test may perhaps furnish the teacher with a rough-and-ready yardstick telling him "how much" has been understood, but for education as well as for purposes of research the question "how much of what?" is also of interest. This information is usually not revealed by the total score.[5] A possible approach to testing comprehension is by devising different *kinds* of questions and assigning scores to each kind. For defining these kinds of questions, a conceptual analysis of reading comprehension seems indicated. Taking questions actually appearing in comprehension tests as a starting point, it will be found that some questions ask about what is stated in the text *explicitly*, while others ask about what is *implied* in the text (cf. Harris, 1948). Further, the answer to a given question may require combining information presented in different places of the text, or it may be found at one place (which may be a sentence, or two consecutive sentences, or a whole paragraph). The above two sets of categories, which are used here for illustrative purposes, allow

[3] Cf., e.g. Ledbetter (1947), but see Stroud and Henderson (1943). The size of the correlations has been found to vary with kind of reader (Carlson, 1949; Robinson and Hall 1941); and kind of reading material (Tinker, 1939; Robinson, 1940; Robinson and Hall, 1941, Letson, 1958).

[4] A suggestion for a different measure has been mentioned in 3.1.2, note 9. Poulton (1960) has suggested rate of comprehension as a measure.

[5] Only when items are constructed so as to measure only one aspect of comprehension, and the resulting test constitutes a Guttman scale, can the answers to individual items of the test be reconstructed from the total score.

of four combinations which define four types of questions, and further distinctions may be developed.[6]

Recall tests have occasionally been employed in readability research (e.g., Klare *et al.*, 1954; Wason, 1962). These may perhaps be looked upon as a variant of comprehension tests, because the quality of recall will be influenced by the degree of comprehension achieved. The measure of comprehension may be number of words recalled, the judgment of an observer that recall is correct, or the number of readings required till correct recall is achieved (cf. 6.3.2, 6.3.4). While recall tests seem to be especially suitable for short texts, they may be used with longer texts; recall of content rather than verbatim rendering of the text is then expected of the subject. The text is sometimes divided into "items" for each of which the correctness of recall is assessed (cf. Levitt's methodological study of this method, 1956).

8.1.5. *The "Cloze" Procedure*

The "cloze" procedure is a newcomer in readability research. Its limitations have apparently not always been recognized, and it will be well worth, therefore, to discuss it here at some length. Developed by Taylor (1953, 1956, 1957), the procedure consists in deleting either *1/n* of the total number of words in a text (determining randomly which words are to be deleted) or every *n*-th word in it, and letting the subject supply the missing words. In readability research this technique may be used in two ways:

(i) As a comprehension test: The subject reads a text and then is given the same text with part of the words deleted. The number of words correctly supplied by him serves as a measure of comprehension.[7]

[6] Research along these lines has been carried out at the Israel Institute of Applied Social Research, Jerusalem, under Louis Guttman. Such "facet design" (cf. Guttman, 1965), has been applied also to the construction of alternative answers in the multiple-choice test, and may be expected ultimately to lead to a prediction of interrelationships between different subtests.

[7] If the subject supplies a synonym of the deleted word or any word which does not change the meaning of the sentence, he ought perhaps to be credited

(ii) As a measure of readability: Subjects are asked to fill in the missing words of a text which they have *not* read before. The number of correct responses is a function of the redundancy of the text, and redundancy may be assumed to be related to reading ease.[8] In general, therefore, the "cloze" score will rise with readability. However, redundancy of the text is not the only factor to influence readability; word length and semantic factors, for instance, also appear to affect reading difficulty, and it is not quite clear to what extent these influence the "cloze" score.[9] Nevertheless, it has been claimed that "cloze" scores provide a good overall measure of readability.[10]

"Cloze" tests are easy to construct and to score. They are well suited for testing groups of subjects. There is no problem as to what kind of comprehension is measured when the technique is

---

with a correct substitution. But this would greatly complicate the process of scoring and will usually not be worth while, since it is to be expected that the score for literal substitutions will be monotonically related to the score which takes into account synonyms. For one of the few studies which use the latter score see Morrison and Brown (1957).

[8] Tachistoscopic recognition thresholds for verbal material and intelligibility of words have been found to be influenced by redundancy (e.g., Bruce, 1958; Frick, 1953; Howes, 1957; Howes and Solomon, 1951; Miller *et al.*, 1951; King-Ellison and Jenkins, 1954). Preston (1935) has shown speed of word perception to be related to speed of reading. For a more direct investigation of the relationship between reading rate and amount of information see Pierce and Karlin (1957).

[9] Word length has been shown to affect recognition thresholds in interaction with word frequency (Mc Ginnies *et al.*, 1952; Newbigging, 1961). Pierce and Karlin (1957) found word length, in addition to word frequency, to affect reading speed. Cf. also Rosenzweig Postman (1958). The role of semantic factors in readability still remains to be investigated (see also 7.1, above). One factor, isolated by Bloomer (1961), is the "concreteness" of words, which was found to correlate with reading and spelling ease. The relationship between the syntactic factors discussed in previous chapters and redundancy is still in need of clarification; cf. 1.3.

[10] Cf. Taylor (1956, 1957); Osgood and Sebeok (1954, p. 112), and also section 1.4 note 5. Chapanis (1954) conducted a study of the redundancy of English by means of letter deletion, and suggested that the relationship between redundancy and readability need not always hold. In another study (Rubenstein and Aborn, 1958), redundancy was measured by letting subjects predict successive words in a text, and a high correlation with readability scores (by a formula) was found.

used as a comprehension test (cf. 8.1.4). All this may tempt one to adopt "cloze" procedure as the standard technique of readability research. However, its indiscriminate use must be warned against, as will be shown in the following.

Suppose the researcher addresses himself to the problem of the effect of some linguistic variable that involves the word order of the sentence (such as is the case for most variables discussed in the present study). He proceeds to construct parallel versions of the text differing in this variable so as to compare their "cloze" scores. The *same* words must be deleted in all versions; otherwise the "cloze" score will reflect not the effect of the variable in question, but the difficulty of the specific words deleted. Now, as a result of the different word order in the parallel versions, it will be the case that (a) the number of words intervening between two deletions in one version will often not be the same as the number of words for the corresponding two deletions in the parallel version; and (b) corresponding deletions will often appear in the environment of different words in the parallel versions. Each one of these facts may contribute to the deletions being more difficult to fill in for one of the versions.

By way of example, let us look at the following two versions of a text:

(1) After the two friends had eaten their meal, they set out to explore the neighboring forest.
(2) The two friends, after they had eaten their meal, set out to explore the neighboring forest.

To investigate the effect of the different sentence structure by means of a comprehension test utilizing the "cloze" procedure one might decide to delete, say every second word of (1):

(1a) After _____ two _____ had _____ their _____, they _____ out _____ explore _____ neighboring _____.

Deleting the *same* words in (2), we obtain:

(2a) _____ two _____, after they had _____ their _____, _____ out _____ explore _____ neighboring _____.

The deletions are differently spaced in (1a) and in (2a), and this in itself may affect the ability of subjects to supply the correct words (though in the above example this is perhaps not intuitively so). Also, " _____ out" may be easier to complete when occurring after "they", than after "their _____", (and, again, it is inessential whether or not this holds true for this specific example). The matter is even more serious, if the two versions do not contain exactly the same words, as may be the case when the effect of sentence structure is under study (cf. 4.1.2, note 3). Essentially the same difficulties are attendant upon the use of "cloze" score as a direct test of readability.

The effects of such small differences in the location of deletions might perhaps be assumed to cancel each other when extensive texts involving a great number of deletions are employed. But wherever there is a question of studying the fine-grain structure of a sentence, there arises the suspicion that the factors discussed above may somehow have a systematic effect on the "cloze" score. Unless it can be shown with reasonable certainty that this is not the case, the use of the "cloze" technique had best be avoided.

8.1.6. *Efficiency of Eye-Movements*

An efficient reader reads several words at a glance. Photographic recordings of eye-movements reveal that his eyes make fewer fixations and fewer regressions (reversals to words occurring earlier) per line than those of slow and inefficient readers. Readability also influences eye-movements: For a given reader the number of fixations and regressions increases with difficulty of reading material (Carmichael and Dearborn, 1947; Tinker, 1958). The use of these measures in readability research is rendered difficult by the need for elaborate recording instruments, (cf. Zachrisson, 1965, pp. 54-55) which preclude, moreover, the use of group tests. Therefore only few readability studies have employed eye-movements as a measure (cf. Ledbetter, 1947; Klare *et al.*, 1957).

An indirect way of assessing the efficiency of eye-movements is by means of the eye-voice span (cf. 2.2.3). In some studies the size of the span was found to be affected by the order of approximation to English (2.2.1), and it remains to be seen to what extent this measure will be useful in readability research.[11]

8.1.7. *Judgments of Readability*

Several studies have obtained the potential reader's judgments as to the readability of a given text (e.g., McCracken, 1959). Others have turned to such authorities as teachers, editors and librarians for ratings of readability (Chall, 1958, pp. 10-12; Nahshon, 1957). The relationship of these ratings to more objective criteria of reading difficulty remains to be investigated (cf. also McCracken, 1959). Conceivably, the ratings may be influenced by prevailing stereotypes as to what makes a text readable. Presumably, they will consider highly nested sentences, such as the one in figure 5.6 (see section 5.3.1) to be quite unreadable, whereas experimental results show that this is far from being the case (5.3.2, 5.3.4, 5.4.2).

But the subjectively felt difficulty of texts may be of interest no less than the objectively measured difficulty. It is the former which determines to a large extent "readability" in the sense of acceptability of the material, and its popularity (Klare *et al.*, 1954; Bernstein, 1955-56).

8.1.8. *Measurement of Effort*

In discussing our experiments on the effect of nesting, the results of which were mostly negative, the possibility was pointed out that the linguistic variable affects not overt performance, as measured by reading rates and comprehension scores, but the amount of effort which the reader invests to make up for the increase in difficulty (5.3.5). Such an explanation has been offered by Klare

[11] In a methodological study on the eye-voice span (measured as in 2.2.3) for texts with degrees of nesting (cf. 5.1.3) 0, 1 and 3, no effect of this variable on the length of the span could be demonstrated. It will be remembered, that reading rate is also little affected by degree of nesting (5.3.2, 5.3.4).

et al. (1954) for his unexpected finding that subjects recalled less of the easier version of his experimental text. To follow up this suggestion, use would have to be made of reliable measures of effort. While no such technique has yet been perfected, recent studies indicate that electromyographic recordings of muscle potential may prove useful for this purpose (Smith *et al.*, 1954; Wallerstein, 1954; Edfeldt, 1959; Keller and McGuigan, 1963; cf. also Bélanger, 1957).

Involuntary blinking has been assumed to be an index of effort, and blink rate has been used in legibility research by Luckiesh and Moss (1940) and by Carmichael and Dearborn (1947) to measure fatigue. Not all investigators agree that this measure is a valid one (Bitterman, 1945; Tinker, 1946; Wood and Bitterman, 1950). It remains to be seen whether this technique can be adapted to the study of readability, and in particular, with short stretches of text.[12]

Conceivably, further investigations may show that the additional effort required in reading supposedly difficult material is so small that it can not be measured reliably. If so, it is possible that under particularly difficult reading conditions the effect of effort will be more pronounced. (Cf. 5.3.5, note 13, where possibilities of increasing the overload of the system have been outlined).

8.2. A FRAMEWORK FOR RESEARCH

The goal of most readability research has been an applied one. This has manifested itself not only in research methodology (cf. 1.1) but also in the way this area has been outlined. "Readability", by definition, refers to (a) written text and not orally presented material, and (b) decoding behavior as contrasted with encoding

[12] Negative results were obtained in an experiment on the effect of nesting on blink rate (5.3.4, note 10). It seems possible that with sentences of complex structure blink rate will be largely a function of the number of eye-reversals. Thus it has been found that in visual-motor tasks blinking occurs shortly before and after performance, and that during performance it tends to occur at moments when least attention is required (Drew, 1951; Poulton and Gregory, 1952). If this conjecture proves correct, it would seem more expedient to study eye-reversals directly.

behavior; further, (c) research has been carried out almost exclusively with normal populations. It is suggested that these limitations – adopted, no doubt, for practical reasons – are unnecessary, and that a more comprehensively defined area of research may be expected to lead to an integration of results and a theoretically more meaningful picture. The following remarks will serve to elaborate this point.

(a) In addition to readability, "listenability" has been studied by some investigators (see Klare's review, 1963, p. 148f.; see also Larsen and Feder, 1940, and Harwood, 1955). In the present study it has occasionally been found expedient to deal with auditorily presented material (e.g., 3.1.1, 3.3.2, 3.4.1). Studies of readability and of listenability ought to be mutually illuminating.

(b) The educator is primarily interested in decoding behavior, and reading in particular. But theoretically, there seems to be no virtue in divorcing the investigation of decoding from that of encoding. Discussion of encoding behavior has repeatedly been resorted to in this study (e.g. 3.2, 3.5, 6.2.4) in dealing with theoretical issues.

(c) Grade school pupils, and sometimes college students, have served as subjects in most readability studies. It appears that research with subjects suffering from various language and speech disorders may furnish valuable insights into linguistic processes.[13] But so far there have been only tenuous connections between such studies and readability research.

In short, readability research has much to gain by integration within the wider framework of what may be termed *communicability* research. This term will be defined here by a so-called faceted definition (cf. Guttman, 1965) as follows:

[13] Recent research on aphasia (Osgood and Miron, 1963, pp. 88-142) appears to have much relevance for an explanation of linguistic processes in the hearer and speaker. Furthermore, some recent investigations indicate that the frequency of stuttering is determined, inter alia, by linguistic factors such as word frequency, information load and sentence position (Quarrington *et al.*, 1962; Conway and Quarrington, 1963; Quarrington, 1965; Schlesinger *et al.*, 1965). Another example is a case-study of language disorder by Lenneberg (1953/54) which has important implications for psycholinguistic theory. Cf. also 3.5.

Communicability is

$$\text{the} \begin{Bmatrix}\text{ease}\\ \text{readiness}\end{Bmatrix} \text{with which linguistic material}$$

$$\text{in} \begin{Bmatrix}\text{written}\\ \text{spoken}\end{Bmatrix} \text{form with (given)} \begin{Bmatrix}\text{cognitive}\\ \text{emotional}\end{Bmatrix} \text{characteristics}$$

$$\text{of} \begin{Bmatrix}\text{content}\\ \text{style}\end{Bmatrix} \text{is} \begin{Bmatrix}\text{decoded}\\ \text{encoded}\end{Bmatrix} \text{by members of a (given) population.}$$

Sub-areas of communicability are defined by different combinations of elements of the above facets. Thus, listenability is a sub-area of communicability: it pertains to the decoding of linguistic material in spoken form; and "the ease with which linguistic material in spoken form is encoded ... " refers to studies of verbal fluency. A few remarks explaining the definition will be in order here:

(a) "Ease" and "readiness" are two aspects which readability research has been concerned with (cf. 1.1, note 1). Studies of reading interests and reading preferences are included in the above definition under "readiness with which linguistic material in written form ... is decoded ... " "Ease" and "readiness" are the dependent variables of communicability research.

(b) The characteristics of linguistic material are the determinants of communicability, and fall under four categories.

> (1) "Cognitive characteristics of content": Some kinds of content are more difficult to read about (or hear about, or talk about) than others, and may, or may not, be approached with less readiness by a given reader (or listener, or talker, respectively).
>
> (2) "Emotional characteristics of content": Some topics may be interesting, emotionally arousing and pleasant, while others may be unpleasant, embarrassing, or threatening. These attributes of the content of linguistic material might be related not only to the "readiness" with which it is decoded (or encoded), but also to the "ease" with which this is done (as is shown by investigations of perceptual defence).

(3) "Cognitive characteristics of style": Most readability studies, the present one included, have dealt with the effect of these.[14] Studies of written and spoken style deal mainly with cognitive characteristics, such as sentence length and type-token ratio (see, e.g., the review by Pool, 1959).

(4) "Emotional characteristics of style": A study by Osgood and Walker (1959) on the effect of drive state on the style of encoding is an example of research falling under this heading.

(c) The last phrase of the definition includes a facet written in brackets, the elements of which have not been enumerated. Here we refer to studies of different populations, which, as stated above, may contribute to our understanding of linguistic processes.

The framework presented here is meant to include the effects of any attribute of linguistic material on communicative behavior at the "semantic level" (cf. Shannon and Weaver, 1949, p. 114). The "technical level" – which includes problems of legibility, audibility, and the like – and problems of the effectiveness of communication remain outside its scope, because it is felt that, at the present state of our knowledge, there is little which research in the latter two areas can contribute to the theory of communicability as defined in the above.

[14] Many variables remain to be investigated. Thus it has been argued that static sentences in which the main clause precedes the subordinate clause will be easier to read for the slow reader than dynamic sentences because in the former the reader gets a "meaningful concept" at an earlier stage (Wonderly no date). The same writer suggests that the head-plus-modifier construction (*maison grande*) may be easier for the inexperienced reader than the modifier-plus-head construction (*grande maison*) because in the former the transition probabilities after the first word are higher. Rothkopf (1963) has recently compared the effect on recall of different word orders, as exemplified by: *The Lanes are the ghosts who protect a family shrine* vs. *The ghosts who protect the family shrine are the Lanes.* His results were inconclusive, but this and other factors deserve further study. Only a beginning has been made in assessing the effects of sentence complexity; other measures, than those used in the present study such as the node-to-terminal-node ratio suggested by Miller and Chomsky (1963, p. 480), might be fruitful here. Finally the interesting problem of syntactic ambiguity (Chomsky, 1957, pp. 81-90) has so far received only little attention from psycholinguists; to the best of my knowledge only one study has been conducted in this area (Marshall, unpublished).

In contrast, there are many theoretical connections between different subareas of communicability research, between listening and reading, between encoding and decoding, between language behavior of normals and of aphasics, and so forth. By losing sight of these, the investigator may deprive himself of arriving at new insights. It is our conviction that both theory and applications may gain by research designed within a broad conceptual framework.

REFERENCES

Aborn, M., and Rubenstein, H., "Information theory and immediate recall" *J. exp. Psychol.*, 44 (1952), 260 266.

——, "Word class distribution in sentences of fixed length", *Language*, 32 (1956), 666-674.

Aborn, M., Rubenstein, H., and Sterling, T. D., "Sources of constraint upon words in sentences", *J. exp. Psychol.*, 57 (1959), 171-180.

Allan, M. D., "Memorizing, recoding and perceptual organization", *Brit. J. Psychol.*, 52 (1961), 25-30.

Bar-Hillel, Y., *Language and Information: Selected Essays on their Theory and Application* (Reading, Mass. and Jerusalem, Addison-Wesley and Jerusalem Academic Press, 1964).

Barik, H. C., and Fillenbaum, S., "Negative reinforcement of two grammatical response classes", *Canad. J. Psychol.*, 15 (1961), 107-115.

Barik, H. C., and Lambert, W. E., "Conditioning of complex verbal sequences", *Canad. J. Psychol.*, 14 (1960), 87-95.

Bélanger, D., "'Gradients' musculaires et processus mentaux superieurs", *Canad. J. Psychol.*, 11 (1957), 113-122.

Bellugi, U., and Brown, R. (Eds.), "The acquisition of language", *Monogr. society res. child dev.*, 29 (1964), 1-192.

Bernstein, M. R., "Relationship between interest and reading comprehension", *J. ed. Research*, 49 (1955/56), 283-288.

Bitterman, M. E., "Heart rate and frequency of blinking as indices of visual efficiency", *J. exp. Psychol.*, 35 (1945), 279-292.

Bloomer, R., "Concepts of meaning and the reading and spelling difficulty of words", *J. ed. Research*, 54 (1961), 178-182.

Bloomer, R. H., "Level of abstraction as a function of modifier load", *J. ed. Research*, 52 (1959), 269-272.

Blumenthal, A., "Prompted recall of sentences" (unpublished ms.).

——, and Boakes, R., "Prompted recall of sentences: A further study" (unpublished ms.).

Brinton, J. E., and Danielson, W. A., "A factor analysis of language elements affecting readability", *Journalism Quarterly*, 35 (1958), 420-426.

Broadbent, D. E., *Perception and Communication* (London, Pergamon, 1958).

Brown, J., "Some tests of the decay theory of immediate memory", *Quart. J. exp. Psychol.*, 10 (1958), 12-21.

Bruce, D. J., "The effect of listeners' anticipation on the intelligibility of heard speech", *Language & Speech*, 1 (1958), 79-97.

Bryan, W. L., and Harter, N., "Studies in the telegraphic language", *Psychol. Review*, 6 (1899), 345-375.

Buswell, G. T., "An experimental study of the eye-voice-span in reading", *Supplem. ed. Monogr.*, no. 17 (Chicago, Univ. of Chicago Press, 1920).

Carlson, T. R., "The relationship between speed and accuracy of comprehension", *J. ed. Research*, 42 (1949), 500-512.

Carmichael, L., and Dearborn, W. F., *Reading and Visual Fatigue* (Boston, Houghton-Mifflin, 1947).

Carrol, J. B., *The Study of Language: A Survey of Linguistics and Related Disciplines in America* (Cambridge, Harvard University Press, 1953).

Chall, J. S., *Readability: An appraisal of research and application* (Columbus, Ohio, Ohio State Univ., 1958).

Chansky, N. M., "Threat, anxiety and reading behavior", *J. ed. Research*, 51 (1958), 333-340.

Chapanis, A., "The reconstruction of abbreviated printed messages", *J. exp. Psychol.*, 48 (1954), 496-510.

Cherry, C., *On Human Communication* (New York, Technology Press and Wiley, 1957).

Chomsky, N., "On certain formal properties of grammars", *Information and Control*, 2 (1953), 133-167.

——, *Syntactic Structures* (The Hague, Mouton, 1957).

——, "On the notion 'rule of grammar'", in: R. Jakobson (Ed.), *Structure of Language and its Mathematical Aspects* (American Mathematical Society, 1961), pp. 6-24 (a).

——, "Some methodological remarks on generative grammars", *Word*, 17 (1961), 219-239 (b).

——, "Explanatory models in linguistics", in: E. Nagel, P. Suppes, & A. Tarsky (Eds.), *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress* (Stanford, Calif., Stanford Univ. Press, 1962).

——, and Miller, G. A., "Introduction to the formal analysis of natural languages", in: R. D. Luce, R. R. Bush, & E. Galanter (Eds.) *Handbook of Mathematical Psychology* (New York, Wiley, 1963), vol. 2, pp. 269-323.

Clifton, C. Jr., Kurcz, I., and Jenkins, J. J., "Grammatical relations as determinants of sentence similarity", *Studies in Verbal Behavior*, Report No. 11 (Univ. of Minnesota, 1964).

Clifton, C. Jr., and Odom, P., "Similarity relations among certain English sentence constructions" (unpublished ms.).

Coleman, E. B., "Improving comprehensibility by shortening sentences", *J. appl. Psychol.*, 46 (1962), 131-134.

——, "The comprehensibility of several grammatical transformations", *J. appl. Psychol.*, 48 (1964), 186-190.

——, "Learning of prose written in four grammatical transformations", *J. appl. Psychol.* (in press).

——, and Kim, I., "Comparison of several types of typography in English", *J. appl. Psychol.*, 45 (1961), 262-267.

Conrad, R., "Very brief delay of immediate recall", *Quart. J. exp. Psychol.*, 12 (1960), 45-47.

——, and Hille, B. H., "The decay theory of immediate memory and paced recall", *Canad. J. Psychol.*, 12 (1958), 1-6.

Conway, J. K., and Quarrington, B. J., "Positional effects in the stuttering of contextually organized verbal material", *J. abn. soc. Psychol.*, 67 (1963), 299-303.

Cox, D. R., *Planning of Experiments* (New York, Wiley, 1958).

Crossman, E. R. F. W., "Information and serial order in human immediate memory", in: Cherry (Ed.), *Information Theory: Fourth London Symposium* (Washington, Butterworth, 1961).

Davis, F. B., "Fundamental factors of comprehension in reading", *Psychometrika*, 9 (1944), 185-197.

Derrick, C., *Three aspects of reading comprehension as measured by tests of different lengths* (Princeton, N. J., Educ. Testing Service, RB53-8, 1953).

Doob, L. W., "The effect upon recall of syntactical constructions in German", *J. soc. Psychol.*, 57 (1962), 133-141.

Drew, G. C., "Variations in reflex blink-rate during visual-motor tasks", *Quart. J. exp. Psychol.*, 3 (1951), 73-88.

Edfeldt, A. W., *Silent Speech and Silent Reading* (Stockholm, Almqvist-Wiksell, 1959).

Eifermann, R., "Negation: a linguistic variable", *Acta Psychologica*, 17 (1960), 258-273.

Epstein, I., *Hegyoney Lashon* (Tel Aviv, Am Oved, 1947).

Epstein, W., "The influence of syntactical structure on learning", *Amer. J. Psychol.*, 74 (1961), 80-85.

——, "A further study of the influence of syntactical structure on learning", *Amer. J. Psychol.*, 75 (1962), 121-126.

——, "Temporal schemata in syntactically structured material", *J. general Psychol.*, 68 (1963), 157-164.

Ervin, S. M., and Miller, W. R., "Language development", in: H. W. Stevenson (Ed.), *Child Psychology* (Chicago, Univ. Chicago Press, 1963).

Fairbanks, G., "The relation between eye movements and voice in the oral reading of good and silent readers", *Psychol. Monogr.*, 48 (1937) (Whole No. 215), 78-107.

Fisher, M. S., "Language patterns of preschool children", *Child Developm. Monogr.*, 15 (1934), 1-88.

Fodor, J., and Katz, J. J. (Eds.), *The Structure of Language: Readings in the Philosophy of Language* (New York, Prentice Hall, 1964).

Fónagy, I., "Communication in poetry", *Word*, 17 (1961), 194-218.

Fraser, C., Bellugi, U., and Brown, R., "Control of grammar in imitation, comprehension and production", *J. verb. learning verb. behav.*, 2 (1963), 121-135.

Frick, F. C., "Some perceptual problems from the point of view of information theory", in: B. Mcmillan (Ed.), *Current Trends in Information Theory* (Pittsburgh, Univ. of Pittsburgh Press, 1953), 76-91.

Glanzer, M., "Grammatical category: a rote learning and word association analysis", *J. verb. learning verb. behav.*, 1 (1962), 31-41.

Goldman-Eisler, F., "Speech production and the predictability of words in context", *Quart. J. exp. Psychol.*, 10 (1958), 96-106.

——, "Hesitation, information and levels of speech production", in: A. V. S. de Reuck, M. O'connor (Eds.), *Disorders of Language* (London, Churchill, 1964).

Goodglass, H., and Hunt, J., "Grammatical complexity and aphasic speech", *Word*, 14 (1958), 197-207.

Gorman, A. M., "Recognition memory of nouns as a function of abstractness and frequency", *J. exp. Psychol.*, 61 (1961), 23-29.

Gough, P. B., "The verification of sentences: The effects of delay of evidence and sentence length" (unpublished ms.).

——, "Grammatical transformations and the speed of understanding" (unpublished ms.).

Griffin, P. F., "Reader comprehension of news stories: a preliminary study", *Journalism Quarterly*, 46 (1949), 389-396.

Guttman, L., "A faceted definition of intelligence", *Scripta Hierosolymitana, Studies in Psychology*, 14 (Jerusalem, Magnes Press, 1965), 166-181.

Halle, M., and Stevens, K. N., "Analysis by synthesis", *Proceedings of the Seminar in Speech Transmission and Processing.* AFCRC-TR-59-198 (1959).

Harris, C. W., "The measurement of comprehension of literature", 56, *School Review* (1948), 280-289.

Harwood, K. A., "Listenability and readability", *Speech Monogr.*, 22 (1955), 49-53.

Harwood, F. W., "Quantitative study of syntax of the speech of Australian children", *Language and Speech*, 2 (1959), 236-271.

Hebb, D. O., and Bindra, D., "Scientific writing and the general problem of communication", *Amer. Psychologist*, 7 (1952), 569-573.

Hill, A. A., "Grammaticality", *Word*, 17 (1961), 1-10.

Hockett, C. F., "Grammar for the hearer", in: R. Jakobson (Ed.), *Structure of Language and its Mathematical Aspects* (American Mathematical Society, 1961), 220-236.

Hogan, Y. A., "Copying redundant messages", *J. exp. Psychol.*, 62 (1961), 153-157.

Holland, B. F., "The effect of length and structure of sentences on the silent reading process", *Psychol. Bull.*, 30 (1933), 668-669.

Howes, D., "On the relation between intelligibility and frequency of occurrence of English words", *J. accoust. Soc. Amer.*, 29 (1957), 296-305.

Howes, D. H., and Solomon, R. L., "Visual duration thresholds as a function of word probability", *J. exp. Psychol.*, 41 (1951), 401-410.

Huttenlocher, J., "Children's language: word-phrase relationship", *Science*, 143 (1964), 3603.

Jones, M. H., and Carterette, E. C., "Redundancy in children's free-reading choices", *J. verb learning verb. behav.*, 2 (1963), 489-493.

Kainz, F., *Psychologie der Sprache* (Stuttgart, F. Enke, 1956).

Katz, J. J., and Fodor, J. A., "The structure of a semantic theory", *Language*, 39 (1963), 170-210.

Keller, B , and McGuigan, F. J., "Silent speech during silent reading", *Amer. Psychologist*, 18 (1963), 362. (Abstract).

Kershner, A. M., "Speed of reading in an adult population under differential conditions", *J. appl. Psychol.*, 48 (1964), 25-28.

King-Ellison, P., and Jenkins, J. J., "The durational threshold of visual recognition as a function of word-frequency", *Amer. J. Psych.*, 67 (1954), 700-703.

Klare, G. R., *The Measurement of Readability* (Ames, Iowa, Iowa State Univ. Press, 1963).

——, Mabry, J. E., and Gustafson, L. M., "The relationship of verbal communication variables to immediate and delayed retention and to acceptability of technical training materials", *USAF Person. Training Center*, AFPTRC-TR-54-103, 1954.

——, Mabry, J. E., and Gustafson, L. M., "The relationship of style difficulty to immediate retention and to acceptability of technical material", *J. ed. Psychol.*, 46 (1955), 287-295.

——, Shuford, E. H. & Nichols, W. H., "The relationship of style difficulty, practice, and ability to efficiency of reading and to retention", *J. appl. Psychol.*, 41 (1957), 222-226.

Ladefoged, P., and Broadbent, D. E., "Perception of sequence in auditory events", *Quart. J. exp. Psychol.*, 12 (1960), 162-170.

Lane, H., and Schneider, B., "Some discriminative properties of syntactic structures", *J. verb. learning verb. behav.*, 2 (1963), 457-461.

Larsen, R. P., and Feder, D. D., "Common and differential factors in reading and hearing comprehension", *J. ed. Psychol.*, 31 (1940), 241-252.

Lawson, E. A., "A note on the influence of different orders of approximation to the English language on eye-voice-span, *Quart. J. exp. Psychol.*, 13 (1961), 53-55.

Ledbetter, F. G., "Reading reactions for varied types of subject matter: an analytical study of the eye movements of eleventh grade pupils", *J. ed. Research*, 41 (1947), 102-115.

Lenneberg, E. H., "Understanding language without ability to speak: a case report", *J. abn. soc. Psychol.*, 65 (1962), 419-425.

Leopold, W. F., "Patterning in children's language learning", *Language Learning*, 5 (1953/54), 1-14.

Letson, C. T., "Speed and comprehension in reading", *J. ed. Research*, 52 (1958), 49-53.

Levitt, E. E., "A methodological study of the preparation of connected verbal stimuli for quantitative memory experiments", *J. exp. Psychol.*, 52 (1956), 33-38.

Lloyd, K. E., Reid, L. S., and Feallock, J. B., "Short term retention as a function of the average number of items presented", *J. exp. Psychol.*, 60 (1960), 201-207.

Luckiesh, M., and Moss, F. K., "A summary of researches involving blink rate as a criterion of ease of seeing", *Transact. Illumin. Engin.*, 35 (1940), 19-32.

McClay, H., and Osgood, C. E., "Hesitation phenomena in spontaneous English speech", *Word*, 15 (1959), 19-44.

McClay, H., and Sleator, M. D., "Responses to language: judgments of grammaticalness", *Intern. J. Amer. Linguistics*, 26 (1960), 275-282.

McCracken, R. A., "An experiment with contrived readability in fifth and sixth grades", *J. ed. Research*, 52 (1959), 277-278.

McGinnies, E., Comer, P. B., and Lacey, O. L., "Visual Recognition thresholds as a function of word length and word frequency", *J. exp. Psychol.*, 44 (1952), 65-69.

Mandler, G., "Associative frequency and associative prepotency as measures of response to nonsense syllables", *Amer. J. Psychol.*, 68 (1955), 662-665.

Mandler, G., and Mandler, J. M., "Serial position effects in sentences", *J. verb. learning verb. behav.*, 3 (1964), 195-202.

Marks, L. E., and Miller, G. A., "The role of semantic and syntactic constraints in the memorization of English sentences", *J. verb. learning verb. behav.*, 3 (1964), 1-5.

Marshall, J. C., Unpublished Ph. D. thesis. University of Oxford.

Mehler, J., "Some effects of grammatical transformations on the recall of English sentences", *J. verb. learning verb. behav.*, 2 (1963), 346-351.

Mehler, J., "How some sentences are remembered", Unpublished Ph. D. thesis, Harvard University, 1964.

——, and Miller, G. A., "Retroactive interference in the recall of simple sentences", *British J. Psychol.*, 55 (1964), 295-303.

Menyuk, P., "A preliminary evaluation of grammatical capacity in children", *J. verb. learning verb. behav.*, 2 (1963), 429-439. (a).

——, "Syntactic structures in the language of children", *Child Development*, 45 (1963), 407-423. (b).

——, "Syntactic rules used by children from pre-school through first grade", *Child Development*, 35 (1964), 533-546.

Miller, G. A., *Language and Communication* (New York, McGraw-Hill., 1951).

——, "The magical number seven plus or minus two: some limits on our capacity for processing information", *Psychol. Review*, 63 (1956), 81-97.

——, "Free recall of redundant strings of letters", *J. exp. Psychol.*, 56 (1958), 485-491.

——, "How long is a sentence?", in: *Proceed. of the sixteenth Intern. Congress of Psychology* (Bonn, 1960), pp. 543-545.

——, "Some psychological studies of grammar", *Amer. Psychologist*, 17 (1962), 748-762.

——, and Chomsky, N., "Finitary models of language users", in: R. D. Luce, R. R. Bush & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (New York, Wiley, 1963), Vol. 2, pp. 419-493.

——, Galanter, E., and Pribram, K. H., *Plans and the Structure of Behavior* (New York,, Holt-Dryden, 1960).

——, Heise, G. A., and Lichten, W., "The intelligibility of speech as a function of the context of the test materials", *J. exp. Psychol.*, 41 (1951), 329-335.

——, and Isard, S., "Some perceptual consequences of linguistic rules", *J. verb· learning verb. behav.*, 2 (1963), 217-228.

——, and Isard, S., "Free recall of self-embedded English sentences", *Information and Control*, 7 (1964), 292-303.

——, and McKean, K. O., "A chronometric study of some relations between sentences", *Quart. J. exp. Psychol.*, 16 (1964), 298-307.

——, and Selfridge, J., "Verbal context and the recall of meaningful material", *Amer. J. Psychol.*, 63 (1950), 176-185.

Miller, J. G., "Information input overload", in: M. C. Yovitz, G. T. Jacobi, and G. D. Goldstein (Eds.), *Self-Organizing Systems* (Washington, Spartan Books, 1962), pp. 61-78.

Moray, N., and Taylor, A. M., "The effect of redundancy in shadowing one of two dichotic messages", *Language and Speech*, 1 (1958), 102-109.

Morrison, H. M., and Black, J. W., "Prediction of missing words in sentences", *J. speech Hearing Disord.*, 22 (1957), 236-240.

Morton, J., "The effects of context upon speech of reading, eye movements and eye-voice span", *Quart. J. exp. Psychol.*, 16 (1964), 340-354.

Nahinsky, I. D., "The influence of certain typographical arrangements upon span of visual comprehension", 40, *J. appl. Psychol.* (1956), 37-39.

Nahshon, S., "Readability measurement of Hebrew prose", unpublished Ph. D. thesis, Teachers College Columbia University, 1957.

Newbigging, P. L., "The perceptual redintegration of frequent and infrequent words", *Canad. J. Psychol.*, 15 (1961), 123-132.

Nicol, M. A., and Miller, K. M., "Word redundancy in written English", *Austr. J. Psychol.*, 11 (1959), 81-91.

North, A. J., and Jenkins, L. B., "Reading speech and comprehension as a function of typography", *J. appl. Psychol.*, 35 (1951), 225-228.

Osgood, C. E., "A behavioristic analysis of perception and language as cognitive phenomena", in: *Contemporary Approaches to Cognition; Colorado Symposium* (Cambridge, Mass., Harvard Univ. Press, 1957), pp. 75-119.

——, "On understanding and creating sentences", *American Psychologist*, 18 (1963), 735-751.

——, "Psycholinguistics", in: S. Koch (Ed.), *Psychology: A Study of a Science* (New York, McGraw-Hill, 1963), Vol. 6, pp. 244-317.

Osgood, C. E., and Miron, M. S. (Eds.), *Approaches to the Study of Aphasia* (Urbana, University of Illinois Press, 1963).

——, and Sebeok, T. A. (Eds.), "Psycholinguistics: a survey of theory and research problems", *J. abn. soc. Psychol.*, 1954, supplement.

——, and Walker, E., "Motivation and language behavior: a content analysis of suicide notes", *J. abn. soc. Psychol.*, 54 (1959), 58-67.

Pierce, J. R., and Karlin, J. E., "Reading rates and the information rate of a human channel", *Bell Systems Techn. J.*, 36 (1957), 497-516.

Pool, I., *Trends in Content Analysis* (Urbana, University of Illinois Press, 1959).

Postman, L., and Adams, P. A., "Studies in incidental learning: VIII. The effects of contextual determination", *J. exp. Psychol.*, 59 (1960), 153-164.

——, and Rosenzweig, M. R., "Practice and transfer in the visual and auditory recognition of verbal stimuli", *Amer. J. Psychol.*, 69 (1956), 209-226.

Pollack, I., Johnson, L. B., and Knaff, P. R., "Running memory span", *J. exp. Psychol.*, 57 (1959), 137-146.

Poulton, E. C., "A note on printing to make comprehension easier", *Ergonomics*, 3 (1960), 245-248.

——, and Gregory, R. L., "Blinking during visual tracking", *Quart. J. exp. Psychol.*, 4 (1952), 57-65.

Preston, K. A., "The speed of word perception and its relation to reading ability", *J. general Psychol.*, 13 (1935), 199-203.

Putnam, H., "Some issues in the theory of grammar", in: R. Jakobson (Ed.), *Structure of Language and its Mathematical Aspects* (Mathematical Society, 1961), pp. 25-42.

Quarrington, B., "Stuttering as a function of the information value of sentence position of words", *J. abn. Psychol.*, 70 (1965), 221-224.

——, Conway, J., and Siegel, N., "An experimental study of some properties of stuttered words", *J. Speech Hearing Res.*, 5 (1962), 381-394.

Reid, L. S., Lloyd, K. E., Brackett, H. R., and Hawkins, W. F., "Short term retention as a function of average storage load and average load reduction", *J. exp. Psychol.*, 62 (1961), 518-522.

Riffaterre, M., "Criteria for style analysis", *Word*, 15 (1959), 154-174.

Robinson, F. P., "'Speed *versus* comprehension in reading' – a discussion", *J. ed. Psychol.*, 31 (1940), 554-558.

——, and Hall, P., "Studies of higher-level reading abilities", *J. ed. Psychol.*, 32 (1941), 241-252.

Rosenzweig, M. R., and Postman, L., "Frequency of usage and the perception of words", *Science*, 127 (1958), 263-266.

Rothkopf, E. Z., "Learning from written sentences: within-sentence order in the acquisition of name-clause equivalences", *J. verb. learning verb. behav.*, 2 (1963), 470-475.

Rubenstein, H., and Aborn, M., "Immediate recall as a function of degree of organization and length of study period", *J. exp. Psychol.*, 48 (1954), 146-152.

——, and Aborn, M., "Learning, prediction, and readability", *J. appl. Psychol.*, 42 (1958), 28-32.

——, and Aborn, M., "Psycholinguistics", *Ann. Rev. Psychol.*, 1 (1960), 291-322.

——, and Pickett, J. M., "Intelligibility of words in sentences", *J. accoust. Soc. Amer.*, 30 (1958), 670.

Savin, H. B., and Perchonok, E., "Grammatical structure and the immediate recall of English sentences", (unpublished ms.).

Schlesinger, I. M., "The effect of stimulus redundancy in a card sorting task", *Scripta Hierosolymitana: Studies in psychology*, 14 (1965), 216-230. (a).

——, "Discriminability as a function of figure complexity", *J. general Psychol.*, 73 (1965), 21-29. (b).

——, Forte, M., Fried, B., and Melkman, R., "Stuttering, information load, and response strength", *J. Speech Hear. Disord.*, 30 (1965), 32-36.

Schramm, W., "Measuring another dimension of newspaper readership", *Journalism Quarterly*, 24 (1947), 293-306.

Shannon, C. E., and Weaver, W., *Mathematical Theory of Communication* (Urbana, Univ. Illinois Press, 1949).

Sharp, H. C., "The effect of contextual constraint upon recall of verbal passages", *Amer. J. Psychol.*, 71 (1958), 568-577.

Shipstone, E. I., "Some variables affecting pattern conception", *Psychol. Monogr.*, 74 (1960), (17, Whole no. 504), 42-60.

Siegel, S., *Nonparametric Statistics for the Behavioral Sciences* (New York, McGraw-Hill, 1956).

Slobin, D. I., "Grammatical transformations and sentence comprehension in childhood and adulthood" (unpublished ms.).

——, "Soviet psycholinguistics", in: N. O'Connor (Ed.), *Present Day Russian Psychology* (in press).

Smith, A. A., Malmo, R. B., and Shagass, C., "An electromyographic study of listening and talking", *Canad. J. Psychol.*, 8 (1954), 219-227.

Stevens, S. S., and Stone, G., "Psychological writing, easy and hard", *Amer. Psychologist*, 2 (1947), 230-235.

Stolurow, L. M., Newman, J. R., "A factional analysis of objective features of printed language presumably related to reading difficulty", 52, *J. ed. Res.* (1959), 243-251.

Stroud, J. B., and Henderson, M., "Rate of reading and learning by reading", *J. ed. Psychol.*, 34 (1943), 193-205.

Taylor, W., "'Cloze procedure', a new tool for measuring readability", *Journalism Quart.*, 30 (1953), 415-433.

——, "Recent developments in the use of 'cloze procedure'", *Journalism Quart.*, 33 (1956), 42-48.

Taylor, W. L., "'Cloze' readability scores and indices of individual differences in comprehension and aptitude", *J. appl. Psychol.*, 41 (1957), 19-26.

Thurstone, L. L., "Note on a reanalysis of Davis's reading tests", *Psychometrika*, 11 (1946), 185-188.

Tinker, M. A., "Speed versus comprehension in reading as affected by level of difficulty", *J. ed. Research*, 30 (1939), 81-94.

——, "Validity of frequency of blinking as a criterion of readability", *J. exp. Psychol.*, 36 (1946), 453-460.

——, "Recent studies of eye-movements in reading", *Psychol. Bull.*, 4 (1958), 215-231.

Tulving, E., and Patkau, J. E., "Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material", *Canad. J. Psychol.*, 16 (1962), 83-95.

Vernon, P. E., "The determinants of reading comprehension". *Educ. Psycholog. Measurement*, 22 (1962), 269-286.

Wallerstein, H., "An electromyographic study of attentive listening", *Canad. J. Psychol.*, 8 (1954), 228-238.

Wason, P. C., "The processing of positive and negative information", *Quart. J. exp. Psychol.*, 11 (1959), 92-107.

——, "Response to affirmative and negative binary statements", *Brit. J. Psychol.*, 52 (1961), 133-142.

——, "The retention of material presented through précis", *J. Commun.*, 12 (1962), 36-43.

——, and Jones, S., "Negatives: denotation and connotation", *Brit. J. Psychol.*, 54 (1963), 229-307.

——, "The contexts of plausible denial", *J. verb. learning verb. behav.*, 4 (1965), 7-11.

Weir, R. H., *Language in the Crib* (The Hague, Mouton, 1962).

Wisseman, H., "Die Rolle des Grammatischen beim Verstehen des Satzsinnes", *Proceed. 16th Internat. Congress Psychol.* (Bonn, 1960).

Wonderly, W. L., "Some factors of meaningfulness in reading matter for inexperienced readers", *A William Cameron Townsend en el XXV Aniversario del I.L.V.* (n. d.).

Wood, C. L., and Bitterman, M. E., "Blinking as a measure of effort on visual work", *Amer. J. Psychol.*, 63 (1950), 584-588.

Yngve, V. H., "A model and an hypothesis for language structure", *Proceed. Amer. Philos. Soc.*, 104 (1960), 444-466.

Zachrisson, B., *Studies in the Legibility of Printed Text* (Uppsala, Almqvist & Wiksells, 1965).